Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Testing Independence and Conditional Independence in High Dimensions via Coordinatewise Gaussianization

Jinyuan Chang, Yue Du, Jing He & Qiwei Yao

To cite this article: Jinyuan Chang, Yue Du, Jing He & Qiwei Yao (10 Mar 2026): Testing Independence and Conditional Independence in High Dimensions via Coordinatewise Gaussianization, Journal of the American Statistical Association, DOI: 10.1080/01621459.2026.2637891

To link to this article: https://doi.org/10.1080/01621459.2026.2637891

Testing Independence and Conditional Independence in High Dimensions via Coordinatewise Gaussianization

Jinyuan Chang [1,2,3], Yue Du[1], Jing He[4], and Qiwei Yao[5]

[1] *Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Chengdu, China*
[2] *State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*
[3] *China Center for Economic Research, Peking University, Beijing, China*
[4] *Institute of Statistical Interdisciplinary Research, Southwestern University of Finance and Economics, Chengdu, China*
[5] *Department of Statistics, The London School of Economics and Political Science, London, U.K.*

*q.yao@lse.ac.uk

Abstract

We propose new statistical tests, in high-dimensional settings, for testing the independence of two random vectors and their conditional independence given a third random vector. The key idea is simple, i.e., we first transform each component variable to the standard normal via its marginal empirical distribution, and we then test for independence and conditional independence of the transformed random vectors using appropriate L_∞ -type test statistics. While we are testing some necessary conditions of the independence or the conditional independence, the new tests outperform the 13 frequently used testing methods in a large scale simulation comparison. The advantage of the new tests can be summarized as follows: (i) they do not require any moment conditions, (ii) they allow arbitrary dependence structures of the components among the random vectors, and (iii) they allow the dimensions of random vectors to diverge at the exponential rates of the sample size. The critical values of the proposed tests are determined by a computationally efficient multiplier bootstrap procedure. Theoretical analysis shows that the sizes of the proposed tests can be well controlled by the nominal significance level, and the proposed tests are also consistent under certain local alternatives. The finite sample performance of the new tests is illustrated via extensive simulation studies and a real data application.

Keywords: Conditional independence test, coordinatewise Gaussianization, Gaussian approximation, high-dimensional statistical inference, independence test, multiplier bootstrap.

1 Introduction

Let $\mathbf{X} \in \mathbb{R}^p$, $\mathbf{Y} \in \mathbb{R}^q$ and $\mathbf{Z} \in \mathbb{R}^m$ be three random vectors. Given samples $\{(\mathbf{X}_i, \mathbf{Y}_i, \mathbf{Z}_i)\}_{i=1}^n$ with $(\mathbf{X}_i, \mathbf{Y}_i, \mathbf{Z}_i) \overset{\text{i.i.d.}}{\sim} (\mathbf{X}, \mathbf{Y}, \mathbf{Z})$, we are interested in the following two hypothesis testing problems:

• (Hypothesis testing for independence)

$$\mathbb{H}_0 : \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \quad \text{versus} \quad \mathbb{H}_1 : \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}. \quad (1)$$

• (Hypothesis testing for conditional independence)

$$\mathbb{H}_0 : \mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z} \quad \text{versus} \quad \mathbb{H}_1 : \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}. \quad (2)$$

These two testing problems are of direct application in, among others, building statistical models including feature selection and simplification, causal inference, and understanding complex relationships in machine learning and data analysis for various practical problems. Due to their immense importance, a large number of the testing methods have been developed. In spite of this, we argue that there is still a justification for the proposed tests in this paper. Indeed the existing methods have demonstrated the successes under various settings and conditions, but none of them is predominately better than the others. Though it is prohibitively difficult, if not impossible, to construct a universally optimal test, we propose a new test, for each of (1) and (2) respectively, under some mild conditions in high-dimensional settings, and they uniformly outperform the 13 frequently used tests in the extensive simulation studies.

Our new tests are based on coordinatewise Gaussianization and Gaussian approximation (Chernozhukov et al., 2017; Chang et al., 2024a) in the high-dimensional settings. Assuming all the marginal distributions of $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ are continuous, we transform each component of $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ to a standard normal random variable by its distribution function. Let $\mathbf{U}, \mathbf{V}$ and $\mathbf{W}$ be the transformed vectors of, respectively, $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$. We adopt the maximum absolute pairwise sample covariance between the components of $\mathbf{U}$ and those of $\mathbf{V}$ as the statistic for testing independence hypothesis (1). Under the null hypothesis $\mathbb{H}_0$ in (1), all the covariances between the components of $\mathbf{U}$ and those of $\mathbf{V}$ are 0. But the converse is not necessarily true. For testing conditional independence hypothesis (2), we first fit regression models of $\mathbf{U}$ and $\mathbf{V}$ on $\mathbf{W}$, and then adopt the maximum absolute pairwise sample covariance between the components of the residuals for $\mathbf{U}$ and those for $\mathbf{V}$ as the statistic. Again we are testing a necessary condition under the null hypothesis $\mathbb{H}_0$ in (2). Nevertheless, the extensive simulation studies in Section 7 show that the proposed tests uniformly outperform the 13 frequently used tests.

The null-distributions of the test statistics are evaluated in terms of the Gaussian approximation technique, which is implemented by a computationally efficient multiplier bootstrap scheme for computing the critical values of the tests. Our theoretical analysis shows that the sizes of the new tests can be correctly controlled by the prescribed nominal significance level, and they are also consistent under certain local alternatives.

The advantage exhibited by the proposed tests can be summarized as follows: (a) They require no moment conditions on $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, and, hence, can be applied to heavy-tailed distributions. (b) They allow arbitrary dependence structures among the components of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$. (c) The dimensions of $\mathbf{X}$ and $\mathbf{Y}$ can diverge at the exponential rates of the sample size, and the dimension of $\mathbf{Z}$ can diverge at a polynomial rate of the sample size.

The coordinatewise Gaussianization is a widely used technique in statistical analysis, especially in high-dimensional settings. See, for example, Liu et al. (2009), Xue and Zou (2012) and Mai and Zou (2015) for applications of coordinatewise Gaussianization in high-dimensional Gaussian graphical models and sufficient dimension reduction, and Mai et al. (2023) for the theoretical guarantee of the coordinatewise Gaussianization methods.

The literature on the tests of independence and conditional independence is large. The independence test has been well studied in the low-dimensional scenario. For example, Spearman (1904), Pearson (1920), Kendall (1938), Blum et al. (1961), and Reshef et al. (2011) propose various dependence measures when $p=q=1$. Wilks (1935), Hotelling (1936), Puri and Sen (1971), Hettmansperger and Oja (1994), Gieser and Randles (1997), Taskinen et al. (2003) and Taskinen et al. (2005) investigate the tests under the Gaussian or elliptically symmetric distributions with fixed (p,q). Gretton et al. (2008) consider a test based on the Hilbert-Schmidt independence criterion (HSIC). Bergsma and Dassios (2014) propose a consistent test based on a sign covariance. Lyons (2013) and Jakobsen (2017) deal with the tests in more general metric spaces. In the high-dimensional scenario with $p,q \gg n$, the distance correlations for characterizing the dependence between $\mathbf{X}$ and $\mathbf{Y}$ have been proposed and the associated testing procedures for (1) have been studied. See Székely et al. (2007), Székely and Rizzo (2013), Zhu et al. (2020) and Gao et al. (2021). All the tests aforementioned require certain moment conditions on $\mathbf{X}$ and $\mathbf{Y}$. To alleviate the moment restrictions, a projection correlation based test is considered by Zhu et al. (2017), and some rank-based tests are presented by Heller et al. (2013), Shi et al. (2022) and Deb and Sen (2023).

Testing independence (1) is a special case of testing whether $\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(\ell)}$ are mutually independent with $\ell=2$, where $\mathbf{X}^{(1)} \in \mathbb{R}^{p_1},\ldots,\mathbf{X}^{(\ell)} \in \mathbb{R}^{p_\ell}$ are ℓ random vectors. Many existing works in the literature focus on this more general setting. When $p_1=\cdots=p_\ell=1$, Pfister et al. (2018) extend the HSIC test (Gretton et al., 2008) to ℓ-variate HSIC for $\ell>2$. See also Matteson and Tsay (2017). Han et al. (2017), Leung and Drton (2018) and Yao et al. (2018) propose mutual independence tests for $\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(\ell)}$ when $\ell \gg n$. When $p_1,\ldots,p_\ell>1$, Jin and Matteson (2018) propose a test based on generalized distance covariance. Chakraborty and Zhang (2019) use joint distance covariance to quantify and to test the joint independence among ℓ random vectors. See Chang et al. (2024b) for a more general form of this testing problem.

Testing conditional independence (2) is more challenging, which relies on the properties of the controlling variables $\mathbf{Z}$. There is also abundant literature on the conditional independence tests for fixed (p,q,m). For the simplest case with $p=q=1$ and fixed m, partial correlation (Lawrance, 1976) is the most commonly used measure for the conditional dependence between two normal variables with the effects of controlling variables being removed. However, in the non-Gaussian case, zero partial correlation coefficient is not necessarily equivalent to conditional independence. Various nonparametric tests have been developed in the literature, including Kendall (1942), Goodman (1959), Veraverbeke et al. (2011), Azadkia and Chatterjee (2021) and Otneim and Tjøstheim (2022). When $p,q,m>1$, Su and White (2008), Corradi et al. (2012) and Huang et al. (2016) construct the tests by comparing the conditional distributions under the null and the alternative hypotheses. Su and White (2007) and Wang and Hong (2018) introduce tests based on the conditional characteristic functions. Fukumizu et al. (2008), Zhang et al. (2011), Doran et al. (2014) and Strobl et al. (2019) explore extensively various kernel based methods. Runge (2018) propose a test based on conditional mutual information. When *p*, *q* or *m* is potentially large, Berrett et al. (2020) introduce a conditional permutation test, Székely and Rizzo (2014), Wang et al. (2015) and Fan et al. (2020) construct the tests based on the extended conditional distance correlations. Furthermore, Shah and Peters (2020) propose the generalized covariance measure based on the sample covariance between the residuals of the regressions $\mathbf{X}$ and $\mathbf{Y}$ on $\mathbf{Z}$. Although both the test statistics of Shah and Peters (2020) and ours for conditional independence hypothesis (2) are based on the residuals of some regressions, there are several fundamental differences worth noting. First, since Shah and Peters (2020) consider the regressions $\mathbf{X}$ and $\mathbf{Y}$ on $\mathbf{Z}$ directly, it is necessary to impose certain moment conditions on the elements of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, whereas our proposed method considers the regressions $\mathbf{U}$ and $\mathbf{V}$ on $\mathbf{W}$ which essentially eliminates the moment conditions on the elements of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ through coordinatewise Gaussianization, and can offer an advantage in dealing with heavy-tailed data. Second, the methods used to solve the regression problems are different, where Shah and Peters (2020) use kernel ridge regression, and we employ the feedforward neural network. Third, the theoretical guarantee of Shah and Peters (2020) requires that *m* is fixed, while our proposed method allows *m* to diverge with the sample size. Extensive simulation studies in Section 7.2 show that their method may not work for heavy-tailed data, but the proposed method performs exceptionally well with both high-dimensional and heavy-tailed data. For dependent data, Zhou et al. (2022) propose a conditional independence test based on a projection approach.

The rest of the paper is organized as follows. Section 2 introduces the coordinatewise Gaussianization technique. Section 3 introduces the proposed independence test. Section 4 introduces the proposed conditional independence tests based on nonparametric regressions and linear regressions, respectively. Section 5 provides a computationally efficient multiplier bootstrap scheme for computing the critical values of the proposed tests. Section 6 investigates the associated theoretical properties of the proposed tests. Section 7 evaluates the finite-sample performance of the proposed tests via extensive simulation studies. All technical proofs and a real data example are relegated to the supplementary material. The used real data and the code for implementing our proposed tests are available at the GitHub repository: https://github.com/JinyuanChang-Lab/CoordinatewiseGaussianizationTest.

Notation. The notation $I(\cdot)$ denotes the indicator function. For any positive integer *k*, write $[k]=\{1,\ldots,k\}$, and denote by $\mathbf{I}_k$ the $k\times k$ identity matrix. For any $a,b\in\mathbb{R}$, let $\lceil a\rceil$ and $\lfloor a\rfloor$ denote, respectively, the smallest integer greater than or equal to *a*, and the largest integer less than or equal to *a*, and let $a\vee b$ and $a\wedge b$ denote, respectively, the larger and smaller number between *a* and *b*. For a vector $\mathbf{a}=(a_1,\ldots,a_k)^\top\in\mathbb{R}^k$, let $|\mathbf{a}|_0=\sum_{i=1}^k I(a_i\neq 0)$, $|\mathbf{a}|_1=\sum_{i=1}^k |a_i|$, $|\mathbf{a}|_2=(\sum_{i=1}^k a_i^2)^{1/2}$, and $|\mathbf{a}|_\infty=\max_{i\in[k]}|a_i|$ be its L_0-norm, L_1-norm, L_2-norm and L_∞-norm, respectively. For a matrix $\mathbf{A}=(A_{i,j})_{k_1\times k_2}$, we write $|\mathbf{A}|_\infty=\max_{i\in[k_1],\,j\in[k_2]}|A_{i,j}|$. Denote by $\otimes$ the Kronecker product operator between matrices. For any set $\mathcal{S}$, let $|\mathcal{S}|$ denote its cardinality. Let $\mathcal{N}(\boldsymbol{\mu},\mathbf{B})$, $U(a,b)$ and $t(c)$ denote, respectively, multi-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{B}$, the uniform distribution on $[a,b]$, and the *t*-distribution with *c* degrees of freedom. Let $\Phi(\cdot)$ be the cumulative distribution function of the standard normal distribution $\mathcal{N}(0,1)$. For any two sequences of positive numbers $\{a_k\}$ and $\{b_k\}$, we write $a_k \lesssim b_k$ or $b_k \gtrsim a_k$ if $\limsup_{k\to\infty} a_k / b_k < \infty$, and write $a_k \ll b_k$ or $b_k \gg a_k$ if $\limsup_{k\to\infty} a_k / b_k = 0$. Moreover, $a_k \asymp b_k$ means that $a_k \lesssim b_k$ and $b_k \lesssim a_k$ hold simultaneously. The sets of natural numbers, natural numbers including 0 and real numbers are denoted by $\mathbb{N}$, $\mathbb{N}_0$ and $\mathbb{R}$, respectively.

2 Coordinatewise Gaussianization

Let $\mathbf{X} = (X_1,\ldots,X_p)^\top \sim F_{\mathbf{X}}$, $\mathbf{Y} = (Y_1,\ldots,Y_q)^\top \sim F_{\mathbf{Y}}$ and $\mathbf{Z} = (Z_1,\ldots,Z_m)^\top \sim F_{\mathbf{Z}}$ be three generic random vectors. For each $j \in [p]$, $k \in [q]$ and $l \in [m]$, denote by $F_{\mathbf{X},j}(\cdot)$, $F_{\mathbf{Y},k}(\cdot)$ and $F_{\mathbf{Z},l}(\cdot)$, respectively, the distribution functions of X_j, Y_k and Z_l. Assume all $F_{\mathbf{X},j}(\cdot)$, $F_{\mathbf{Y},k}(\cdot)$ and $F_{\mathbf{Z},l}(\cdot)$ are continuous. Then $U_j \equiv \Phi^{-1}\{F_{\mathbf{X},j}(X_j)\}$, $V_k \equiv \Phi^{-1}\{F_{\mathbf{Y},k}(Y_k)\}$ and $W_l \equiv \Phi^{-1}\{F_{\mathbf{Z},l}(Z_l)\}$ are the standard normal random variables. Put $\mathbf{U} = (U_1,\ldots,U_p)^\top$, $\mathbf{V} = (V_1,\ldots,V_q)^\top$ and $\mathbf{W} = (W_1,\ldots,W_m)^\top$. Since $\Phi^{-1}\{F_{\mathbf{X},j}(\cdot)\}$, $\Phi^{-1}\{F_{\mathbf{Y},k}(\cdot)\}$ and $\Phi^{-1}\{F_{\mathbf{Z},l}(\cdot)\}$ are strictly monotone mappings, the hypotheses (1) and (2) are equivalent to, respectively,

$$\mathbb{H}_0 : \mathbf{U} \perp\!\!\!\perp \mathbf{V} \quad \text{versus} \quad \mathbb{H}_1 : \mathbf{U} \not\!\perp\!\!\!\perp \mathbf{V}, \quad (3)$$

and

$$\mathbb{H}_0 : \mathbf{U} \perp\!\!\!\perp \mathbf{V} | \mathbf{W} \quad \text{versus} \quad \mathbb{H}_1 : \mathbf{U} \not\!\perp\!\!\!\perp \mathbf{V} | \mathbf{W}. \quad (4)$$

For each $i \in [n]$, write $\mathbf{X}_i = (X_{i,1},\ldots,X_{i,p})^\top$, $\mathbf{Y}_i = (Y_{i,1},\ldots,Y_{i,q})^\top$ and $\mathbf{Z}_i = (Z_{i,1},\ldots,Z_{i,m})^\top$, and define $\mathbf{U}_i = (U_{i,1},\ldots,U_{i,p})^\top$, $\mathbf{V}_i = (V_{i,1},\ldots,V_{i,q})^\top$ and $\mathbf{W}_i = (W_{i,1},\ldots,W_{i,m})^\top$ with $U_{i,j} = \Phi^{-1}\{F_{\mathbf{X},j}(X_{i,j})\}$, $V_{i,k} = \Phi^{-1}\{F_{\mathbf{Y},k}(Y_{i,k})\}$ and $W_{i,l} = \Phi^{-1}\{F_{\mathbf{Z},l}(Z_{i,l})\}$. Write $\mathcal{X}_n = \{\mathbf{X}_1,\ldots,\mathbf{X}_n\}$, $\mathcal{Y}_n = \{\mathbf{Y}_1,\ldots,\mathbf{Y}_n\}$ and $\mathcal{Z}_n = \{\mathbf{Z}_1,\ldots,\mathbf{Z}_n\}$. Given $(\mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n)$, we can approximate $\mathbf{U}_i$, $\mathbf{V}_i$ and $\mathbf{W}_i$, respectively, by $\mathbf{U}_i = (\hat{U}_{i,1},\ldots,\hat{U}_{i,p})^\top$, $\mathbf{V}_i = (\hat{V}_{i,1},\ldots,\hat{V}_{i,q})^\top$ and $\mathbf{W}_i = (\hat{W}_{i,1},\ldots,\hat{W}_{i,m})^\top$ with

$$\hat{U}_{i,j} = \Phi^{-1}\{\frac{n\hat{F}_{\mathbf{X},j}(X_{i,j})}{n+1}\}, \hat{V}_{i,k} = \Phi^{-1}\{\frac{n\hat{F}_{\mathbf{Y},k}(Y_{i,k})}{n+1}\}, \hat{W}_{i,l} = \Phi^{-1}\{\frac{n\hat{F}_{\mathbf{Z},l}(Z_{i,l})}{n+1}\}, \quad (5)$$

where $\hat{F}_{\mathbf{X},j}(\cdot) = n^{-1}\sum_{s=1}^{n} I(X_{s,j} \le \cdot)$, $\hat{F}_{\mathbf{Y},k}(\cdot) = n^{-1}\sum_{s=1}^{n} I(Y_{s,k} \le \cdot)$ and $\hat{F}_{\mathbf{Z},l}(\cdot) = n^{-1}\sum_{s=1}^{n} I(Z_{s,l} \le \cdot)$.
Multiplying them by $n(n+1)^{-1}$ in (5) is to guarantee $|\hat{U}_{i,j}| < +\infty$, $|\hat{V}_{i,k}| < +\infty$ and $|\hat{W}_{i,l}| < +\infty$. In

Sections 3 and 4, we will propose testing procedures for (1) and (2) based on coordinatewise Gaussianization.

3 Testing for Independence

Note that $\boldsymbol{\gamma}_i \equiv \mathbf{U}_i \otimes \mathbf{V}_i$ is a d-dimensional random vector with $d = pq$, and $\mathbb{E}(\boldsymbol{\gamma}_i) = \mathbf{0}$ under the null hypothesis $\mathbb{H}_0$ in (3). For given $\{(\mathbf{U}_i, \mathbf{V}_i)\}_{i=1}^n$, several studies have considered testing whether $\mathbb{E}(\boldsymbol{\gamma}_i) = \mathbf{0}$ holds; see, e.g., Chang et al. (2017), Yang et al. (2024), and the references therein. In practice, however, $\{(\mathbf{U}_i, \mathbf{V}_i)\}_{i=1}^n$ are unobservable, so we need to construct feasible statistics based on $\{(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i)\}_{i=1}^n$. Let $\mathbf{S}_n = n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\gamma}}_i$ with $\hat{\boldsymbol{\gamma}}_i = \hat{\mathbf{U}}_i \otimes \hat{\mathbf{V}}_i$, where the components of $\hat{\mathbf{U}}_i$ and $\hat{\mathbf{V}}_i$ are specified in (5). The components of $\mathbf{S}_n$ can be viewed as all the pairwise sample covariances between the components of $\mathbf{U}$ and those of $\mathbf{V}$. We consider the test statistic

$$H_n = \sqrt{n}\,|\mathbf{S}_n|_\infty$$

for (3), and reject $\mathbb{H}_0$ at the significance level $\alpha \in (0,1)$ if $H_n > \mathrm{cv}_{\mathrm{ind},\alpha}$, where $\mathrm{cv}_{\mathrm{ind},\alpha}$ is the critical value satisfying $\mathbb{P}(H_n > \mathrm{cv}_{\mathrm{ind},\alpha}) = \alpha$ under $\mathbb{H}_0$.

Let $\Sigma = \mathrm{Cov}(\boldsymbol{\gamma}_i)$, which can be estimated by $\hat{\Sigma} = n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\gamma}}_i \hat{\boldsymbol{\gamma}}_i^\top - \bar{\hat{\boldsymbol{\gamma}}}\bar{\hat{\boldsymbol{\gamma}}}^\top$ with $\bar{\hat{\boldsymbol{\gamma}}} = n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\gamma}}_i$. For any $\alpha \in (0,1)$, Proposition 1 in Appendix A of the supplementary material indicates that $\mathrm{cv}_{\mathrm{ind},\alpha}$ can be approximated by

$$\hat{\mathrm{cv}}_{\mathrm{ind},\alpha} = \inf\{t \geq 0 : \mathbb{P}(|\boldsymbol{\xi}|_\infty \leq t \,|\, \mathcal{X}_n, \mathcal{Y}_n) \geq 1-\alpha\} \quad (6)$$

for $\boldsymbol{\xi} \,|\, \mathcal{X}_n, \mathcal{Y}_n \sim \mathcal{N}(\mathbf{0}, \hat{\Sigma})$. Section 5 will introduce a multiplier bootstrap procedure to determine the critical value for the test, which is computationally efficient in practice. Notice that coordinatewise Gaussianization eliminates the need for moment conditions on $\mathbf{X}$ and $\mathbf{Y}$, which enables the proposed independence test to handle heavy-tailed distributions.

Remark 1 .

The L_∞-type test statistic H_n can be extended to a more general test statistic:

$$\tilde{H}_n = \max_{1 \leq j_1 < \cdots < j_T \leq d} \sum_{t=1}^T \sqrt{n}\,|\hat{S}_{n,j_t}|, \quad (7)$$

where $\mathbf{S}_n = (\hat{S}_{n,1},\ldots,\hat{S}_{n,d})^\top = n^{-1}\sum_{i=1}^{n}\hat{\boldsymbol{\gamma}}_i$. Based on the Gaussian approximation technique (Chernozhukov et al., 2017; Chang et al., 2024a), the associated critical value can be selected as $\inf\{t \ge 0 : \mathbb{P}(\max_{1\le j_1<\cdots<j_T\le d} \sum_{t=1}^{T}|\hat{\xi}_{j_t}| \le t|\mathcal{X}_n,\mathcal{Y}_n) \ge 1-\alpha\}$ with $\boldsymbol{\xi} = (\hat{\xi}_1,\ldots,\hat{\xi}_d)^\top$. When $T=1$, this new test statistic is identical to H_n. Using the test statistic with the form (7) has some technical advantages. More specifically, if T is fixed or diverges slower than some certain rate, the associated test procedure is valid even if p and q diverge exponentially fast with n, which does not require any structural assumptions imposed on the covariance matrices of $\mathbf{U}$ and $\mathbf{V}$. These advantages are quite important in practice.

Remark 2 .

Notice that U_j and V_k are Gaussian random variables for each $j\in[p]$ and $k\in[q]$. Our testing procedure actually selects $|\mathbb{E}(\mathbf{U}\mathbf{V}^\top)|_\infty$ as a measure for $\mathbf{U} \perp\!\!\!\perp \mathbf{V}$. If $\mathbf{U}$ and $\mathbf{V}$ are jointly Gaussian, then $|\mathbb{E}(\mathbf{U}\mathbf{V}^\top)|_\infty = 0$ if and only if $\mathbf{U} \perp\!\!\!\perp \mathbf{V}$. If $\mathbf{U}$ and $\mathbf{V}$ are not jointly Gaussian, our procedure essentially tests a necessary condition of the independence. However, regardless of whether $\mathbf{U}$ and $\mathbf{V}$ are jointly Gaussian or not, Theorem 1 in Section 6.1 shows that our proposed test can always control the size at the significance level. When $\mathbf{U} \not\!\perp\!\!\!\perp \mathbf{V}$, Theorem 2 in Section 6.1 shows that the power of our proposed test will depend on the magnitude of $|\mathbb{E}(\mathbf{U}\mathbf{V}^\top)|_\infty$. We admit that $|\mathbb{E}(\mathbf{U}\mathbf{V}^\top)|_\infty$ may be equal to zero when $\mathbf{U} \not\!\perp\!\!\!\perp \mathbf{V}$. In this case, we may consider to choose a different measure such that $\mathbf{U} \perp\!\!\!\perp \mathbf{V}$ if and only if this measure between $\mathbf{U}$ and $\mathbf{V}$ equals to zero. See Section R.1 in the supplementary material for details.

4 Testing for Conditional Independence

Given $(\mathbf{U}_i, \mathbf{V}_i, \mathbf{W}_i)$, we consider two regression models:

$$\mathbf{U}_i = \mathbf{f}(\mathbf{W}_i) + \boldsymbol{\varepsilon}_i, \quad \mathbf{V}_i = \mathbf{g}(\mathbf{W}_i) + \boldsymbol{\delta}_i, \ (8)$$

where $\mathbf{f}(\mathbf{W}_i) = \mathbb{E}(\mathbf{U}_i|\mathbf{W}_i)$, and $\mathbf{g}(\mathbf{W}_i) = \mathbb{E}(\mathbf{V}_i|\mathbf{W}_i)$. The null hypothesis $\mathbb{H}_0$ in (4) holds if and only if $\boldsymbol{\varepsilon}_i \perp\!\!\!\perp \boldsymbol{\delta}_i|\mathbf{W}_i$. In general, we can estimate $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ in (8) using feedforward neural networks, which will be introduced in Section 4.1. It is well known that estimating nonparametric regression models using feedforward neural networks requires a substantially large number of observations, especially in high-dimensional scenarios. Alternatively, when the sample size *n* is small, we can further consider to fit the following linear models:

$$\mathbf{U}_i = \mathbf{A}\mathbf{W}_i + \boldsymbol{\varepsilon}_i, \quad \mathbf{V}_i = \mathbf{B}\mathbf{W}_i + \boldsymbol{\delta}_i, \ (9)$$

with $\mathbb{E}(\boldsymbol{\varepsilon}_i|\mathbf{W}_i) = \mathbf{0}$ and $\mathbb{E}(\boldsymbol{\delta}_i|\mathbf{W}_i) = \mathbf{0}$. If $(\mathbf{U}_i, \mathbf{V}_i, \mathbf{W}_i)$ is jointly normal, (8) reduces to the linear equations in (9), and the null hypothesis $\mathbb{H}_0$ in (4) holds if and only if $\mathrm{Cov}(\boldsymbol{\varepsilon}_i, \boldsymbol{\delta}_i) = \mathbf{0}$. We will proceed with the linear representation (9) in Section 4.2. The simulation results in Section 7.2 indicate that the proposed conditional independence test based on the linear regressions performs well in most scenarios, and outperforms the proposed conditional independence test based on the nonparametric regressions in most cases when the sample size n is small.

4.1 Conditional Independence Test based on Nonparametric Regressions

Write $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,p})^\top$ and $\boldsymbol{\delta}_i = (\delta_{i,1}, \ldots, \delta_{i,q})^\top$. The component-wise forms of (8) are as follows:

$$U_{i,j} = f_j(\mathbf{W}_i) + \varepsilon_{i,j}, \qquad V_{i,k} = g_k(\mathbf{W}_i) + \delta_{i,k}, \quad (10)$$

where $f_j(\mathbf{W}_i) = \mathbb{E}(U_{i,j} | \mathbf{W}_i)$ and $g_k(\mathbf{W}_i) = \mathbb{E}(V_{i,k} | \mathbf{W}_i)$. Recall $\mathbf{U}_i = (U_{i,1}, \ldots, U_{i,p})^\top$, $\mathbf{V}_i = (V_{i,1}, \ldots, V_{i,q})^\top$ and $\mathbf{W}_i = (W_{i,1}, \ldots, W_{i,m})^\top$ with $U_{i,j} = \Phi^{-1}\{F_{\mathbf{X},j}(X_{i,j})\}$, $V_{i,k} = \Phi^{-1}\{F_{\mathbf{Y},k}(Y_{i,k})\}$ and $W_{i,l} = \Phi^{-1}\{F_{\mathbf{Z},l}(Z_{i,l})\}$. Let $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$ be three disjoint subsets of *[n]* with $|\mathcal{D}_1| = n_1 \asymp n$, $|\mathcal{D}_2| = n_2 \asymp n$ and $|\mathcal{D}_3| = n_3 \asymp n^\kappa$ for some constant $\kappa \in (0,1)$. Write $\mathcal{W}_{\mathcal{D}_j} = \{(\mathbf{X}_i, \mathbf{Y}_i, \mathbf{Z}_i) : i \in \mathcal{D}_j\}$. Our testing procedure includes three steps: Step 1 estimates $F_{\mathbf{X},j}(\cdot)$, $F_{\mathbf{Y},k}(\cdot)$ and $F_{\mathbf{Z},l}(\cdot)$ based on $\mathcal{W}_{\mathcal{D}_1}$, Step 2 estimates f_j and g_k based on $\mathcal{W}_{\mathcal{D}_2}$, and Step 3 calculates the test statistic and critical value based on $\mathcal{W}_{\mathcal{D}_3}$. See Section 4.1.1 for details. Section 4.1.2 will propose a data-driven procedure to select (n_1, n_2, n_3) in practice.

4.1.1 Testing Procedure

Given the subsamples $\mathcal{W}_{\mathcal{D}_1}$, the empirical distribution functions $\hat{F}_{\mathbf{X},j}(\cdot) = n_1^{-1} \sum_{s \in \mathcal{D}_1} I(X_{s,j} \le \cdot)$, $\hat{F}_{\mathbf{Y},k}(\cdot) = n_1^{-1} \sum_{s \in \mathcal{D}_1} I(Y_{s,k} \le \cdot)$ and $\hat{F}_{\mathbf{Z},l}(\cdot) = n_1^{-1} \sum_{s \in \mathcal{D}_1} I(Z_{s,l} \le \cdot)$ provide the natural estimates for $F_{\mathbf{X},j}(\cdot)$, $F_{\mathbf{Y},k}(\cdot)$ and $F_{\mathbf{Z},l}(\cdot)$. Since $\hat{F}_{\mathbf{X},j}(X_{i,j})$ may be equal to 0 or 1 for $i \in \mathcal{D}_2 \cup \mathcal{D}_3$, we consider its truncated version as follows:

$$\begin{aligned}\hat{F}^{(w)}_{\mathbf{X},j}(\cdot) \quad &= \frac{1}{n_1} I\{\hat{F}_{\mathbf{X},j}(\cdot) \le \frac{1}{n_1}\} + \hat{F}_{\mathbf{X},j}(\cdot) I\{\frac{1}{n_1} < \hat{F}_{\mathbf{X},j}(\cdot) \le \frac{n_1 - 1}{n_1}\} \\ &+ \frac{n_1 - 1}{n_1} I\{\hat{F}_{\mathbf{X},j}(\cdot) > \frac{n_1 - 1}{n_1}\}.\end{aligned} \quad (11)$$

Analogously, we can define $\hat{F}^{(w)}_{\mathbf{Y},k}(\cdot)$ and $\hat{F}^{(w)}_{\mathbf{Z},l}(\cdot)$ in the same manner. Then, for each $i \in [n]$, we can approximate $\mathbf{U}_i$, $\mathbf{V}_i$ and $\mathbf{W}_i$, respectively, by $\mathbf{U}_i^{(w)} = (\hat{U}^{(w)}_{i,1}, \ldots, \hat{U}^{(w)}_{i,p})^\top$, $\mathbf{V}_i^{(w)} = (\hat{V}^{(w)}_{i,1}, \ldots, \hat{V}^{(w)}_{i,q})^\top$ and $\mathbf{W}_i^{(w)} = (\hat{W}^{(w)}_{i,1}, \ldots, \hat{W}^{(w)}_{i,m})^\top$ with $\hat{U}^{(w)}_{i,j} = \Phi^{-1}\{\hat{F}^{(w)}_{\mathbf{X},j}(X_{i,j})\}$,

$\hat{V}_{i,k}^{(w)} = \Phi^{-1}\{\hat{F}_{\mathbf{Y},k}^{(w)}(Y_{i,k})\}$ and $\hat{W}_{i,l}^{(w)} = \Phi^{-1}\{\hat{F}_{\mathbf{Z},l}^{(w)}(Z_{i,l})\}$, which guarantee $|\hat{U}_{i,j}^{(w)}| < +\infty$, $|\hat{V}_{i,k}^{(w)}| < +\infty$ and $|\hat{W}_{i,l}^{(w)}| < +\infty$.

Given an integer $\ell \geq 0$, let $\mathcal{H}^{(\ell)}$ be the hierarchical neural networks proposed by Bauer and Kohler (2019). See (6.2) in Section 6.2 for its definition. Write

$$T_{\tilde{\beta}_n}\mathcal{H}^{(\ell)} = \{T_{\tilde{\beta}_n} h : h \in \mathcal{H}^{(\ell)}\},$$

where $(T_{\tilde{\beta}_n} h)(\boldsymbol{x}) = \{|h(\boldsymbol{x})| \wedge \tilde{\beta}_n\}\text{sign}\{h(\boldsymbol{x})\}$ with $\tilde{\beta}_n = (\log n)\log^{1/2}(\tilde{d}n)$ and $\tilde{d} = p \vee q \vee m$. Given $\{(\mathbf{U}_i^{(w)}, \mathbf{V}_i^{(w)}, \mathbf{W}_i^{(w)})\}_{i\in\mathcal{D}_2}$, we can estimate f_j and g_k as

$$\begin{aligned} \hat{f}_j(\cdot) &= \arg\min_{h\in T_{\tilde{\beta}_n}\mathcal{H}^{(\ell)}} \frac{1}{n_2}\sum_{i\in\mathcal{D}_2} |\hat{U}_{i,j}^{(w)} - h(\mathbf{W}_i^{(w)})|^2, \\ \hat{g}_k(\cdot) &= \arg\min_{h\in T_{\tilde{\beta}_n}\mathcal{H}^{(\ell)}} \frac{1}{n_2}\sum_{i\in\mathcal{D}_2} |\hat{V}_{i,k}^{(w)} - h(\mathbf{W}_i^{(w)})|^2. \end{aligned} \quad (12)$$

Given $\{(\mathbf{U}_i^{(w)}, \mathbf{V}_i^{(w)}, \mathbf{W}_i^{(w)})\}_{i\in\mathcal{D}_3}$, let $\Omega_n = n_3^{-1}\sum_{i\in\mathcal{D}_3}\boldsymbol{\eta}_i$ with $\boldsymbol{\eta}_i = \tilde{\boldsymbol{\varepsilon}}_i \otimes \boldsymbol{\delta}_i$, where $\tilde{\boldsymbol{\varepsilon}}_i = (\varepsilon_{i,1},\ldots,\varepsilon_{i,p})^\top$ and $\boldsymbol{\delta}_i = (\tilde{\delta}_{i,1},\ldots,\tilde{\delta}_{i,q})^\top$ with $\tilde{\varepsilon}_{i,j} = \hat{U}_{i,j}^{(w)} - \hat{f}_j(\mathbf{W}_i^{(w)})$ and $\tilde{\delta}_{i,k} = \hat{V}_{i,k}^{(w)} - \hat{g}_k(\mathbf{W}_i^{(w)})$ for $\hat{f}_j(\cdot)$ and $\hat{g}_k(\cdot)$ specified in (12). We consider the test statistic

$$\tilde{G}_n = \sqrt{n_3}\,|\Omega_n|_\infty$$

for (4), and reject $\mathbb{H}_0$ at the significance level $\alpha \in (0,1)$ if $\tilde{G}_n > \text{cv}_{\text{cind},\alpha}$, where $\text{cv}_{\text{cind},\alpha}$ is the critical value satisfying $\mathbb{P}(\tilde{G}_n > \text{cv}_{\text{cind},\alpha}) = \alpha$ under $\mathbb{H}_0$.

Let $\Theta = \text{Cov}(\boldsymbol{\eta}_i)$ for $\boldsymbol{\eta}_i = \boldsymbol{\varepsilon}_i \otimes \boldsymbol{\delta}_i$, which can be estimated by $\Theta = n_3^{-1}\sum_{i\in\mathcal{D}_3}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^\top - \bar{\boldsymbol{\eta}}\bar{\boldsymbol{\eta}}^\top$ with $\bar{\boldsymbol{\eta}} = n_3^{-1}\sum_{i\in\mathcal{D}_3}\boldsymbol{\eta}_i$. For any $\alpha \in (0,1)$, Proposition 2 in Appendix B of the supplementary material indicates that $\text{cv}_{\text{cind},\alpha}$ can be approximated by

$$\text{cv}_{\text{cind},\alpha} = \inf\{t \geq 0 : \mathbb{P}(|\boldsymbol{\zeta}|_\infty \leq t | \mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n) \geq 1-\alpha\} \quad (13)$$

for $\boldsymbol{\zeta} | \mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n \sim \mathcal{N}(\boldsymbol{0}, \Theta)$. Section 5 will introduce a multiplier bootstrap procedure to determine the critical value for the test, which is more computationally efficient. Our theoretical analysis in Section 6.2 shows that the proposed conditional independence test based on nonparametric regressions has three advantages: (a) no moment conditions on $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ are required, (b) it allows arbitrary dependence structures among the components of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, and (c) it allows the dimensions of $\mathbf{X}$ and $\mathbf{Y}$ to grow exponentially with the sample size *n*, while allowing the dimension of $\mathbf{Z}$ to grow polynomially with the sample size *n*.

Remark 3 .

Notice that the key requirement for the validity of our proposed method is

$$|\frac{1}{\sqrt{n_3}}\sum_{i\in\mathcal{D}_3}\tilde{\boldsymbol{\varepsilon}}_i\boldsymbol{\delta}_i^\top-\frac{1}{\sqrt{n_3}}\sum_{i\in\mathcal{D}_3}\boldsymbol{\varepsilon}_i\boldsymbol{\delta}_i^\top|_\infty=o_{\mathrm{p}}(1).\quad(14)$$

As shown in Section 6.2, the estimated functions $\hat{f}_j$ and $\hat{g}_k$ from the hierarchical neural networks $\mathcal{H}^{(\ell)}$ can satisfy (14) if f_j and g_k are (ϑ,C)-smooth functions (Bauer and Kohler, 2019). However, $\mathcal{H}^{(\ell)}$ is not necessary for our proposed method. More generally, any alternative function class can be used in place of $\mathcal{H}^{(\ell)}$, as long as the resulting estimates $\hat{f}_j$ and $\hat{g}_k$ satisfy (14).

4.1.2 A Data-driven Procedure for Selecting (n_1,n_2,n_3)

To implement the testing procedure for conditional independence proposed in Section 4.1.1, we need to determine (n_1,n_2,n_3) in practice. Our theory requires $n_1\asymp n$, $n_2\asymp n$ and $n_3\asymp n^\kappa$ for some constant $\kappa\in(0,1)$. Since the test statistic $\tilde{G}_n$ is constructed based on n_3 samples, the selection of n_3 will play a key role in the size control of the proposed test. On the other hand, due to $n_1,n_2\gg n_3$, the approximation errors caused by $(\mathbf{U}_i^{(w)},\mathbf{V}_i^{(w)},\mathbf{W}_i^{(w)})$ to $(\mathbf{U}_i,\mathbf{V}_i,\mathbf{W}_i)$ in Step 1 and $(\hat{f}_j,\hat{g}_k)$ to (f_j,g_k) in Step 2 will be negligible in constructing the theoretical properties of $\tilde{G}_n$. Hence, we mainly focus on the selection of n_3. In practice, we always set $\mathcal{W}_{\mathcal{D}_1}=\{(\mathbf{X}_i,\mathbf{Y}_i,\mathbf{Z}_i)\}_{i=1}^{n_1}$ and $\mathcal{W}_{\mathcal{D}_2}=\{(\mathbf{X}_i,\mathbf{Y}_i,\mathbf{Z}_i)\}_{i=n_1+1}^{n_1+n_2}$ with $n_1=\lfloor n/3\rfloor$ and $n_2=\lfloor n/2\rfloor$, and target on selecting some samples from $\{(\mathbf{X}_i,\mathbf{Y}_i,\mathbf{Z}_i)\}_{i=n_1+n_2+1}^{n}$ to form $\mathcal{W}_{\mathcal{D}_3}$. More specifically, given $\mathcal{W}_{\mathcal{D}_1}\cup\mathcal{W}_{\mathcal{D}_2}$, we can obtain the estimate $(\hat{f}_j,\hat{g}_k)$. Then, for each $i\in[n]$, we have $\tilde{\boldsymbol{\varepsilon}}_i=(\varepsilon_{i,1},\ldots,\varepsilon_{i,p})^\top$ and $\boldsymbol{\delta}_i=(\tilde{\delta}_{i,1},\ldots,\tilde{\delta}_{i,q})^\top$ with $\tilde{\varepsilon}_{i,j}=\hat{U}_{i,j}^{(w)}-\hat{f}_j(\mathbf{W}_i^{(w)})$ and $\tilde{\delta}_{i,k}=\hat{V}_{i,k}^{(w)}-\hat{g}_k(\mathbf{W}_i^{(w)})$. Based on the idea of bootstrap, we present in Algorithm 1 a data-driven procedure for selecting n_3 in practice.

Algorithm 1 Selection of optimal n_3

Input: (i) the number of repetitions *B*; and (ii) the significance level α, (iii) the estimated functions $\{\hat{f}_j\}_{j=1}^p$ and $\{\hat{g}_k\}_{k=1}^q$.

1: **for** $b \in [B]$ **do**

2: Generate $\{\varsigma_{1,i,k}^{(b)}\}_{i,k=1}^{n}$, $\{\varsigma_{2,i,k}^{(b)}\}_{i,k=1}^{n}$ and $\{\varsigma_{3,i,k}^{(b)}\}_{i,k=1}^{n}$ independently from $\mathcal{N}(0,1)$. Compute $\mathbf{W}_i^{(b)} = n^{-1/2}\sum_{k=1}^{n} \varsigma_{1,i,k}^{(b)} \mathbf{W}_k^{(w)}$, $\boldsymbol{\varepsilon}_i^{(b)} = n^{-1/2}\sum_{k=1}^{n} \varsigma_{2,i,k}^{(b)} \tilde{\boldsymbol{\varepsilon}}_k$ and $\boldsymbol{\delta}_i^{(b)} = n^{-1/2}\sum_{k=1}^{n} \varsigma_{3,i,k}^{(b)} \boldsymbol{\delta}_k$ for each $i \in [n]$. Write $\boldsymbol{\varepsilon}_i^{(b)} = (\varepsilon_{i,1}^{(b)}, \ldots, \varepsilon_{i,p}^{(b)})^\top$ and $\boldsymbol{\delta}_i^{(b)} = (\delta_{i,1}^{(b)}, \ldots, \delta_{i,q}^{(b)})^\top$.

3: Calculate $\mathbf{U}_i^{(b)} = (U_{i,1}^{(b)}, \ldots, U_{i,p}^{(b)})^\top$ and $\mathbf{V}_i^{(b)} = (V_{i,1}^{(b)}, \ldots, V_{i,q}^{(b)})^\top$ with $U_{i,j}^{(b)} = \hat{f}_j(\mathbf{W}_i^{(b)}) + \varepsilon_{i,j}^{(b)}$ and $V_{i,k}^{(b)} = \hat{g}_k(\mathbf{W}_i^{(b)}) + \delta_{i,k}^{(b)}$ for each $i \in [n]$.

4: Construct $\mathbf{U}_i^{(b)} = (\hat{U}_{i,1}^{(b)}, \ldots, \hat{U}_{i,p}^{(b)})^\top$ with $\hat{U}_{i,j}^{(b)} = U_{i,j}^{(b)} I(|U_{i,j}^{(b)}| \le M_1) + M_1 \cdot \mathrm{sign}(U_{i,j}^{(b)}) I(|U_{i,j}^{(b)}| > M_1)$ and $M_1 = \Phi^{-1}(1-n_1^{-1})$. Analogously, construct $\mathbf{V}_i^{(b)} = (\hat{V}_{i,1}^{(b)}, \ldots, \hat{V}_{i,q}^{(b)})^\top$ and $\mathbf{W}_i^{(b)} = (\hat{W}_{i,1}^{(b)}, \ldots, \hat{W}_{i,m}^{(b)})^\top$ in the same manner as $\mathbf{U}_i^{(b)}$ but with replacing $\mathbf{U}_i^{(b)}$ by $\mathbf{V}_i^{(b)}$ and $\mathbf{W}_i^{(b)}$, respectively.

5: For each $j \in [p]$ and $k \in [q]$, calculate $\hat{f}_j^{(b)}$ and $\hat{g}_k^{(b)}$ in the same manner as $\hat{f}_j$ and $\hat{g}_k$ specified in (12) but with replacing $\{(\mathbf{U}_i^{(w)}, \mathbf{V}_i^{(w)}, \mathbf{W}_i^{(w)})\}_{i \in \mathcal{D}_2}$ by $\{(\mathbf{U}_i^{(b)}, \mathbf{V}_i^{(b)}, \mathbf{W}_i^{(b)})\}_{i \in \mathcal{D}_2}$.

6: For each $i \in [n] \backslash [n_1 + n_2]$, calculate $\boldsymbol{\eta}_i^{(b)} = \tilde{\boldsymbol{\varepsilon}}_i^{(b)} \otimes \boldsymbol{\delta}_i^{(b)}$ with $\tilde{\boldsymbol{\varepsilon}}_i^{(b)} = (\varepsilon_{i,1}^{(b)}, \ldots, \varepsilon_{i,p}^{(b)})^\top$ and $\boldsymbol{\delta}_i^{(b)} = (\tilde{\delta}_{i,1}^{(b)}, \ldots, \tilde{\delta}_{i,q}^{(b)})^\top$, where $\varepsilon_{i,j}^{(b)} = \hat{U}_{i,j}^{(b)} - \hat{f}_j^{(b)}(\mathbf{W}_i^{(b)})$ and $\tilde{\delta}_{i,k}^{(b)} = \hat{V}_{i,k}^{(b)} - \hat{g}_k^{(b)}(\mathbf{W}_i^{(b)})$.

7: **for** $\tilde{\ell} \in [n - n_1 - n_2]$ **do**

8: Calculate the test statistic $\tilde{G}_{\tilde{\ell}}^{(b)} = \sqrt{\tilde{\ell}}\, |\Omega_{\tilde{\ell}}^{(b)}|_\infty$ with $\Omega_{\tilde{\ell}}^{(b)} = \tilde{\ell}^{-1} \sum_{i=n_1+n_2+1}^{n_1+n_2+\tilde{\ell}} \boldsymbol{\eta}_i^{(b)}$.

9: Calculate the critical value $\mathrm{cv}_{\mathrm{cind},\alpha}^{(b)}$ in the same manner as $\mathrm{cv}_{\mathrm{cind},\alpha}$ defined in (13) but with replacing $\{\boldsymbol{\eta}_i\}_{i \in \mathcal{D}_3}$ by $\{\boldsymbol{\eta}_i^{(b)}\}_{i=n_1+n_2+1}^{n_1+n_2+\tilde{\ell}}$.

10: Calculate $a_b(\tilde{\ell}) = I\{\tilde{G}_{\tilde{\ell}}^{(b)} > \mathrm{cv}_{\mathrm{cind},\alpha}^{(b)}\}$.

11: **end for**

12: **end for**

13: For each $\tilde{\ell} \in [n - n_1 - n_2]$, calculate $\bar{a}(\tilde{\ell}) = B^{-1} \sum_{b=1}^{B} a_b(\tilde{\ell})$.

Output: $n_3^{\mathrm{opt}} = \arg\min_{\tilde{\ell} \in [n-n_1-n_2]} |\bar{a}(\tilde{\ell}) - \alpha|$.

Remark 4 .

The proposed conditional independence test based on nonparametric regressions employs a single sample-splitting. As shown in Section R.4 in the supplementary material, the performance of the proposed method under null hypothesis is empirically insensitive to how to split all the samples into three disjoint parts as long as the size conditions of the three parts are satisfied. The sample-splitting is introduced primarily for theoretical convenience in establishing the asymptotic properties of the proposed method as $n \to \infty$. Algorithm 1 is intended to ensure that the proposed method can achieve good size control in finite samples. It requires training additional Bpq neural networks, which is computationally intensive. When n is small, sample-splitting may also lead to power loss. In this case, we can consider in practice using the full sample at each step of the testing procedure. Numerical studies in Section R.4 of the supplementary material show that our proposed test with full sample works well when n is small ($n \le 100$). When n is large, in order to improve the computational efficiency, we may consider setting $n_3 = n - n_1 - n_2$ directly in practice.

4.2 Conditional Independence Test based on Linear Regressions

Recall $\mathbf{U}_i = (\hat{U}_{i,1}, \ldots, \hat{U}_{i,p})^\top$, $\mathbf{V}_i = (\hat{V}_{i,1}, \ldots, \hat{V}_{i,q})^\top$ and $\mathbf{W}_i = (\hat{W}_{i,1}, \ldots, \hat{W}_{i,m})^\top$ with $\hat{U}_{i,j}$, $\hat{V}_{i,k}$ and $\hat{W}_{i,l}$ specified in (5). For $(\mathbf{A}, \mathbf{B})$ specified in (9), we write $\mathbf{A} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p)^\top$ and $\mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q)^\top$. We can estimate $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_k$ by the following Lasso estimators:

$$
\begin{aligned}
\boldsymbol{\alpha}_j &= \arg\min_{\boldsymbol{\alpha} \in \mathbb{R}^m} \{ \frac{1}{n} \sum_{i=1}^{n} (\hat{U}_{i,j} - \boldsymbol{\alpha}^\top \mathbf{W}_i)^2 + 2\lambda_{\boldsymbol{\alpha},j} |\boldsymbol{\alpha}|_1 \}, \\
\boldsymbol{\beta}_k &= \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^m} \{ \frac{1}{n} \sum_{i=1}^{n} (\hat{V}_{i,k} - \boldsymbol{\beta}^\top \mathbf{W}_i)^2 + 2\lambda_{\boldsymbol{\beta},k} |\boldsymbol{\beta}|_1 \},
\end{aligned} \tag{15}
$$

where $\lambda_{\boldsymbol{\alpha},j}$ and $\lambda_{\boldsymbol{\beta},k}$ are the regularization parameters. Let $\Omega_n = n^{-1} \sum_{i=1}^{n} \boldsymbol{\eta}_i$ with $\boldsymbol{\eta}_i = \hat{\boldsymbol{\varepsilon}}_i \otimes \boldsymbol{\delta}_i$, where $\hat{\boldsymbol{\varepsilon}}_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,p})^\top$ and $\boldsymbol{\delta}_i = (\hat{\delta}_{i,1}, \ldots, \hat{\delta}_{i,q})^\top$ with $\varepsilon_{i,j} = \hat{U}_{i,j} - \boldsymbol{\alpha}_j^\top \mathbf{W}_i$ and $\hat{\delta}_{i,k} = \hat{V}_{i,k} - \boldsymbol{\beta}_k^\top \mathbf{W}_i$. We consider the test statistic

$$
\hat{G}_n = \sqrt{n} \, |\Omega_n|_\infty
$$

for (4), and reject $\mathbb{H}_0$ at the significance level $\alpha \in (0,1)$ if $\hat{G}_n > \mathrm{cv}^*_{\mathrm{cind},\alpha}$, where $\mathrm{cv}^*_{\mathrm{cind},\alpha}$ is the critical value satisfying $\mathbb{P}(\hat{G}_n > \mathrm{cv}^*_{\mathrm{cind},\alpha}) = \alpha$ under $\mathbb{H}_0$.

Recall $\Theta = \mathrm{Cov}(\boldsymbol{\eta}_i)$, which can be estimated by $\Theta = n^{-1}\sum_{i=1}^{n}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\top} - \bar{\boldsymbol{\eta}}\bar{\boldsymbol{\eta}}^{\top}$ with $\bar{\boldsymbol{\eta}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{\eta}_i$. For any $\alpha \in (0,1)$, Proposition 3 in Appendix C of the supplementary material indicates that $\mathrm{cv}_{\mathrm{cind},\alpha}^{*}$ can be approximated by

$$\mathrm{cv}_{\mathrm{cind},\alpha}^{*} = \inf\{t \geq 0 : \mathbb{P}(|\boldsymbol{\zeta}|_{\infty} \leq t | \mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n) \geq 1-\alpha\} \quad (16)$$

for $\boldsymbol{\zeta} | \mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n \sim \mathcal{N}(\boldsymbol{0}, \Theta)$. Section 5 will introduce a multiplier bootstrap procedure to determine the critical value for the test, which is more computationally efficient. The proposed conditional independence test based on linear regressions shares similar advantages with its nonparametric counterpart discussed in Section 4.1.1. The main difference is that the current setting allows the dimensions of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ to grow exponentially with the sample size *n*. See our theoretical analysis in Section 6.3 for details.

5 Multiplier Bootstrap Procedure

To implement the proposed tests, we need to generate bootstrap samples of three *d*-dimensional normal random vectors $\boldsymbol{\xi} | \mathcal{X}_n, \mathcal{Y}_n \sim \mathcal{N}(\boldsymbol{0}, \Sigma)$, $\boldsymbol{\zeta} | \mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n \sim \mathcal{N}(\boldsymbol{0}, \Theta)$, and $\boldsymbol{\zeta} | \mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n \sim \mathcal{N}(\boldsymbol{0}, \Theta)$. Let $\varpi_1, \ldots, \varpi_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, which are independent of $(\mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n)$. Then

$$\boldsymbol{\xi}^{\dagger} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varpi_i(\hat{\boldsymbol{\gamma}}_i - \bar{\hat{\boldsymbol{\gamma}}}),\ \boldsymbol{\zeta}^{\dagger} = \frac{1}{\sqrt{n_3}}\sum_{i\in\mathcal{D}_3}\varpi_i(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}) \text{ and } \boldsymbol{\zeta}^{\dagger} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varpi_i(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}) \quad (17)$$

satisfy $\boldsymbol{\xi}^{\dagger} | \mathcal{X}_n, \mathcal{Y}_n \sim \mathcal{N}(\boldsymbol{0}, \Sigma)$, $\boldsymbol{\zeta}^{\dagger} | \mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n \sim \mathcal{N}(\boldsymbol{0}, \Theta)$, and $\boldsymbol{\zeta}^{\dagger} | \mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n \sim \mathcal{N}(\boldsymbol{0}, \Theta)$. For any $\alpha \in (0,1)$, the critical values defined in (6), (13) and (16) are equal to, respectively,

$$\begin{aligned}
\mathrm{cv}_{\mathrm{ind},\alpha} &= \inf\{t \geq 0 : \mathbb{P}(|\hat{\boldsymbol{\xi}}^{\dagger}|_{\infty} \leq t | \mathcal{X}_n, \mathcal{Y}_n) \geq 1-\alpha\}, \\
\mathrm{cv}_{\mathrm{cind},\alpha} &= \inf\{t \geq 0 : \mathbb{P}(|\tilde{\boldsymbol{\zeta}}^{\dagger}|_{\infty} \leq t | \mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n) \geq 1-\alpha\}, \quad (18) \\
\mathrm{cv}_{\mathrm{cind},\alpha}^{*} &= \inf\{t \geq 0 : \mathbb{P}(|\hat{\boldsymbol{\zeta}}^{\dagger}|_{\infty} \leq t | \mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n) \geq 1-\alpha\}.
\end{aligned}$$

Empirically, $\mathrm{cv}_{\mathrm{ind},\alpha}$ can be selected as the $\lfloor N\alpha \rfloor$-th largest value among $|\boldsymbol{\xi}_1^{\dagger}|_{\infty}, \ldots, |\boldsymbol{\xi}_N^{\dagger}|_{\infty}$, where *N* is a sufficiently large integer, and $\boldsymbol{\xi}_1^{\dagger}, \ldots, \boldsymbol{\xi}_N^{\dagger}$ are generated independently by (17). Analogously, $\mathrm{cv}_{\mathrm{cind},\alpha}$ and $\mathrm{cv}_{\mathrm{cind},\alpha}^{*}$ can be determined in the same manner.

Recall $d = pq$. When the dimension *d* is large and the sample size *n* is small, the Gaussian approximation specified above may lead to size distortions. See the numerical results in Section

7. To improve the finite sample performance, we may consider two other types of multipliers $\{\varpi_i\}_{i=1}^n$ in (17) advocated by Deng and Zhang (2020):

• Mammen's multiplier (Mammen, 1993): $\mathbb{P}\{\varpi_i = (1 \pm \sqrt{5})/2\} = (\sqrt{5} \mp 1)/(2\sqrt{5})$.

• Rademacher multiplier: $\mathbb{P}(\varpi_i = \pm 1) = 1/2$.

Theorem 7 in Section 6.4 shows that $\mathrm{cv}_{\mathrm{ind},\alpha}$, $\mathrm{cv}_{\mathrm{cind},\alpha}$ and $\mathrm{cv}^{*}_{\mathrm{cind},\alpha}$ defined in (18) with either Mammen's multiplier or Rademacher multiplier are also asymptotically valid critical values. Our extensive simulation studies in Section 7 indicate that the Rademacher multiplier provides more accurate approximations in finite samples. Hence we recommend using Rademacher multiplier ϖ_i in (17).

6 Theoretical Analysis

In this section, we provide the theoretical analysis for the proposed independence test and conditional independence tests.

6.1 Independence Test

Theorem 1 .

Let $p \lesssim n^{\varkappa_1}$ *and* $q \lesssim n^{\varkappa_2}$ *for any given constants* $\varkappa_1 \geq 0$ *and* $\varkappa_2 \geq 0$. *Under the null hypothesis* $\mathbb{H}_0$ *in (3), then*

$$\mathbb{P}(H_n > \mathrm{cv}_{\mathrm{ind},\alpha}) \to \alpha$$

as $n \to \infty$.

Theorem 1 shows that the size of the proposed independence test can be correctly controlled by the significance level $\alpha \in (0,1)$. Recall $d = pq$. Proposition 1 in Appendix A of the supplementary material indicates that Theorem 1 actually holds provided that $\log d \ll n^{\tilde{c}_1}$ for some constant $\tilde{c}_1 \in (0,1)$. Assuming $p \lesssim n^{\varkappa_1}$ and $q \lesssim n^{\varkappa_2}$ is just to simplify the presentation. Write $\Sigma = \mathrm{Cov}(\gamma_i) := (\Sigma_{i,j})_{d \times d}$. Theorem 2 shows that the proposed independence test is consistent under certain local alternatives imposed on the magnitude of $|\mathbb{E}(\mathbf{U}_i \mathbf{V}_i^{\top})|_{\infty}$.

Theorem 2 .

Let $p \lesssim n^{\varkappa_1}$ *and* $q \lesssim n^{\varkappa_2}$ *for any given constants* $\varkappa_1 \geq 0$ *and* $\varkappa_2 \geq 0$. *Under the alternative hypothesis* $\mathbb{H}_1$ *in (3), if* $\min_{j \in [d]} \Sigma_{j,j} \geq c_1$ *for some universal constant* $c_1 > 0$, *and*

$$|\mathbb{E}(\mathbf{U}_i\mathbf{V}_i^{\top})|_{\infty} \geq 4\sqrt{6}(1+\nu_n)n^{-1/2}(\log d)^{1/2}(\log n)/\sqrt{5}$$

with $\nu_n \geq c_2$, *where* $c_2 > 0$ *is an arbitrarily prescribed universal constant, then*

$$\mathbb{P}(H_n > \mathrm{cv}_{\mathrm{ind},\alpha}) \geq 1 - 2d^{-\nu_n/2-\nu_n^2/16} - o(1).$$

If either *p* or *q* diverges with the sample size *n*, as long as $|\mathbb{E}(\mathbf{U}_i\mathbf{V}_i^{\top})|_{\infty} \geq Cn^{-1/2}(\log d)^{1/2}\log n$ under the alternative hypothesis $\mathbb{H}_1$ in (3) for some universal constant $C > 4\sqrt{6/5}$, Theorem 2 implies that the proposed independence test is a consistent test in the sense that its power approaches 1. If *d* is a fixed constant, as long as $|\mathbb{E}(\mathbf{U}_i\mathbf{V}_i^{\top})|_{\infty} \gg n^{-1/2}\log n$ under the alternative hypothesis $\mathbb{H}_1$ in (3), the proposed independence test is also a consistent test. As shown in Section A.3 of the supplementary material, Theorem 2 actually holds provided that $\log d \ll n^{\tilde{c}_2}$ for some constant $\tilde{c}_2 \in (0,1)$. Together with Theorem 1, we know that, even if the dimensions of $\mathbf{X}$ and $\mathbf{Y}$ diverge exponentially with the sample size *n*, the proposed independence test can still correctly control the Type I error at the significance level $\alpha \in (0,1)$ and also have power approaching 1 under certain local alternatives.

6.2 Conditional Independence Test based on Nonparametric Regressions

To establish the theoretical guarantee of the proposed conditional independence test based on nonparametric regressions, we assume that the regression functions f_j and g_k in (10) satisfy the (ϑ, C)-smooth generalized hierarchical interaction model, which was introduced in Bauer and Kohler (2019). This function class covers a wide variety of models frequently used in nonparametric regression, such as additive models, single-index models, and interaction models, which is enough for capturing the nonlinear dependence between $(\mathbf{U},\mathbf{V})$ and $\mathbf{W}$ in practice. Bauer and Kohler (2019) establish the convergence rate of the regression estimates by using feedforward neural network under the (ϑ, C)-smooth generalized hierarchical interaction model assumption, which provides the foundation of our theoretical results. See Bauer and Kohler (2019) for more discussions. For the sake of completeness, we first introduce the definition of (ϑ, C)-smooth generalized hierarchical interaction model.

Definition 1 ((ϑ, C)-smooth function).

Let $\vartheta = \tilde{\vartheta} + s$ *for some* $\tilde{\vartheta} \in \mathbb{N}_0$ *and* $s \in (0,1]$. *A function* $h: \mathbb{R}^m \to \mathbb{R}$ *is called* (ϑ, C)*-smooth, if for every* $\boldsymbol{r} = (r_1, \ldots, r_m)^{\top} \in \mathbb{N}_0^m$ *with* $\sum_{j=1}^m r_j = \tilde{\vartheta}$, *the partial derivative* $(\partial^{\tilde{\vartheta}} h)/(\partial^{r_1} x_1 \cdots \partial^{r_m} x_m)$ *exists and satisfies*

$$\left|\frac{\partial^{\tilde{\vartheta}} h}{\partial^{r_1} x_1 \cdots \partial^{r_m} x_m}(\boldsymbol{x}) - \frac{\partial^{\tilde{\vartheta}} h}{\partial^{r_1} x_1 \cdots \partial^{r_m} x_m}(\boldsymbol{z})\right| \leq C|\boldsymbol{x} - \boldsymbol{z}|_2^s$$

for all $\boldsymbol{x} = (x_1, \ldots, x_m)^\top \in \mathbb{R}^m$ *and* $\boldsymbol{z} = (z_1, \ldots, z_m)^\top \in \mathbb{R}^m$.

Definition 2 ((ϑ, C) -smooth generalized hierarchical interaction model).

Let $m \in \mathbb{N}$, $m_* \in [m]$ *and* $f : \mathbb{R}^m \to \mathbb{R}$.

(i) *We say that f satisfies a generalized hierarchical interaction model of order* m_* *and level 0, if there exist* $h_1 : \mathbb{R}^{m_*} \to \mathbb{R}$ *and* $\phi_1, \ldots, \phi_{m_*} \in \mathbb{R}^m$ *such that* $f(\boldsymbol{x}) = h_1(\phi_1^\top \boldsymbol{x}, \ldots, \phi_{m_*}^\top \boldsymbol{x})$ *for all* $\boldsymbol{x} \in \mathbb{R}^m$.

(ii) *We say that f satisfies a generalized hierarchical interaction model of order* m_* *and level* $l+1$*, if there exist* $K \in \mathbb{N}$, $h_k : \mathbb{R}^{m_*} \to \mathbb{R}$ $(k \in [K])$ *and* $\tilde{h}_{1,k}, \ldots, \tilde{h}_{m_*,k} : \mathbb{R}^m \to \mathbb{R}$ $(k \in [K])$ *such that* $\tilde{h}_{1,k}, \ldots, \tilde{h}_{m_*,k}$ $(k \in [K])$ *satisfy a generalized hierarchical interaction model of order* m_* *and level* l*, and* $f(\boldsymbol{x}) = \sum_{k=1}^{K} h_k\{\tilde{h}_{1,k}(\boldsymbol{x}), \ldots, \tilde{h}_{m_*,k}(\boldsymbol{x})\}$ *for all* $\boldsymbol{x} \in \mathbb{R}^m$.

(iii) *We say that the generalized hierarchical interaction model defined above is* (ϑ, C) *-smooth, if all functions occurring in its definition are* (ϑ, C) *-smooth according to Definition* 1 .

Definition 3 .

Let $\mathcal{F}(m, m_*, l, K, \vartheta, L, C, \tilde{C})$ *be the set of functions* $f : \mathbb{R}^m \to \mathbb{R}$*, which satisfy the following conditions: f satisfies a* (ϑ, C)*-smooth generalized hierarchical interaction model of order* m_* *and level l as in Definition 2 with* $K \in \mathbb{N}$, $\vartheta = \tilde{\vartheta} + s$ *for some* $\tilde{\vartheta} \in \mathbb{N}_0$ *and* $s \in (0,1]$. *All partial derivatives of order less than or equal to* $\tilde{\vartheta}$ *of the functions* h_k, $\tilde{h}_{j,k}$ *given in Definition 2(ii) are bounded, that is, each such function h satisfies*

$$\max_{\substack{j_1, \ldots, j_m \in \{0\} \cup [\tilde{\vartheta}], \\ j_1 + \cdots + j_m \le \tilde{\vartheta}}} \left| \frac{\partial^{j_1 + \cdots + j_m} h}{\partial^{j_1} x_1 \cdots \partial^{j_m} x_m} \right|_\infty \le \tilde{C}$$

for some constant $\tilde{C} > 0$. *And let all functions* h_k *be Lipschitz continuous with Lipschitz constant* $L > 0$.

Bauer and Kohler (2019) recommend using the hierarchical neural networks to estimate the (ϑ, C)-smooth generalized hierarchical interaction model. Write $\boldsymbol{x} = (x_1, \ldots, x_m)^\top \in \mathbb{R}^m$. For $M_* \in \mathbb{N}$, $m_* \in [m]$ and $\tilde{\alpha}_n > 0$, we denote by $\mathcal{F}^{\mathrm{NN}}_{M_*, m_*, m, \tilde{\alpha}_n}$ the set of all functions $h : \mathbb{R}^m \to \mathbb{R}$ that satisfy

$$h(\boldsymbol{x})=\sum_{i=1}^{M_*}\mu_i\sigma\{\sum_{j=1}^{4m_*}\lambda_{i,j}\sigma(\sum_{v=1}^{m}\theta_{i,j,v}x_v+\theta_{i,j,0})+\lambda_{i,0}\}+\mu_0$$

for some $\mu_i,\lambda_{i,j},\theta_{i,j,v}\in\mathbb{R}$, where $\sigma(x)=(1+e^{-x})^{-1}$ for any $x\in\mathbb{R}$, $|\mu_i|\le\tilde{\alpha}_n$, $|\lambda_{i,j}|\le\tilde{\alpha}_n$ and $|\theta_{i,j,v}|\le\tilde{\alpha}_n$ for each $i\in[M_*]\cup\{0\}$, $j\in[4m_*]\cup\{0\}$ and $v\in[m]\cup\{0\}$. For $l=0$, the space of hierarchical neural networks is defined by $\mathcal{H}^{(0)}=\mathcal{F}^{\rm NN}_{M_*,m_*,m,\tilde{\alpha}_n}$. For $l\ge1$, define recursively

$$\begin{aligned}\mathcal{H}^{(l)}=\{f:\mathbb{R}^m\to\mathbb{R}: f(\boldsymbol{x})=\sum_{k=1}^{K}h_k\{\tilde{h}_{1,k}(\boldsymbol{x}),\ldots,\tilde{h}_{m_*,k}(\boldsymbol{x})\}\\ \text{for some } h_k\in\mathcal{F}^{\rm NN}_{M_*,m_*,m_*,\tilde{\alpha}_n} \text{ and } \tilde{h}_{j,k}\in\mathcal{H}^{(l-1)}\}\end{aligned} \quad (19)$$

with $K\in\mathbb{N}$. Then, we impose the following condition on the regression models (10).

Condition 1 .

For each $j\in[p]$ *and* $k\in[q]$, *the functions* $f_j,g_k\in\mathcal{F}(m,m_*,\ell,K,\vartheta,L,C,\tilde{C})$ *with finite positive integers* m_*, ℓ *and K, and some positive constants* ϑ, *L, C and* $\tilde{C}$.

Condition 1 is commonly assumed in the existing works of nonparametric regressions using deep neural networks, where they usually assume that the distribution of the predictor is supported on a bounded set. In our setting, although the predictor $\mathbf{W}_i$ has unbounded support, as shown in Equation (K.7) in the supplementary material, we have $\mathbf{W}_i^{(w)}\in[-\sqrt{2\log n_1},\sqrt{2\log n_1}]^m$ for sufficiently large n_1.

Recall $\Theta=\mathrm{Cov}(\boldsymbol{\eta}_i)$ with $\Theta=(\Theta_{i,j})_{d\times d}$. Write $\varrho=\vartheta+2m_*\tilde{\vartheta}+3m_*$, and $(\tilde{\alpha}_n,M_*)$ specified in (6.2) as

$$\tilde{\alpha}_n=n^{c_3} \text{ and } M_*=c_4\left\lceil n^{m_*/(4\vartheta+m_*)}(m^2\log n)^{m_*(2\tilde{\vartheta}+3)/(2\vartheta)}\right\rceil$$

for some sufficiently large constants $c_3>0$ and $c_4>0$. Recall $n_3\asymp n^{\kappa}$ for some constant $\kappa\in(0,1)$. Theorem 3 shows that the size of the proposed conditional independence test based on nonparametric regressions can be correctly controlled by the significance level $\alpha\in(0,1)$.

Theorem 3 .

Let Condition 1 hold with $p\lesssim n^{\varkappa_1}$, $q\lesssim n^{\varkappa_2}$ *and* $m\lesssim n^{\varkappa_3}$ *for any given constants*

$$\varkappa_1\ge0,\ \varkappa_2\ge0 \text{ and } 0\le\varkappa_3<\min\{\frac{\vartheta}{\varrho}(\frac{4\vartheta}{4\vartheta+m_*}-\kappa),\frac{1-\kappa}{2},\frac{\kappa}{4}\}. \quad (20)$$

Under the null hypothesis $\mathbb{H}_0$ *in (4), if* $\min_{j\in[d]}\Theta_{j,j}\ge c_5$ *for some universal constant* $c_5>0$, *then*

$$\mathbb{P}(\tilde{G}_n>\mathrm{cv}_{\mathrm{cind},\alpha})\to\alpha$$

as $n\to\infty$.

Recall $d=pq$. To obtain Theorem 3, Proposition 2 in Appendix B of the supplementary material indicates that *d* needs to satisfy $\log d\ll n^{\tilde{c}_3}$ for some constant $\tilde{c}_3\in(0,1)$. Assuming $p\lesssim n^{\varkappa_1}$ and $q\lesssim n^{\varkappa_2}$ is just to simplify the presentation. Write $\lambda(d,\alpha)=(2\log d)^{1/2}+\{2\log(1/\alpha)\}^{1/2}$. Theorem 4 shows that the proposed conditional independence test based on nonparametric regressions is consistent under certain local alternatives imposed on the magnitude of $|\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\delta}_i^\top)|_\infty$.

Theorem 4 .

Let $n_3\ge n^\kappa$ *and Condition 1 hold with* $p\lesssim n^{\varkappa_1}$, $q\lesssim n^{\varkappa_2}$ *and* $m\lesssim n^{\varkappa_3}$ *for any given constants* $\varkappa_1$, $\varkappa_2$ *and* $\varkappa_3$ *satisfying (20). Under the alternative hypothesis* $\mathbb{H}_1$ *in (4), if* $\min_{j\in[d]}\Theta_{j,j}\ge c_5$ *for some universal constant* $c_5>0$, *and*

$$|\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\delta}_i^\top)|_\infty\ge(1+\tilde{\epsilon}_n)n^{-\kappa/2}\lambda(d,\alpha)\max_{j\in[d]}\Theta_{j,j}^{1/2}$$

with $\tilde{\epsilon}_n>0$ *satisfying* $\tilde{\epsilon}_n^2\log d\to\infty$ *as* $n\to\infty$, *then*

$$\mathbb{P}(\tilde{G}_n>\mathrm{cv}_{\mathrm{cind},\alpha})\to 1$$

as $n\to\infty$.

As long as $|\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\delta}_i^\top)|_\infty\ge Cn^{-\kappa/2}(\log d)^{1/2}$ under the alternative hypothesis $\mathbb{H}_1$ in (4) for some universal constant $C>1$, Theorem 4 implies that the proposed conditional independence test based on nonparametric regressions is a consistent test in the sense that its power approaches 1. As shown in Section B.3 of the supplementary material, to obtain Theorem 4, *d* needs to satisfy $\log d\ll n^{\tilde{c}_4}$ for some constant $\tilde{c}_4\in(0,1)$. Together with Theorem 3, we know that, even if the dimensions of $\mathbf{X}$ and $\mathbf{Y}$ diverge exponentially with the sample size *n*, and the dimension of $\mathbf{Z}$ diverges polynomially with the sample size *n*, the proposed conditional independence test based on nonparametric regressions can still correctly control the Type I error at the significance level $\alpha\in(0,1)$ and also have power approaching 1 under certain local alternatives.

6.3 Conditional Independence Test based on Linear Regressions

Let $\Sigma_W = \mathrm{Cov}(\mathbf{W})$. To establish the theoretical guarantee of the proposed conditional independence test based on linear regressions, we impose the following condition on the regression models (9) and the regularization parameters $\lambda_{\boldsymbol{\alpha},j}$ and $\lambda_{\boldsymbol{\beta},k}$ involved in (15). Let $s = \max_{j\in[p]} |\boldsymbol{\alpha}_j|_0 \vee \max_{k\in[q]} |\boldsymbol{\beta}_k|_0$.

Condition 2 .

(i) *There exist universal constants* $c_6 > 0$ *and* $c_7 > 0$ *such that* $\mathbb{P}(|\boldsymbol{\alpha}_j^\top \mathbf{W}_i| > x) \le c_6 e^{-c_7 x^2}$ *and* $\mathbb{P}(|\boldsymbol{\beta}_k^\top \mathbf{W}_i| > x) \le c_6 e^{-c_7 x^2}$ *for any* $x > 0$, $i \in [n]$, $j \in [p]$ *and* $k \in [q]$. (ii) *The smallest eigenvalue of* Σ_W *is uniformly bounded away from zero.* (iii) *There exist two sufficiently large constants* $c_8 > 0$ *and* $c_9 > 0$ *such that* $c_8 n^{-1/2} \log^{1/2}(pm) \le \lambda_{\boldsymbol{\alpha},j} \le c_9 n^{-1/2} \log^{1/2}(pm)$ *and* $c_8 n^{-1/2} \log^{1/2}(qm) \le \lambda_{\boldsymbol{\beta},k} \le c_9 n^{-1/2} \log^{1/2}(qm)$ *for any* $j \in [p]$ *and* $k \in [q]$.

Write $\Theta = (\Theta_{i,j})_{d\times d}$. Theorem 5 shows that the size of the proposed conditional independence test based on linear regressions can be correctly controlled by the significance level $\alpha \in (0,1)$.

Theorem 5 .

Let Condition 2 hold with $p \lesssim n^{\varkappa_1}$, $q \lesssim n^{\varkappa_2}$ *and* $m \lesssim n^{\varkappa_3}$ *for any given constants* $\varkappa_1 \ge 0$, $\varkappa_2 \ge 0$ *and* $\varkappa_3 \ge 0$. *Under (9) and the null hypothesis* $\mathbb{H}_0$ *in (4), if* $s \ll n^{1/5}(\log n)^{-3}$ *and* $\min_{j\in[d]} \Theta_{j,j} \ge c_5$ *for some universal constant* $c_5 > 0$, *then*

$$\mathbb{P}(\hat{G}_n > \overset{*}{\mathrm{cv}}_{\mathrm{cind},\alpha}) \to \alpha$$

as $n \to \infty$.

Recall $\tilde{d} = p \vee q \vee m$. Proposition 3 in Appendix C of the supplementary material indicates that Theorem 5 actually holds provided that $\log \tilde{d} \ll n^{\tilde{c}_5}$ for some constant $\tilde{c}_5 \in (0,1)$. Assuming $p \lesssim n^{\varkappa_1}$, $q \lesssim n^{\varkappa_2}$ and $m \lesssim n^{\varkappa_3}$ is just to simplify the presentation. Theorem 6 shows that the proposed conditional independence test based on linear regressions is consistent under certain local alternatives imposed on the magnitude of $|\mathbb{E}(\boldsymbol{\varepsilon}_i \boldsymbol{\delta}_i^\top)|_\infty$.

Theorem 6 .

Let Condition 2 hold with $p \lesssim n^{\varkappa_1}$, $q \lesssim n^{\varkappa_2}$ *and* $m \lesssim n^{\varkappa_3}$ *for any given constants* $\varkappa_1 \ge 0$, $\varkappa_2 \ge 0$ *and* $\varkappa_3 \ge 0$. *Under (9) and the alternative hypothesis* $\mathbb{H}_1$ *in (4), if* $s \ll n^{1/5}(\log n)^{-1/2}$, $\min_{j\in[d]} \Theta_{j,j} \ge c_5$ *for some universal constant* $c_5 > 0$, *and*

$$|\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\delta}_i^\top)|_\infty \ge 12\sqrt{3\tilde{c}^{-1}}(\sqrt{2}+u_n)n^{-1/2}(\log\tilde{d})^{1/2}(\log n)/5$$

with $\tilde{c}=(1\wedge c_7)/4$ *and* $u_n \ge c_{10}$, *where* $c_{10}>0$ *is an arbitrarily prescribed universal constant, then*

$$\mathbb{P}(\hat{G}_n > \mathrm{cv}^*_{\mathrm{cind},\alpha}) \ge 1-2\tilde{d}^{-\sqrt{2}u_n/2-u_n^2/16}-o(1).$$

If either *p*, *q* or *m* diverges with the sample size *n*, as long as $|\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\delta}_i^\top)|_\infty \ge Cn^{-1/2}(\log\tilde{d})^{1/2}\log n$ under (9) and the alternative hypothesis $\mathbb{H}_1$ in (4) for some universal constant $C>12\sqrt{6}/(5\sqrt{\tilde{c}})$, Theorem 6 implies that the proposed conditional independence test based on linear regressions is a consistent test in the sense that its power approaches 1. If $\tilde{d}$ is a fixed constant, as long as $|\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\delta}_i^\top)|_\infty \gg n^{-1/2}\log n$ under the alternative hypothesis $\mathbb{H}_1$ in (4), the proposed conditional independence test based on linear regressions is also consistent. As shown in Section C.3 of the supplementary material, Theorem 6 actually holds provided that $\log\tilde{d} \ll n^{\tilde{c}_6}$ for some constant $\tilde{c}_6 \in (0,1)$. Together with Theorem 5, we know that, even if the dimensions of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ diverge exponentially with the sample size *n*, the proposed conditional independence test based on linear regressions can still correctly control the Type I error at the significance level $\alpha\in(0,1)$ and also have power approaching 1 under certain local alternatives.

6.4 Multiplier Bootstrap Procedure

Theorem 7 shows that the null-distributions of the test statistics H_n, $\tilde{G}_n$ and $\hat{G}_n$ can be approximated, respectively, by the distributions of $\boldsymbol{\xi}^\dagger$, $\boldsymbol{\zeta}^\dagger$ and $\boldsymbol{\zeta}^\dagger$ defined in (17) with either Mammen's multiplier or Rademacher multiplier.

Theorem 7 .

Let $\boldsymbol{\xi}^\dagger$, $\boldsymbol{\zeta}^\dagger$ *and* $\boldsymbol{\zeta}^\dagger$ *be defined in (17), with either Mammen's multiplier or Rademacher multiplier. Then the following three assertions hold.*

(i) *Let the conditions of Theorem 1 hold. Under the null hypothesis* $\mathbb{H}_0$ *in (3), then*

$$\sup_{z>0}|\mathbb{P}(H_n>z)-\mathbb{P}(|\boldsymbol{\xi}^\dagger|_\infty > z|\mathcal{X}_n,\mathcal{Y}_n)| = o_{\mathrm{p}}(1)$$

as $n\to\infty$.

(ii) *Let the conditions of Theorem 3 hold. Under the null hypothesis* $\mathbb{H}_0$ *in (4), then*

$$\sup_{z>0} |\mathbb{P}(\tilde{G}_n > z) - \mathbb{P}(|\zeta^{\dagger}|_\infty > z | \mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n)| = o_{\mathrm{p}}(1)$$

as $n \to \infty$.

(iii) *Let the conditions of Theorem 5 hold. Under (9) and the null hypothesis* $\mathbb{H}_0$ *in (4), if* $s \ll n^{1/6}(\log n)^{-13/6}$*, then*

$$\sup_{z>0} |\mathbb{P}(\hat{G}_n > z) - \mathbb{P}(|\zeta^{\dagger}|_\infty > z | \mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n)| = o_{\mathrm{p}}(1)$$

as $n \to \infty$.

7 Simulations

In this section, we conduct numerical studies to evaluate the finite-sample performance of the proposed independence test and conditional independence tests. To implement the proposed tests, we always use the multiplier bootstrap procedure introduced in Section 5 to calculate the associated critical values with $N = 5000$. We compare the performance of the three multipliers, i.e., Gaussian multiplier, Rademacher multiplier and Mammen's multiplier. All simulation results are based on 2000 replications and at the nominal significance level $\alpha = 0.05$.

7.1 Independence Test

In this subsection, we evaluate the performance of the proposed independence test via five simulated examples which characterize different types of dependence between the two random vectors $\mathbf{X} = (X_1, \ldots, X_p)^\top \in \mathbb{R}^p$ and $\mathbf{Y} = (Y_1, \ldots, Y_q)^\top \in \mathbb{R}^q$. We always set $p = q$ in Examples 1–5.

Example 1. Draw $X_1, \ldots, X_p, \tilde{Y}_1, \ldots, \tilde{Y}_q \overset{\text{i.i.d.}}{\sim} t(1)$. For $l \in [q]$, let $Y_l = \exp(X_l) I(l \in [K]) + \tilde{Y}_{l-K} I(l \in [q] \setminus [K])$. We set $K \in \{0, p/20, p/10\}$. When $K = 0$, $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$. Otherwise, $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}$.

Example 2. Let $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)^\top$ and $\boldsymbol{\varphi} = (\tilde{\varphi}_1, \ldots, \tilde{\varphi}_q)^\top$ with $\varphi_1, \ldots, \varphi_p, \tilde{\varphi}_1, \ldots, \tilde{\varphi}_q \overset{\text{i.i.d.}}{\sim} t(1)$. Generate $\tau \sim \mathcal{N}(0,1)$ independently of $\boldsymbol{\varphi}$ and $\boldsymbol{\varphi}$. For $j \in [p]$ and $l \in [q]$, let $X_j = 0.2\varphi_j + \tau I(j \in [K])$ and $Y_l = 0.2\tilde{\varphi}_l + \tau I(l \in [K])$. We set $K \in \{0, p/20, p/10\}$. When $K = 0$, $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$. Otherwise, $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}$.

Example 3. Draw $\tilde{X}_1,\ldots,\tilde{X}_p,\tilde{Y}_1,\ldots,\tilde{Y}_q,\tau_1,\ldots,\tau_K \overset{\text{i.i.d.}}{\sim} U(0,2\pi)$. For $j\in[p]$ and $l\in[q]$, let $X_j=\sin^2(\tau_j)I(j\in[K])+\tilde{X}_j I(j\in[p]\backslash[K])$ and $Y_l=\cos^2(\tau_l)I(l\in[K])+\tilde{Y}_l I(l\in[q]\backslash[K])$. We set $K\in\{0,p/20,p/10\}$. When $K=0$, $\mathbf{X}\perp\!\!\!\perp\mathbf{Y}$. Otherwise, $\mathbf{X}\not\!\perp\!\!\!\perp\mathbf{Y}$.

Example 4. Under the null hypothesis $\mathbb{H}_0$ in (1), generate $\boldsymbol{\varphi}=(\varphi_1,\ldots,\varphi_{p+q})^\top\sim\mathcal{N}(\mathbf{0},\mathbf{I}_{p+q})$. For $j\in[p]$ and $l\in[q]$, let $X_j=\varphi_j$ and $Y_l=\varphi_{p+l}$. Under the alternative hypothesis $\mathbb{H}_1$ in (1), generate $\boldsymbol{\varphi}\sim\mathcal{N}(\mathbf{0},\mathbf{R}^*)$, where $\mathbf{R}^*$ is generated as follows. Let

$$\Delta=\begin{pmatrix}\mathbf{0} & \Delta_{12}\\ \Delta_{12}^\top & \mathbf{0}\end{pmatrix}\in\mathbb{R}^{(p+q)\times(p+q)}$$

be a random matrix, where $\Delta_{12}\in\mathbb{R}^{p\times q}$ has only four nonzero entries. We set the locations of the four nonzero entries randomly in Δ_{12}, each with a magnitude randomly drawn from $U(0,1)$. To ensure positivity, let $\mathbf{R}^*=(1+\upsilon)\mathbf{I}_{p+q}+\Delta$ with $\upsilon=\{-\lambda_{\min}(\mathbf{I}_{p+q}+\Delta)+0.05\}I\{\lambda_{\min}(\mathbf{I}_{p+q}+\Delta)\le 0\}$. Then, for $j\in[p]$ and $l\in[q]$, let $X_j=\varphi_j$ and $Y_l=\varphi_{p+l}$.

Example 5. Write $\vartheta=(\vartheta_1,\ldots,\vartheta_{p+q})^\top$. For $j\in[p]$ and $l\in[q]$, let $X_j=\vartheta_j^{1/3}$ and $Y_l=\vartheta_{p+l}^{1/3}$. Under the null hypothesis $\mathbb{H}_0$ in (1), generate $\vartheta\sim\mathcal{N}(\mathbf{0},\mathbf{I}_{p+q})$. Under the alternative hypothesis $\mathbb{H}_1$ in (1), generate $\vartheta\sim\mathcal{N}(\mathbf{0},\mathbf{R}^*)$ with $\mathbf{R}^*$ specified in Example 4.

Example 1 is used in Zhu et al. (2017) for the monotone and nonlinear dependence between $\mathbf{X}$ and $\mathbf{Y}$. Example 2 is similar to the setting (V1) in the supplementary material of Deb and Sen (2023), which characterizes the monotone and linear dependence between $\mathbf{X}$ and $\mathbf{Y}$. Their setting only considers the case with $K=p$, while our Example 2 is more general that can cover the cases with $K\neq p$. In Examples 1 and 2, the distributions of $\mathbf{X}$ and $\mathbf{Y}$ are heavy-tailed. Example 3 is similar to Example A.4(iii) in the supplementary material of Zhu et al. (2020), which characterizes the nonlinear and non-monotone dependence between $\mathbf{X}$ and $\mathbf{Y}$. In comparison to Zhu et al. (2020) that only consider the case with $K=p$, our Example 3 is more general which can cover the cases with $K\neq p$. Examples 4 and 5 extend the simulation settings in Han et al. (2017), respectively, for data generated from the Gaussian distribution and the light-tailed Gaussian copula to the two-sample problem with Δ_{12} being the cross covariance matrix between $\mathbf{X}$ and $\mathbf{Y}$. These two examples can, respectively, characterize the linear and nonlinear dependence between $\mathbf{X}$ and $\mathbf{Y}$ under the sparse alternative.

We also compare the proposed independence test with eight other existing methods: (i) the test based on projection correlation (Pcor) in Zhu et al. (2017), (ii) the test based on ranks of distances (rdCov) in Heller et al. (2013), (iii) the test based on distance correlation (dCor) in Székely and Rizzo (2013), (iv) the *k*-variate HSIC based test (dHSIC) in Pfister et al. (2018), (v) the test based on the rank-based dependence matrix (JdCov_R) in Chakraborty and Zhang

(2019), (vi) the generalized distance covariance based test (GdCov) in Jin and Matteson (2018), (vii) the center-outward ranks and signs based test (Hallin) in Shi et al. (2022), and (viii) the multivariate rank-based test (mrdCov) in Deb and Sen (2023). All simulations are implemented in R. The R codes for implementing the Pcor test are provided by the authors of Zhu et al. (2017). The rdCov, dCor and dHSIC tests are implemented by calling the R-functions `hhg.test`, `dcorT.test` and `dhsic.test` in the `HHG`, `energy` and `dHSIC` packages, respectively. The JdCov_R test is implemented by using the R codes provided in the supplementary material of Chakraborty and Zhang (2019). The GdCov test is implemented by calling the R-function `mdm_test` in the R-package `EDMeasure`. The R codes of the Hallin and mrdCov tests are, respectively, available in the supplementary materials of Shi et al. (2022) and Deb and Sen (2023).

We set $p=q\in\{100,400,1600\}$ and $n\in\{50,100\}$ in the simulations. Table 1 reports the empirical sizes and powers of the proposed independence test and the competing methods. In Example 1 with $K\in\{p/20,p/10\}$, since the dCor, dHSIC, GdCov, Hallin and mrdCov tests return invalid results for more than 20% in the 2000 repetitions due to the heavy tails of the data, the associated results are reported by NA. Such a phenomenon indicates that these five tests may not work for the heavy-tailed data. For the proposed independence test, Rademacher multiplier has the best performance among the three choices of multipliers which can always control the sizes around the nominal significance level 0.05 and also has the highest powers. While Gaussian and Mammen's multipliers are under-sized in some scenarios, they still have quite good power performance in all the settings. For the competing methods, they can always control the sizes around the nominal level 0.05 in all the settings. However, the competing methods (except the JdCov_R test) have no powers in all the settings. The JdCov_R test only has good power performance in Example 2, but it still underperforms the proposed method.

7.2 Conditional Independence Test

In this subsection, we evaluate the performance of the proposed conditional independence tests based on nonparametric regressions (denoted by CI-FNN) and linear regressions (denoted by CI-Lasso), respectively, via five simulated examples which characterize different types of the conditional dependence between $\mathbf{X}=(X_1,\ldots,X_p)^\top\in\mathbb{R}^p$ and $\mathbf{Y}=(Y_1,\ldots,Y_q)^\top\in\mathbb{R}^q$ given $\mathbf{Z}=(Z_1,\ldots,Z_m)^\top\in\mathbb{R}^m$. We always set $p=q$ and $m\le p$ in Examples 6–10.

Example 6. Let $\mathcal{C}=\{(s_{i,1},s_{i,2}):1\le s_{i,1}<s_{i,2}\le m, i\in[\tilde{s}]\}$ with $\tilde{s}=\min\{m(m-1)/2,p,q\}$. Draw $Z_1,\ldots,Z_m,\tilde{X}_1,\ldots,\tilde{X}_{p-\tilde{s}},\tilde{Y}_1,\ldots,\tilde{Y}_{q-\tilde{s}}\overset{\text{i.i.d.}}{\sim}t(2)$. Generate $\tau_1,\ldots,\tau_K\overset{\text{i.i.d.}}{\sim}t(1)$ independently of $\{Z_l\}_{l=1}^m$, $\{\tilde{X}_j\}_{j=1}^{p-\tilde{s}}$ and $\{\tilde{Y}_k\}_{k=1}^{q-\tilde{s}}$. For $j\in[K]$, let $w_j=\tau_j+3\tau_j^3$. For $j\in[p]$ and $k\in[q]$, let $X_j=(Z_{s_{j,1}}Z_{s_{j,2}})I(j\in[\tilde{s}])+\tilde{X}_{j-\tilde{s}}I(j\in[p]\setminus[\tilde{s}])+w_jI(j\in[K])$ and $Y_k=(Z_{s_{k,1}}+Z_{s_{k,2}})I(k\in[\tilde{s}])+\tilde{Y}_{k-\tilde{s}}I(k\in[q]\setminus[\tilde{s}])+w_kI(k\in[K])$. We set $K\in\{0,p/10,p/5\}$. When $K=0$, $\mathbf{X}\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$. Otherwise, $\mathbf{X}\not\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$.

Example 7. Draw $Z_1,\ldots,Z_m \overset{\text{i.i.d.}}{\sim} U(-1,1)$ and $\tilde{X}_1,\ldots,\tilde{X}_{p-m},u_1,\ldots,u_q,\tau_1,\ldots,\tau_{q-m} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Let $\nu_1,\ldots,\nu_m$ be independent random variables that are computed as the sum of 48 i.i.d. random variables from $U(-0.25,0.25)$. Assume $\{Z_l\}_{l=1}^m$, $\{\tilde{X}_j\}_{j=1}^{p-m}$, $\{u_k\}_{k=1}^q$, $\{\tau_s\}_{s=1}^{q-m}$, and $\{\nu_l\}_{l=1}^m$ are mutually independent. For $j\in[p]$ and $k\in[q]$, let
$X_j=(Z_j+0.25Z_j^2+\nu_j)I(j\in[m])+\tilde{X}_{j-m}I(j\in[p]\backslash[m])$ and
$Y_k=(\beta X_k+Z_k+u_k)I(k\in[m])+(\tau_{k-m}+\beta X_k+u_k)I(k\in[q]\backslash[m])$ with $\beta=5\rho/(2\sqrt{1-\rho^2})$. We set $\rho\in\{0,0.7,0.8\}$. When $\rho=0$, $\mathbf{X}\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$. Otherwise, $\mathbf{X}\not\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$.

Example 8. Generate $Z_1,\ldots,Z_m,\tilde{X}_1,\ldots,\tilde{X}_{p-m},\tilde{Y}_1,\ldots,\tilde{Y}_{q-m},\nu_1,\ldots,\nu_m,u_1,\ldots,u_m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Draw $\tau_1,\ldots,\tau_K \overset{\text{i.i.d}}{\sim} t(1)$ independently of $\{Z_l\}_{l=1}^m$, $\{\tilde{X}_j\}_{j=1}^{p-m}$, $\{\tilde{Y}_k\}_{k=1}^{q-m}$, $\{\nu_l\}_{l=1}^m$ and $\{u_l\}_{l=1}^m$. For $j\in[p]$ and $k\in[q]$, let $X_j=\{\varphi_j+\varphi_j^3/3+\tanh(\varphi_j/3)/2\}I(j\in[m])+\tilde{X}_{j-m}I(j\in[p]\backslash[m])+3\tau_jI(j\in[K])$ and $Y_k=\{\tilde{\varphi}_k+\tanh(\tilde{\varphi}_k/3)\}^3I(k\in[m])+\tilde{Y}_{k-m}I(k\in[q]\backslash[m])+3\tau_kI(k\in[K])$ with
$\varphi_j=\{0.7(Z_j^3/5+Z_j/2)+\tanh(\nu_j)\}I(j\in[m])$ and $\tilde{\varphi}_k=\{(Z_k^3/4+Z_k)/3+u_k\}I(k\in[m])$. We set $K\in\{0,p/10,p/5\}$. When $K=0$, $\mathbf{X}\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$. Otherwise, $\mathbf{X}\not\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$.

Example 9. Generate $\{Z_l\}_{l=1}^m$, $\{\tilde{X}_j\}_{j=1}^{p-m}$, $\{\tilde{Y}_k\}_{k=1}^{q-m}$, $\{\nu_l\}_{l=1}^m$ and $\{u_l\}_{l=1}^m$ in the same manner as Example 8. Draw $\tau_1,\ldots,\tau_K \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$ independently of $\{Z_l\}_{l=1}^m$, $\{\tilde{X}_j\}_{j=1}^{p-m}$, $\{\tilde{Y}_k\}_{k=1}^{q-m}$, $\{\nu_l\}_{l=1}^m$ and $\{u_l\}_{l=1}^m$. For $j\in[p]$ and $k\in[q]$, let $\breve{X}_j=(\varphi_j+\varphi_j^3/3)I(j\in[m])+\tilde{X}_{j-m}I(j\in[p]\backslash[m])$ and
$\breve{Y}_k=\{\tilde{\varphi}_k+\tanh(\tilde{\varphi}_k/3)\}I(k\in[m])+\tilde{Y}_{k-m}I(k\in[q]\backslash[m])$ with
$\varphi_j=\{0.5(Z_j^3/7+Z_j/2)+\tanh(\nu_j)\}I(j\in[m])$ and $\tilde{\varphi}_k=\{(Z_k^3/2+Z_k)/3+u_k\}I(k\in[m])$. Then, let $X_j=\{0.5\breve{X}_j+3\cosh(\tau_j)\}I(j\in[K])+\breve{X}_jI(j\in[p]\backslash[K])$ and
$Y_k=\{0.5\breve{Y}_k+3\cosh(\tau_k^2)\}I(k\in[K])+\breve{Y}_kI(k\in[q]\backslash[K])$ for $j\in[p]$ and $k\in[q]$. We set $K\in\{0,p/10,p/5\}$. When $K=0$, $\mathbf{X}\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$. Otherwise, $\mathbf{X}\not\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$.

Example 10. Draw $Z_1,\ldots,Z_m,\tilde{X}_1,\ldots,\tilde{X}_{p-L},\tilde{Y}_1,\ldots,\tilde{Y}_{q-L},\nu_1,\ldots,\nu_L,u_1,\ldots,u_L$, $\tau_1,\ldots,\tau_K \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ with $L=p/4$ and $K\le L$. Let $\tilde{Z}=m^{-1}\sum_{i=1}^m Z_i$. For $j\in[p]$ and $k\in[q]$, let
$X_j=\tanh\{\tilde{Z}+\nu_j+3\tau_jI(j\in[K])\}I(j\in[L])+\tilde{X}_{j-L}I(j\in[p]\backslash[L])$ and
$Y_k=\{\tilde{Z}+u_k+3\tau_kI(k\in[K])\}^3I(k\in[L])+\tilde{Y}_{k-L}I(k\in[p]\backslash[L])$. We set $K\in\{0,p/10,p/5\}$. When $K=0$, $\mathbf{X}\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$. Otherwise, $\mathbf{X}\not\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$.

Example 6 is similar to Example 10 in Wang et al. (2015), where the latter only considers the fixed-dimensional scenario. Example 7 is similar to DGP1 in Su and White (2012), where the components of $\mathbf{X}$ and $\mathbf{Y}$ are generated by the polynomial regression models on $\mathbf{Z}$. Their setting only considers the case with $p = q = 1$, and our Example 7 is more general which allows $p, q \geq 1$. Given $\mathbf{Z}$, the random vectors $\mathbf{X}$ and $\mathbf{Y}$ are linearly conditional correlated in Examples 6 and 7. Example 8 is similar to the simulation setting provided in the Matlab codes of Zhang et al. (2011), which characterizes the linear conditional dependence between $\mathbf{X}$ and $\mathbf{Y}$ given $\mathbf{Z}$, under the nonlinear regression model settings of $\mathbf{X}$ and $\mathbf{Y}$ on $\mathbf{Z}$. Their setting only considers the case with $p = q = 1$, while our Example 8 can cover more general cases with $p, q \geq 1$. Example 9 extends Example 7 in Wang et al. (2015) that only considers the case with $p = q = 1$ to more general cases with $p, q \geq 1$, which characterizes the nonlinear conditional dependence between $\mathbf{X}$ and $\mathbf{Y}$ given $\mathbf{Z}$ under the nonlinear regression model settings of $\mathbf{X}$ and $\mathbf{Y}$ on $\mathbf{Z}$. Example 10 extends the simulation setting in Runge (2018) which only considers the case with $K = p = 1$ to more general cases with $K \neq p$ and $p \geq 1$.

We also compare the finite-sample performance of the proposed conditional independence tests with five other existing methods: (i) the test based on the generalized covariance measure (GCM) in Shah and Peters (2020), (ii) the test based on the projective approach (PCD) in Zhou et al. (2022), (iii) the randomized conditional independence test (RCIT) in Strobl et al. (2019), (iv) the randomized conditional correlation test (RCoT) in Strobl et al. (2019), and (v) the test based on conditional distance correlation (cdCov) in Wang et al. (2015). All simulations are implemented in R, except that the CI-FNN test is implemented in Python. In the CI-FNN test, $\hat{f}_j$ and $\hat{g}_k$ are estimated by (12) with the parameters $(\ell, K, m_*, M_*) = (0,1,1,32)$. We set $n_1 = \lfloor n/3 \rfloor$, $n_2 = \lfloor n/2 \rfloor$ and $n_3 = n_3^{\text{opt}}$, where n_3^{opt} is selected by Algorithm 1 with $B = 500$. In the CI-Lasso test, the Lasso estimators $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_k$ are obtained by calling the R-functions `glmnet` and `cv.glmnet` in the `glmnet` with the tuning parameters $\lambda_{\boldsymbol{\alpha},j}$ and $\lambda_{\boldsymbol{\beta},k}$ being chosen by the default 10-fold cross validation method. The GCM test is implemented by calling the R-function `gcm.test` in the `GeneralisedCovarianceMeasure` package. The codes of the PCD test are available in the supplementary material of Zhou et al. (2022). The RCIT and RCoT tests are implemented by calling the R-functions `RCIT` and `RCoT` in the `RCIT` package. The cdCov test is implemented by calling the R-function `cdcov.test` in the `cdcsis` package.

We set $p = q \in \{100, 400, 1600\}$, $m = 5$ and $n \in \{100, 200\}$ in the simulations. Table 2 reports the empirical sizes and powers of the proposed conditional independence tests and the competing methods. Since the PCD test would return Inf/NaN values for the test statistics due to the curse of dimensionality issue for the kernel-based methods, the associated results are reported as NA. When the sample size increases from $n = 100$ to $n = 200$, the proposed CI-FNN tests with three multipliers show significant improvements in both size control and power performance. This is consistent with the discussion in Section 4, where it is noted that fitting the feedforward neural network requires a substantial number of samples. Among the three choices of multipliers, same as the discussion in Section 7.1 for the proposed independence test, the proposed CI-FNN test with Rademacher multiplier still has the best performance in all the settings with well-controlled sizes and the highest powers. The CI-FNN tests with Gaussian and Mammen's multipliers are under-sized in most scenarios and exhibit reduced power when the sample size n is small $(n = 100)$. However, when n increases to 200, they still have quite good power performance in all the settings. On the other hand, as discussed in Section 4, when the joint distribution of $(\mathbf{U}_i, \mathbf{V}_i, \mathbf{W}_i)$ is close to normal, the CI-Lasso test can also be applied. It can be observed from Table 2 that the CI-Lasso test with Rademacher multiplier has higher powers in most cases than the CI-FNN test with Rademacher multiplier, particularly when $n = 100$. While the CI-Lasso tests with Gaussian and Mammen's multipliers are under-sized in most scenarios, their power performance is still quite good. Note that there are many more parameters to be estimated when fitting feedforward neural networks in the CI-FNN test than estimating Lasso estimators in the CI-Lasso test. For example, when $m = 5$, to estimate a f_j, fitting a feedforward neural network needs to estimate 4929 parameters, while fitting a linear regression model only needs to estimate 5 parameters. This might also reduce the power performance of the CI-FNN test when n is small. When n increases to 200, Table 2 shows that the power performance of the CI-FNN and CI-Lasso tests becomes comparably good.

For the competing methods, the RCIT, RCoT and cdCov tests fail to control the sizes around the nominal level in all the settings, since good approximation for the null distributions of the RCIT, RCoT and cdCov tests requires considerable sample size (Runge, 2018; Strobl et al., 2019; Wang et al., 2015). The GCM test has good size control in the simulation settings except Example 6. For Examples 6 and 8, the GCM test has no powers. The power performance of the GCM test in Example 9 is inferior to that of the CI-FNN and CI-Lasso tests with Rademacher multiplier. For Examples 7 and 10, the power performance of the GCM test is quite good and comparable to that of the CI-Lasso test.

Supplementary Material

Appendices A–P collect the technical proofs of all main theoretical results. Appendix Q presents the real-data analysis. Appendix R reports additional simulation studies on an extension of the proposed independence test; the effects of coordinatewise Gaussianization; comparisons with covariance-based methods under Gaussian data; and the effect of sample splitting in the proposed nonparametric conditional independence test.

Acknowledgments

We are grateful to the editor, the associate editor, and three referees for their insightful comments and suggestions, which have led to significant improvement of our article.

Disclosure Statement

The authors report there are no competing interest to declare.

Funding

J. Chang and Y. Du were supported in part by the National Natural Science Foundation of China (Grant nos. 72125008 and 72495122). J. He was supported in part by the National Natural Science Foundation of China (Grant nos. 72473114 and 72495121). Q. Yao was supported in part by the U.K. Engineering and Physical Sciences Research Council (Grant nos. EP/V007556/1 and EP/X002195/1).

References

Azadkia, M. and Chatterjee, S. (2021). A simple measure of conditional dependence. *Ann. Statist.*, 49, 3070–3102.

Bauer, B. and Kohler, M. (2019). On deep learning as a remedy for the curse of dimensionality in nonparametric regression. *Ann. Statist.*, 47, 2261–2285.

Bergsma, W. and Dassios, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli*, 20, 1006–1028.

Berrett, T. B., Wang, Y., Barber, R. F. and Samworth, R. J. (2020). The conditional permutation test for independence while controlling for confounders. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 82, 175–197.

Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, 32, 485–498.

Chakraborty, S. and Zhang, X. (2019). Distance metrics for measuring joint dependence with application to causal inference. *J. Amer. Statist. Assoc.*, 114, 1638–1650.

Chang, J., Chen, X. and Wu, M. (2024a). Central limit theorems for high dimensional dependent data. *Bernoulli*, 30, 712–742.

Chang, J., He, J., Kang, J. and Wu, M. (2024b). Statistical inferences for complex dependence of multimodal imaging data. *J. Amer. Statist. Assoc.*, 119, 1486–1499.

Chang, J., Zheng, C., Zhou, W. X. and Zhou, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics*, 73, 1300–1310.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.*, 45, 2309–2352.

Corradi, V., Distaso, W. and Fernandes, M. (2012). International market links and volatility transmission. *J. Econometrics*, 170, 117–141.

Deb, N. and Sen, B. (2023). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.*, 118, 192–207.

Deng, H. and Zhang, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.*, 48, 3643–3671.

Doran, G., Muandet, K., Zhang, K. and Schölkopf, B. (2014). A permutation-based kernel conditional independence test. *Proc. 30th Conf. Uncertainty Artif. Intell.*, 132–141.

Fan, J., Feng, Y. and Xia, L. (2020). A projection-based conditional dependence measure with applications to high-dimensional undirected graphical models. *J. Econometrics*, 218, 119–139.

Fukumizu, K., Gretton, A., Sun, X. and Schölkopf, B. (2008). Kernel measures of conditional dependence. *In NIPS*, 489–496.

Gao, L., Fan, Y., Lv, J. and Shao, Q.-M. (2021). Asymptotic distributions of high-dimensional distance correlation inference. *Ann. Statist.*, 49, 1999–2020.

Gieser, P. W. and Randles, R. H. (1997). A nonparametric test of independence between two vectors. *J. Amer. Statist. Assoc.*, 92, 561–567.

Goodman, L. A. (1959). Partial tests for partial taus. *Biometrika*, 46, 425–432.

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B. and Smola, A. J. (2008). A kernel statistical test of independence. *In NIPS*, 585–592.

Han, F., Chen, S. and Liu, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika*, 104, 813–828.

Heller, R., Heller, Y. and Gorfine, M. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika*, 100, 503–510.

Hettmansperger, T. P. and Oja, H. (1994). Affine invariant multivariate multisample sign tests. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 56, 235–249.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28, 321–377.

Huang, M., Sun, Y. and White, H. (2016). A flexible nonparametric test for conditional independence. *Econometric Theory*, 32, 1434–1482.

Lawrance, A. J. (1976). On conditional and partial correlation. *Amer. Statist.*, 30, 146–149.

Leung, D. and Drton, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.*, 46, 280–307.

Liu, H., Lafferty, J. and Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *J. Mach. Learn. Res.*, 10, 2295–2328.

Lyons, R. (2013). Distance covariance in metric spaces. *Ann. Probab.*, 41, 3284–3305.

Jakobsen, M. E. (2017). Distance covariance in metric spaces: Non-parametric independence testing in metric spaces. arXiv: 1706.03490.

Jin, Z. and Matteson, D. S. (2018). Generalizing distance covariance to measure and test multivariate mutual dependence via complete and incomplete V-statistics. *J. Multivariate Anal.*, 168, 304–322.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30, 81–93.

Kendall, M. G. (1942). Partial rank correlation. *Biometrika*, 32, 277–283.

Mai, Q., He, D. and Zou, H. (2023). Coordinatewise Gaussianization: Theories and applications. *J. Amer. Statist. Assoc.*, 118, 2329–2343.

Mai, Q. and Zou, H. (2015). Nonparametric variable transformation in sufficient dimension reduction. *Technometrics*, 57, 1–10.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.*, 21, 255–285.

Matteson, D. S. and Tsay, R. S. (2017). Independent component analysis via distance covariance. *J. Amer. Statist. Assoc.*, 112, 623–637.

Otneim, H. and Tjøstheim, D. (2022). The locally Gaussian partial correlation. *J. Bus. Econom. Statist.*, 40, 924–936.

Pearson, K. (1920). Notes on the history of correlation. *Biometrika*, 13, 25–45.

Pfister, N., Bühlmann, P., Schölkopf, B. and Peters, J. (2018). Kernel-based tests for joint independence. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 80, 5–31.

Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. New York: Wiley.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M. and Sabeti, P. C. (2011). Detecting novel associations in large data sets. *Science*, 334, 1518–1524.

Runge, J. (2018). Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information. *In AISTATS*, 938–947.

Shah, R. D. and Peters, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.*, 48, 1514–1538.

Shi, H., Drton, M. and Han, F. (2022). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.*, 117, 395–410.

Spearman, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol.*, 15, 72–101.

Strobl, E. V., Zhang, K. and Visweswaran, S. (2019). Approximate kernel-based conditional independence tests for fast non-parametric causal discovery. *J. Causal Inference*, 7, 20180017.

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

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

Su, L. and White, H. (2012). Conditional independence specification testing for dependent processes with local polynomial quantile regression. *Essays in Honor of Jerry Hausman*, 29, 355–434.

Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.*, 35, 2769–2794.

Székely, G. J. and Rizzo, M. L. (2013). The distance correlation t-test of independence in high dimension. *J. Multivariate Anal.*, 117, 193–213.

Székely, G. J. and Rizzo, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.*, 42, 2382–2412.

Taskinen, S., Kankainen, A. and Oja, H. (2003). Sign test of independence between two random vectors. *Statist. Probab. Lett.*, 62, 9–21.

Taskinen, S., Oja, H. and Randles, R. H. (2005). Multivariate nonparametric tests of independence. *J. Amer. Statist. Assoc.*, 100, 916–925.

Veraverbeke, N., Omelka, M. and Gijbels, I. (2011). Estimation of a conditional copula and association measures. *Scand. J. Stat.*, 38, 766–780.

Wang, X. and Hong, Y. (2018). Characteristic function based testing for conditional independence: A nonparametric regression approach. *Econometric Theory*, 34, 815–849.

Wang, X., Pan, W., Hu, W., Tian, Y. and Zhang, H. (2015). Conditional distance correlation. *J. Amer. Statist. Assoc.*, 110, 1726–1734.

Wilks, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica*, 3, 309–326.

Xue, L. and Zou, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *Ann. Statist.*, 40, 2541–2571.

Yang, S., Zheng, S. and Li, R. (2024). A new test for high-dimensional two-sample mean problems with consideration of correlation structure. *Ann. Statist.*, 52, 2217–2240.

Yao, S., Zhang, X. and Shao, X. (2018). Testing mutual independence in high dimension via distance covariance. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 80, 455–480.

Zhang, K., Peters, J., Janzing, D. and Schölkopf, B. (2011). Kernel-based conditional independence test and application in causal discovery. *Proc. 27th Conf. Uncertainty Artif. Intell.*, 804–813.

Zhu, C., Zhang, X., Yao, S. and Shao, X. (2020). Distance-based and RKHS-based dependence metrics in high dimension. *Ann. Statist.*, 48, 3366–3394.

Zhu, L., Xu, K., Li, R. and Zhong, W. (2017). Projection correlation between two random vectors. *Biometrika*, 104, 829–843.

Zhou, Y., Zhang, Y. and Zhu, L. (2022). A projective approach to conditional independence test for dependent processes. *J. Bus. Econom. Statist.*, 40, 398–407.

Table 1: Empirical sizes (the rows with $K=0$ in Examples 1–3 and 'null' in Examples 4–5) and powers (the rows with $K=p/20$ and $p/10$ in Examples 1–3 and 'alternative' in Examples 4–5) of the proposed independence test and the comparing methods in Examples 1–5. All numbers reported below are multiplied by 100. The results reported by 'NA' indicate that the associated tests return invalid results.

			$n=50$											$n=100$										
			Proposed Method											Proposed Method										
	p	Setting	Gaussian	Mammen	Rademacher	Pcor	rdCov	dCor	dHSIC	JdCov_R	GdCov	Hallin	mrdCov	Gaussian	Mammen	Rademacher	Pcor	rdCov	dCor	dHSIC	JdCov_R	GdCov	Hallin	mrdCov
Example 1	100	$K=0$	0.4	2.3	6.3	5.3	5.2	3.3	5.1	5.4	5.0	5.4	5.4	0.5	2.6	5.7	4.8	4.8	3.0	4.7	4.9	5.5	5.2	5.1
		$K=p/20$	100.0	100.0	100.0	14.3	6.1	NA	NA	9.5	NA	NA	NA	100.0	100.0	100.0	35.7	8.2	NA	NA	9.4	NA	NA	NA
		$K=p/10$	100.0	100.0	100.0	27.0	13.2	NA	NA	15.4	NA	NA	NA	100.0	100.0	100.0	60.0	23.2	NA	NA	22.0	NA	NA	NA

	400	$K=0$	0.0	1.1	6.1	5.8	5.3	4.4	5.4	5.6	5.8	5.7	4.0	0.0	1.3	5.3	5.8	5.1	4.0	5.4	4.9	6.3	4.8	5.1
		$K=p/20$	100.0	100.0	100.0	6.1	5.4	NA	NA	7.4	NA	NA	NA	100.0	100.0	100.0	7.4	8.2	NA	NA	7.6	NA	NA	NA
		$K=p/10$	100.0	100.0	100.0	6.2	9.0	NA	NA	11.4	NA	NA	NA	100.0	100.0	100.0	10.7	17.7	NA	NA	13.1	NA	NA	NA
	1600	$K=0$	0.0	0.0	6.2	5.3	5.1	4.8	5.5	6.1	5.5	4.4	5.1	0.0	1.0	6.3	5.0	5.1	4.2	4.8	5.5	5.2	5.0	5.8
		$K=p/20$	100.0	100.0	100.0	4.5	4.8	NA	NA	7.5	NA	NA	NA	100.0	100.0	100.0	4.9	6.6	NA	NA	9.2	NA	NA	NA
		$K=p/10$	100.0	100.0	100.0	3.4	7.4	NA	NA	11.8	NA	NA	NA	100.0	100.0	100.0	3.7	12.5	NA	NA	12.2	NA	NA	NA
Example 2	100	$K=0$	0.4	1.4	4.5	5.1	4.2	3.6	4.4	5.5	5.8	5.2	5.5	1.1	3.1	5.7	4.6	4.8	3.5	5.6	5.5	5.2	4.7	5.3
		$K=p/20$	94.9	99.4	100.0	5.9	5.9	3.9	5.5	30.6	5.4	5.2	5.0	100.0	100.0	100.0	6.2	5.0	3.7	5.1	58.0	5.2	5.6	5.1
		$K=p/10$	99.8	100.0	100.0	6.6	5.2	4.0	5.7	100.0	5.7	5.9	5.2	100.0	100.0	100.0	9.3	4.7	3.4	5.3	100.0	4.7	5.4	5.5
	400	$K=0$	0.0	1.1	4.4	4.7	4.3	4.2	4.5	5.6	5.1	4.9	5.1	0.2	1.4	5.4	5.7	5.7	3.9	5.9	4.3	5.9	4.6	4.8
		$K=p/20$	98.9	100.0	100.0	5.4	5.0	4.2	5.7	94.8	5.5	5.2	8.1	100.0	100.0	100.0	6.1	5.0	4.3	5.2	99.9	5.7	4.5	5.4
			100	100	100.	5.	4.	4.	4.	10	5.	5.	29.	100	100	100.	5.	4.	4.	5.	10	5.	5.	39.

		$K=p/10$	0	.0	0	1	9	4	7	0.0	8	5	9	.0	.0	0	8	7	1	0	0.0	2	5	5
	1600	$K=0$	0.0	0.4	5.9	4.9	5.5	3.6	5.2	5.3	5.8	4.7	5.1	0.0	0.9	5.0	4.4	4.4	3.3	5.0	6.0	4.9	5.2	5.1
		$K=p/20$	96.2	100.0	100.0	5.7	5.5	4.8	5.0	100.0	5.0	6.3	13.8	100.0	100.0	100.0	5.1	4.0	3.7	5.3	100.0	5.3	5.3	22.6
		$K=p/10$	99.8	100.0	100.0	5.8	4.6	4.3	5.6	100.0	5.2	4.9	40.2	100.0	100.0	100.0	5.6	4.4	3.9	6.0	100.0	5.8	4.8	66.0
Example 3	100	$K=0$	0.1	1.7	5.7	6.1	5.3	6.1	5.8	5.7	5.6	5.2	5.3	1.1	2.3	5.9	4.8	5.7	4.8	4.7	4.8	5.6	5.1	5.2
		$K=p/20$	100.0	100.0	100.0	4.7	5.8	4.7	4.4	7.8	5.3	5.6	5.3	100.0	100.0	100.0	5.1	4.3	5.4	5.3	8.4	5.2	5.6	4.8
		$K=p/10$	100.0	100.0	100.0	5.4	4.4	5.5	5.3	11.2	5.0	4.7	6.1	100.0	100.0	100.0	5.6	5.1	5.3	5.6	22.6	5.1	4.7	5.8
	400	$K=0$	0.1	0.8	5.4	5.5	5.4	5.5	5.5	6.7	5.3	4.9	5.2	0.3	1.5	5.9	4.8	4.8	4.9	4.5	5.4	5.9	5.5	4.7
		$K=p/20$	100.0	100.0	100.0	4.9	5.2	4.9	4.7	6.8	4.9	4.4	4.7	100.0	100.0	100.0	5.5	4.6	5.4	5.2	8.6	5.1	4.8	5.9
		$K=p/10$	100.0	100.0	100.0	5.0	4.6	4.8	4.8	12.2	4.8	5.4	5.5	100.0	100.0	100.0	5.4	4.4	4.9	4.9	11.4	5.4	5.0	6.2
	1600	$K=0$	0.0	0.3	6.1	5.1	5.0	5.2	5.1	6.2	5.3	4.9	5.0	0.0	1.2	6.2	6.1	5.1	5.9	5.7	5.1	4.9	5.0	5.4
		$K=p/20$	100.0	100.0	100.0	5.2	5.3	4.8	4.7	6.0	6.2	6.0	4.8	100.0	100.0	100.0	4.5	4.7	4.8	4.8	9.8	4.8	5.3	5.3

		$K = p/10$	100.0	100.0	100.0	5.4	4.5	5.1	5.2	9.6	5.0	5.1	4.6	100.0	100.0	100.0	5.8	5.1	5.7	5.9	12.5	5.2	5.1	5.2
Example 4	100	null	0.1	1.3	7.3	4.9	5.0	4.9	5.0	5.4	6.0	5.7	4.9	0.4	1.6	5.6	5.2	4.9	4.9	4.9	4.4	5.5	5.9	5.2
		alternative	76.6	84.2	89.4	12.8	5.8	12.5	12.9	5.1	7.2	4.7	5.8	95.2	96.4	97.1	26.5	6.3	26.8	25.6	5.6	8.9	5.8	6.0
	400	null	0.0	0.4	7.5	3.9	4.9	3.6	3.7	5.8	5.3	4.8	6.6	0.0	1.7	5.5	5.1	4.9	4.9	4.7	5.1	5.5	4.7	6.0
		alternative	62.8	75.4	83.4	7.0	4.5	7.0	6.8	6.1	5.6	4.5	5.2	91.1	93.0	94.7	7.4	4.4	7.5	7.3	4.5	5.2	5.7	4.6
	1600	null	0.0	0.0	7.0	5.3	5.3	5.3	5.5	5.1	5.1	5.1	5.0	0.0	0.9	6.4	4.5	3.3	4.5	4.9	5.6	5.3	5.0	5.4
		alternative	42.7	65.6	78.8	4.6	4.3	4.5	4.2	4.2	5.4	5.4	5.1	88.2	91.1	92.4	6.0	4.3	5.8	5.9	4.6	5.7	4.8	4.4
Example 5	100	null	0.0	1.2	6.8	5.0	4.5	5.0	5.0	5.3	5.6	4.9	5.7	0.6	1.9	5.9	5.2	4.9	4.8	4.5	6.0	5.1	5.5	5.5
		alternative	77.5	83.9	88.7	11.7	5.5	11.5	11.3	5.6	6.4	6.0	4.8	95.1	96.0	96.9	22.0	7.0	21.5	21.3	6.4	9.2	5.6	5.7
	400	null	0.0	0.5	7.0	6.1	4.8	6.0	5.9	4.7	5.0	5.6	6.7	0.1	1.5	5.9	5.6	5.3	5.1	5.2	5.4	6.0	5.3	5.7
		alternative	62.5	74.5	83.3	6.8	5.3	6.7	6.8	5.6	5.9	5.1	5.3	90.8	93.3	94.7	8.0	5.9	7.8	7.5	6.0	5.0	5.2	4.7
	1600	null	0.0	0.2	6.8	5.6	5.4	5.2	5.0	4.8	5.4	4.7	5.5	0.0	0.6	6.2	5.9	5.3	6.0	6.0	5.6	5.1	5.5	4.3
		alternativ	44.	65.	77.9	5.	5.	5.	5.	6.0	5.	5.	4.8	87.	90.	92.6	5.	4.	5.	5.	4.6	5.	5.	5.2

		e	5	9		7	3	5	7		0	3		7	9		0	9	0	0		8	5	

Table 2: Empirical sizes (the rows with $K=0$ in Examples 6 and 8–10, and $\rho=0$ in Example 7) and powers (the rows with $K=p/10$ and $p/5$ in Examples 6 and 8–10, $\rho=0.7$ and 0.8 in Example 7) of the proposed conditional independence tests and the comparing methods in Examples 6–10. All numbers reported below are multiplied by 100. The results reported by ‘NA’ indicate that the associated tests return invalid results.

	p	Setting	$n=100$: Proposed Methods, Gaussian, CI-FNN	Gaussian, CI-Lasso	Mammen, CI-FNN	Mammen, CI-Lasso	Rademacher, CI-FNN	Rademacher, CI-Lasso	GCM	PCD	RCIT	RCoT	cdCov	$n=200$: Proposed Methods, Gaussian, CI-FNN	Gaussian, CI-Lasso	Mammen, CI-FNN	Mammen, CI-Lasso	Rademacher, CI-FNN	Rademacher, CI-Lasso	GCM	PCD	RCIT	RCoT	cdCov
Example 6	100	$K=0$	0.3	0.8	3.2	2.5	7.3	7.1	0.2	1.1	99.6	99.6	35.1	0.9	2.3	2.1	4.1	7.0	6.0	0.2	0.4	19.2	23.4	59.0
		$K=p/10$	25.2	100.0	21.7	100.0	57.9	100.0	0.2	NA	99.6	99.2	18.9	90.2	100.0	90.5	100.0	92.6	100.0	0.1	NA	22.4	21.1	40.6
		$K=p/5$	45.5	100.0	34.6	100.0	79.9	100.0	0.2	NA	99.4	99.4	16.2	96.7	100.0	94.5	100.0	97.6	100.0	0.0	NA	27.2	27.1	39.5
	400	$K=0$	0.0	0.1	2.2	2.0	5.4	6.6	0.1	5.4	99.4	99.1	42.9	0.0	0.6	1.2	2.1	7.6	6.3	0.0	3.2	24.1	20.8	66.3
		$K=p/10$	23.4	100.0	27.6	100.0	70.8	100.0	0.2	NA	99.6	99.8	17.1	100.0	100.0	100.0	99.7	100.0	100.0	0.1	NA	24.3	22.1	40.6
		$K=p/5$	69.6	100.0	95.6	100.0	99.7	100.0	0.3	NA	99.5	99.5	18.7	100.0	100.0	100.0	100.0	100.0	100.0	0.0	NA	24.9	21.9	35.5
	1600	$K=0$	0.0	0.0	1.2	0.1	9.0	8.9	0.0	NA	99.4	99.2	45.4	0.5	0.3	1.5	1.9	8.5	6.7	0.0	3.2	25.0	25.0	73.8

			$n=100$											$n=200$										
		$K=p/10$	42.5	100.0	82.0	100.0	97.0	100.0	0.2	NA	99.2	99.2	15.6	100.0	100.0	100.0	100.0	100.0	100.0	0.0	NA	21.2	20.2	40.4
		$K=p/5$	70.6	100.0	97.2	100.0	99.4	100.0	0.0	NA	98.8	98.6	19.6	100.0	100.0	100.0	100.0	100.0	100.0	0.2	NA	25.0	26.2	42.8
Example 7	100	$\rho=0$	0.3	1.1	1.9	3.4	7.2	7.0	5.4	3.7	100.0	100.0	99.9	0.1	1.4	1.0	3.2	5.1	6.2	4.0	1.8	55.9	54.6	100.0
		$\rho=0.7$	73.4	100.0	47.9	100.0	97.9	100.0	100.0	0.0	100.0	100.0	100.0	99.0	100.0	99.2	100.0	99.9	100.0	100.0	0.0	81.3	83.7	100.0
		$\rho=0.8$	87.0	100.0	60.4	100.0	99.3	100.0	100.0	0.1	100.0	100.0	99.9	99.3	100.0	100.0	100.0	100.0	100.0	100.0	0.0	83.1	85.6	100.0
	400	$\rho=0$	0.2	0.0	2.2	1.5	6.8	6.5	4.9	6.9	99.9	100.0	100.0	0.2	0.3	1.2	1.5	7.0	4.9	4.4	4.7	55.9	55.6	100.0
		$\rho=0.7$	56.6	100.0	47.0	100.0	97.6	100.0	100.0	54.6	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	1.8	65.0	66.8	100.0
		$\rho=0.8$	99.0	100.0	100.0	100.0	100.0	100.0	100.0	98.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	20.5	63.7	66.1	100.0
	1600	$\rho=0$	0.0	0.0	1.2	0.8	8.4	7.6	6.4	8.2	100.0	100.0	100.0	0.3	0.4	1.5	1.9	4.7	6.2	3.8	7.4	55.6	57.4	100.0
		$\rho=0.7$	74.5	100.0	100.0	100.0	100.0	100.0	100.0	NA	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.2	58.0	57.2	100.0
		$\rho=0.8$	96.6	100.0	100.0	100.0	100.0	100.0	100.0	NA	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	NA	59.2	58.2	100.0

			$n=100$											$n=200$										
Example 8	100	$K=0$	0.0	0.5	2.7	2.2	7.8	5.5	4.6	0.5	99.9	99.9	95.7	0.4	1.6	1.7	3.1	5.0	5.4	4.2	0.3	38.4	40.9	99.8
		$K=p/10$	19.9	100.0	17.2	100.0	52.2	100.0	3.7	5.5	100.0	100.0	67.5	84.0	100.0	83.6	100.0	89.0	100.0	4.1	0.3	39.1	42.0	89.2
		$K=p/5$	45.1	100.0	31.3	100.0	80.1	100.0	4.3	15.4	99.9	99.9	67.8	95.5	100.0	95.6	100.0	97.2	100.0	7.1	7.2	43.8	45.9	89.5
	400	$K=0$	0.0	0.0	2.0	1.0	7.5	5.1	4.7	1.6	99.9	100.0	99.3	0.2	1.0	1.4	2.3	6.4	5.3	4.4	1.7	37.3	40.3	100.0
		$K=p/10$	28.4	100.0	27.4	100.0	66.4	100.0	5.0	NA	99.9	100.0	68.0	100.0	100.0	100.0	100.0	100.0	100.0	3.3	NA	41.2	39.9	89.9
		$K=p/5$	76.0	100.0	97.2	100.0	99.5	100.0	6.8	NA	100.0	100.0	69.6	100.0	100.0	100.0	100.0	100.0	100.0	6.0	NA	42.3	42.1	89.6
	1600	$K=0$	0.2	0.0	0.8	0.4	6.4	6.4	8.6	3.0	100.0	100.0	99.6	0.5	0.1	1.5	1.4	7.4	5.7	4.6	2.8	39.0	42.8	100.0
		$K=p/10$	47.5	100.0	84.5	100.0	98.5	100.0	6.2	NA	100.0	100.0	71.6	100.0	100.0	100.0	100.0	100.0	100.0	3.4	NA	40.8	40.0	89.2
		$K=p/5$	76.4	100.0	98.2	100.0	100.0	100.0	7.6	NA	100.0	100.0	73.0	100.0	100.0	100.0	100.0	100.0	100.0	4.6	NA	35.6	42.2	86.9
Example 9	100	$K=0$	0.1	0.6	2.9	2.3	7.1	5.3	3.8	4.0	99.9	100.0	98.7	0.7	1.7	1.7	2.9	5.1	6.2	4.0	3.6	40.4	37.2	100.0
		$K=p/10$	12.3	100.0	12.8	100.0	44.7	100.0	9.0	0.0	99.9	100.0	37.6	85.9	100.0	85.6	100.0	89.6	100.0	5.3	0.0	41.4	42.3	53.1

			$n=100$											$n=200$										
		$K=p/5$	23.1	100.0	20.3	100.0	65.3	100.0	18.4	0.0	100.0	100.0	38.5	94.5	100.0	92.1	100.0	95.6	100.0	9.7	0.0	42.1	40.8	49.3
	400	$K=0$	0.0	0.0	1.8	1.0	5.6	5.5	5.4	4.8	100.0	100.0	99.8	0.2	0.9	1.0	2.1	6.4	5.3	2.3	0.4	39.8	37.1	100.0
		$K=p/10$	19.2	100.0	49.6	100.0	74.8	100.0	18.8	0.1	100.0	100.0	37.5	99.3	100.0	100.0	100.0	100.0	100.0	10.1	0.0	39.6	39.4	47.5
		$K=p/5$	37.1	100.0	77.0	100.0	94.5	100.0	48.1	0.0	100.0	100.0	34.3	100.0	100.0	100.0	100.0	100.0	100.0	20.4	0.0	40.0	39.1	48.6
	1600	$K=0$	0.2	0.0	1.0	0.3	7.0	6.6	6.2	4.8	100.0	100.0	100.0	0.0	0.1	1.3	1.6	4.8	6.0	4.2	3.8	38.0	41.0	100.0
		$K=p/10$	14.0	100.0	51.2	100.0	82.0	100.0	48.4	0.0	100.0	99.8	31.2	99.6	100.0	100.0	100.0	100.0	100.0	20.2	0.2	39.6	38.0	42.4
		$K=p/5$	26.4	100.0	74.4	100.0	95.6	100.0	85.0	0.0	99.8	100.0	30.4	100.0	100.0	100.0	100.0	100.0	100.0	55.4	0.0	41.6	40.0	39.0
Example 10	100	$K=0$	0.3	0.7	2.8	2.6	7.8	5.2	5.4	3.9	100.0	99.8	97.7	0.2	1.7	2.1	3.1	5.9	5.3	4.0	4.0	39.1	41.6	100.0
		$K=p/10$	15.5	100.0	15.4	100.0	48.0	100.0	100.0	4.9	100.0	99.9	93.7	83.9	100.0	85.0	100.0	88.9	100.0	100.0	4.3	40.3	39.0	99.5
		$K=p/5$	24.6	100.0	26.8	100.0	67.3	100.0	100.0	4.8	99.9	100.0	97.7	93.3	100.0	92.2	100.0	96.2	100.0	100.0	5.4	41.1	43.4	99.7
	400	$K=0$	0.2	0.1	1.0	1.3	6.8	6.1	7.3	3.2	100.0	100.0	99.7	0.0	1.2	1.4	2.8	6.2	6.7	3.9	2.6	38.6	42.0	100.0

			$n=100$											$n=200$										
		$K=p/10$	28.3	100.0	58.5	100.0	78.7	100.0	100.0	3.8	100.0	100.0	99.5	99.7	100.0	100.0	100.0	100.0	100.0	100.0	4.0	41.9	40.1	100.0
		$K=p/5$	44.3	100.0	81.4	100.0	95.4	100.0	100.0	5.4	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	3.9	40.0	39.6	100.0
		$K=0$	0.0	0.0	2.0	6.0	8.0	6.8	9.8	1.8	100.0	100.0	100.0	0.0	0.1	0.8	1.6	6.0	6.0	4.2	1.0	41.2	39.4	100.0
	1600	$K=p/10$	16.5	100.0	55.5	100.0	84.5	100.0	100.0	2.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	2.4	37.6	40.0	100.0
		$K=p/5$	28.8	100.0	79.4	100.0	96.6	100.0	100.0	4.2	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	5.2	40.4	40.6	100.0