ARTICLE IN PRESS

Journal of Multivariate Analysis xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Independence tests in the presence of measurement errors: An invariance law

Jinlin Fan [a], Yaowu Zhang [b], Liping Zhu [a,*]

[a] Center for Applied Statistics, Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China
[b] Research Institute for Interdisciplinary Sciences, School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China

ARTICLE INFO

Article history:
Received 14 August 2021
Received in revised form 20 August 2021
Accepted 20 August 2021
Available online xxxx

AMS 2020 subject classifications:
primary 62H20
secondary 62E20

Keywords:
Distance correlation
Independence test
Measurement error
Nonlinear dependence

ABSTRACT

In many scientific areas the observations are collected with measurement errors. We are interested in measuring and testing independence between random vectors which are subject to measurement errors. We modify the weight functions in the classic distance covariance such that, the modified distance covariance between the random vectors of primary interest is the same as its classic version between the surrogate random vectors, which is referred to as the invariance law in the present context. The presence of measurement errors may substantially weaken the degree of nonlinear dependence. An immediate issue arises: The classic distance correlation between the surrogate vectors cannot reach one even if the two random vectors of primary interest are exactly linearly dependent. To address this issue, we propose to estimate the distance variance using repeated measurements. We study the asymptotic properties of the modified distance correlation thoroughly. In addition, we demonstrate its finite-sample performance through extensive simulations and a real-world application.

1. Introduction

Measuring nonlinear dependence and testing statistical independence are fundamental problems in statistics. Let $\mathbf{x} = (X_1, \ldots, X_p)^\top \in \mathbb{R}^p$ and $\mathbf{y} = (Y_1, \ldots, Y_q)^\top \in \mathbb{R}^q$ be two random vectors of primary interest. For example, $\mathbf{x}$ stands for the systolic and diastolic pressures, and $\mathbf{y}$ stands for the degree of exposures to air pollutants. It is of scientific importance to investigate whether the blood pressures are relevant to the exposures to pollutants. However, in such scientific studies, both $\mathbf{x}$ and $\mathbf{y}$ are measured with non-ignorable random errors. Instead of observing $\mathbf{x}$ and $\mathbf{y}$ directly, we observe the surrogate random vectors, $\mathbf{u} \in \mathbb{R}^p$ and $\mathbf{v} \in \mathbb{R}^q$, which are related to $\mathbf{x}$ and $\mathbf{y}$, respectively, through

$$\mathbf{u} = \boldsymbol{\alpha}_{\mathbf{x}} + \mathbf{B_x}^\top \mathbf{x} + \boldsymbol{\varepsilon}_{\mathbf{x}}, \quad \mathbf{v} = \boldsymbol{\alpha}_{\mathbf{y}} + \mathbf{B_y}^\top \mathbf{y} + \boldsymbol{\varepsilon}_{\mathbf{y}}, \tag{1}$$

where $\boldsymbol{\alpha}_{\mathbf{x}} \in \mathbb{R}^p$, $\boldsymbol{\alpha}_{\mathbf{y}} \in \mathbb{R}^q$ are non-random vectors, $\mathbf{B_x} \in \mathbb{R}^{p\times p}$, $\mathbf{B_y} \in \mathbb{R}^{q\times q}$ are full-rank matrices, and $\boldsymbol{\varepsilon}_{\mathbf{x}} \in \mathbb{R}^p$, $\boldsymbol{\varepsilon}_{\mathbf{y}} \in \mathbb{R}^q$ are unobservable random errors. In some real-world applications, $\mathbf{B_x}$ and $\mathbf{B_y}$ can be identity matrices. We assume $\boldsymbol{\varepsilon}_{\mathbf{x}}$, $\boldsymbol{\varepsilon}_{\mathbf{y}}$, and $(\mathbf{x}, \mathbf{y})$ are mutually independent, and allow $\mathbf{x}$ and $\mathbf{y}$ to be dependent. The measurement error model (1) is widely assumed in literature. See, for example, [7] and [12].

Extensive studies have been conducted to investigate how $\mathbf{y}$ dependents on $\mathbf{x}$ in the regression context, where the covariates are measured with linear errors. Ignoring measurement errors will induce non-ignorable bias in parameter

* Corresponding author.
E-mail address: zhu.liping@ruc.edu.cn (L. Zhu).

https://doi.org/10.1016/j.jmva.2021.104818
0047-259X/© 2021 Published by Elsevier Inc.

Please cite this article as: J. Fan, Y. Zhang and L. Zhu, Independence tests in the presence of measurement errors: An invariance law, Journal of Multivariate Analysis (2021) 104818, https://doi.org/10.1016/j.jmva.2021.104818.

estimation, which would lead misleading results in practice. [3] proposed an efficient estimation for linear measurement error model in which $\mathbf{y}$ depends on $\mathbf{x}$ linearly. [23] and [32] considered partially linear model and generalized linear model, respectively. Interested readers may refer to [7] for a comprehensive review. Recent advances include [1,20,24,26] and [34]. In these recent studies, the covariate vectors are allowed to be high dimensional.

The measurement error models have also been considered in the context of sufficient dimension reduction [11], which falls into the framework of semi-parametric regression. [6] observed that, the central subspace of $\mathbf{y}$ given the unobservable covariate vector $\mathbf{x}$, recovered through ordinary least squares and sliced inverse regression [19], is the same as that of $\mathbf{y}$ given a particular linear transformation of $\mathbf{u}$. In their context this surprising phenomenon is referred to as the invariance property, which was later found to be generally applicable to many other sufficient dimension reduction methods [21]. Similar observation has also been made by [10] in the context of the score test.

Measuring nonlinear dependence between $\mathbf{x}$ and $\mathbf{y}$ in the absence of measurement errors has been extensively studied in literature. See, for example, [2,4,8,14,16,30,33,35] and [36]. Interested readers may refer to [31] for a comprehensive review. However, measuring nonlinear dependence between $\mathbf{x}$ and $\mathbf{y}$ in the presence of measurement errors is rarely touched in the literature. To the best of our knowledge, [9] is perhaps the first attempt to measure nonlinear dependence of $\mathbf{x}$ and $\mathbf{y}$ through the distance correlation between $\mathbf{u}$ and $\mathbf{v}$ [33]. In their context, both $\mathbf{x}$ and $\mathbf{y}$ are restricted to be univariate, and the random errors $\boldsymbol{\varepsilon}_{\mathbf{x}}$ and $\boldsymbol{\varepsilon}_{\mathbf{y}}$ are assumed to be normal with zero means and known variances. These requirements indeed limit the usefulness of [9]'s proposal in real-world problems. In addition, the metric proposed by [9] cannot attain one even when $\mathbf{x}$ and $\mathbf{y}$ are perfectly linearly dependent. This is because the variabilities of $\mathbf{u}$ and $\mathbf{v}$ are always larger than those of $\mathbf{x}$ and $\mathbf{y}$ when $\mathbf{B}_{\mathbf{x}} = \mathbf{B}_{\mathbf{y}} = 1$ and $p = q = 1$ in Model (1), due to the presence of measurement errors.

In this paper, we propose to test statistical independence and measure nonlinear dependence between $\mathbf{x}$ and $\mathbf{y}$, the random vectors of primary interest, through distance correlation between the surrogate random vectors $\mathbf{u}$ and $\mathbf{v}$. We allow for non-normal measurement errors and multivariate random vectors. The distance correlation consists of two components: distance covariance and distance variance. Throughout we use $\mathrm{dcov}(\mathbf{u}, \mathbf{v})$ to stand for the classic distance covariance between $\mathbf{u}$ and $\mathbf{v}$, and $\mathrm{dcov}(\mathbf{v}, \mathbf{v})$ to stand for the distance variance of $\mathbf{v}$ [33]. To test statistical independence, it suffices to use distance covariance. In particular, to test statistical independence between $\mathbf{x}$ and $\mathbf{y}$, we suggest to modify the weight functions in the classic distance covariance such that the modified distance covariance between $\mathbf{x}$ and $\mathbf{y}$ is exactly the same as $\mathrm{dcov}(\mathbf{u}, \mathbf{v})$. We refer to this property as an invariance law. To measure the degree of nonlinear dependence, it is required to quantify the variabilities of $\mathbf{x}$ and $\mathbf{y}$, which are usually smaller than those of $\mathbf{u}$ and $\mathbf{v}$. An immediate issue arises: The classic distance correlation between $\mathbf{u}$ and $\mathbf{v}$ cannot attain one even when $\mathbf{x}$ and $\mathbf{y}$ are exactly linearly dependent. In other words, the presence of measurement errors may substantially weaken the degree of nonlinear dependence. We use a toy example to demonstrate this phenomenon. Suppose for now that in Model (1), both $\mathbf{x}$ and $\mathbf{y}$ are univariate and $\mathbf{B}_{\mathbf{x}} = \mathbf{B}_{\mathbf{y}} = 1$. In addition, $\mathbf{x}$ and $\mathbf{y}$ follow bivariate standard normal distribution with correlation coefficient r. The squared distance correlation between $\mathbf{x}$ and $\mathbf{y}$, averaged over 1000 replications and denoted as $\mathrm{dcorr}^2(\mathbf{x}, \mathbf{y})$, is displayed in Fig. 1. We also report $\mathrm{dcorr}^2(\mathbf{u}, \mathbf{v})$ with dashed line in the same figure for the purposes of comparison. It can be clearly seen that, when the correlation coefficient between $\mathbf{x}$ and $\mathbf{y}$, denoted as r, approaches 1 or -1, $\mathrm{dcorr}^2(\mathbf{x}, \mathbf{y})$ is very near to 1. However, $\mathrm{dcorr}^2(\mathbf{u}, \mathbf{v})$ is smaller than 0.2 throughout. This indicates that, directly using $\mathrm{dcorr}^2(\mathbf{u}, \mathbf{v})$ to quantify the degree of nonlinear dependence between $\mathbf{x}$ and $\mathbf{y}$ is substantially biased. To address this issue, we suggest to modify the distance variances slightly, by assuming the repeated measurements, $\widetilde{\mathbf{u}}$ and $\widetilde{\mathbf{v}}$, which are indeed the respective independent copies of $\mathbf{u}$ and $\mathbf{v}$ when $\mathbf{x}$ and $\mathbf{y}$ are fixed, are available. We suggest to use $\mathrm{dcov}(\mathbf{u}, \widetilde{\mathbf{u}})$ and $\mathrm{dcov}(\mathbf{v}, \widetilde{\mathbf{v}})$, to replace $\mathrm{dcov}(\mathbf{u}, \mathbf{u})$ and $\mathrm{dcov}(\mathbf{v}, \mathbf{v})$ in distance correlation, which leads to a new metric, denoted as $\rho^2(\mathbf{x}, \mathbf{y})$ in Fig. 1. It can be clearly seen that, the modified squared distance correlation, $\rho^2(\mathbf{x}, \mathbf{y})$, is indeed very close to $\mathrm{dcorr}^2(\mathbf{x}, \mathbf{y})$, across all r values.

This paper is organized as follows. In Section 2, we introduce an invariance law to test independence in the presence of measurement errors. At the sample level, we propose to estimate the modified distance correlation with U-statistic theory and the distance variances with repeated measurements. The asymptotic properties of these estimates are also studied. In Section 3 we illustrate the finite-sample performance of our proposal through extensive simulations and an application. We conclude this paper with brief discussions in Section 4.

2. A modified distance correlation in the presence of measurement errors

2.1. The population level

Let $\varphi(\cdot)$ be a characteristic function. Throughout our goal is to test H_0: the two random vectors of primary interest, $\mathbf{x}$ and $\mathbf{y}$, are independent. Under H_0, $\varphi_{\mathbf{x},\mathbf{y}}(\mathbf{t}, \mathbf{s}) = \varphi_{\mathbf{x}}(\mathbf{t})\varphi_{\mathbf{y}}(\mathbf{s})$ for all $\mathbf{t} \in \mathbb{R}^p$ and $\mathbf{s} \in \mathbb{R}^q$, or equivalently,

$$\iint \|\varphi_{\mathbf{x},\mathbf{y}}(\mathbf{t}, \mathbf{s}) - \varphi_{\mathbf{x}}(\mathbf{t})\varphi_{\mathbf{y}}(\mathbf{s})\|^2 \omega_1(\mathbf{t})\omega_2(\mathbf{s}) d\mathbf{t} d\mathbf{s}, \tag{2}$$

where $\omega_1(\cdot)$ and $\omega_2(\cdot)$ are positive weight functions such that the integration (2) exists. However, in the present context $\mathbf{x}$ and $\mathbf{y}$ are not observable directly. Instead, we merely observe $\mathbf{u}$ and $\mathbf{v}$, which admit the measurement error model (1).

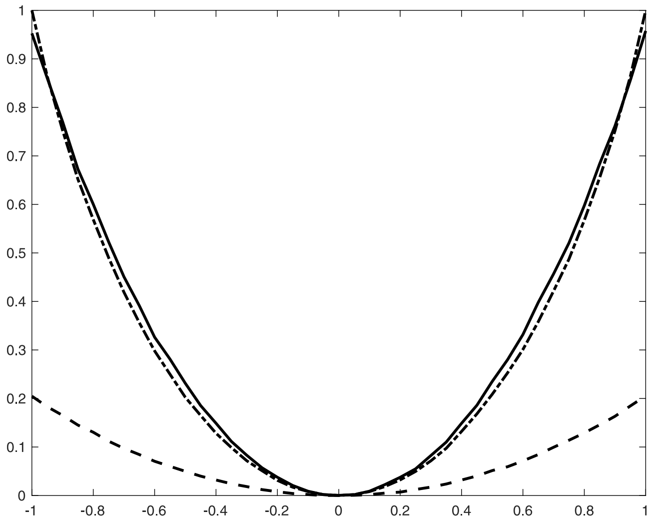

Fig. 1. The vertical axis: the averages over 1000 replications for our proposed metric $\rho^2(\mathbf{x},\mathbf{y})$ (solid line), the distance correlations $\mathrm{dcorr}^2(\mathbf{u},\mathbf{v})$ (dashed line) and $\mathrm{dcorr}^2(\mathbf{x},\mathbf{y})$ (dash-dotted line). The horizontal axis: the Pearson correlation coefficient r between $\mathbf{x}$ and $\mathbf{y}$.

Therefore, we have to work with $\mathbf{u}$ and $\mathbf{v}$ to test independence between $\mathbf{x}$ and $\mathbf{y}$. By Model (1), it follows immediately that

$$\varphi_{\mathbf{u}}(\mathbf{B_x}^{-1}\mathbf{t}) = \exp(i\boldsymbol{\alpha}_{\mathbf{x}}{}^{\top}\mathbf{B_x}^{-1}\mathbf{t})\varphi_{\mathbf{x}}(\mathbf{t})\varphi_{\boldsymbol{\varepsilon}_{\mathbf{x}}}(\mathbf{B_x}^{-1}\mathbf{t}),\quad \varphi_{\mathbf{v}}(\mathbf{B_y}^{-1}\mathbf{s}) = \exp(i\boldsymbol{\alpha}_{\mathbf{y}}{}^{\top}\mathbf{B_y}^{-1}\mathbf{s})\varphi_{\mathbf{y}}(\mathbf{s})\varphi_{\boldsymbol{\varepsilon}_{\mathbf{y}}}(\mathbf{B_y}^{-1}\mathbf{s}),$$
$$\varphi_{\mathbf{u},\mathbf{v}}(\mathbf{B_x}^{-1}\mathbf{t},\mathbf{B_y}^{-1}\mathbf{s}) = \exp(i\boldsymbol{\alpha}_{\mathbf{x}}{}^{\top}\mathbf{B_x}^{-1}\mathbf{t}+i\boldsymbol{\alpha}_{\mathbf{y}}{}^{\top}\mathbf{B_y}^{-1}\mathbf{s})\varphi_{\mathbf{x},\mathbf{y}}(\mathbf{t},\mathbf{s})\varphi_{\boldsymbol{\varepsilon}_{\mathbf{x}}}(\mathbf{B_x}^{-1}\mathbf{t})\varphi_{\boldsymbol{\varepsilon}_{\mathbf{y}}}(\mathbf{B_y}^{-1}\mathbf{s}).$$

Therefore, (2) can be expanded as follows,

$$\iint \|\varphi_{\mathbf{u},\mathbf{v}}(\mathbf{B_x}^{-1}\mathbf{t},\mathbf{B_y}^{-1}\mathbf{s}) - \varphi_{\mathbf{u}}(\mathbf{B_x}^{-1}\mathbf{t})\varphi_{\mathbf{v}}(\mathbf{B_y}^{-1}\mathbf{s})\|^2\|\varphi_{\boldsymbol{\varepsilon}_{\mathbf{x}}}^{-1}(\mathbf{B_x}^{-1}\mathbf{t})\|^2\|\varphi_{\boldsymbol{\varepsilon}_{\mathbf{y}}}^{-1}(\mathbf{B_y}^{-1}\mathbf{s})\|^2\omega_1(\mathbf{t})\omega_2(\mathbf{s})d\mathbf{t}d\mathbf{s}$$
$$= \iint \|\varphi_{\mathbf{u},\mathbf{v}}(\mathbf{t},\mathbf{s}) - \varphi_{\mathbf{u}}(\mathbf{t})\varphi_{\mathbf{v}}(\mathbf{s})\|^2\cdot\left\{\|\varphi_{\boldsymbol{\varepsilon}_{\mathbf{x}}}^{-1}(\mathbf{t})\|^2\omega_1(\mathbf{B_x}\mathbf{t})\ |\mathbf{B_x}|\ \right\}\cdot\left\{\|\varphi_{\boldsymbol{\varepsilon}_{\mathbf{y}}}^{-1}(\mathbf{s})\|^2\omega_2(\mathbf{B_y}\mathbf{s})\ |\mathbf{B_y}|\ \right\}d\mathbf{t}d\mathbf{s}.$$

Throughout we use the same notation $\|\cdot\|$ to stand for the modulus if the argument is a complex number, the norm if the argument is a vector. We use $|\cdot|$ to stand for the determinant if the argument is a matrix. We define $\widetilde{\omega}_1(\mathbf{t}) = \left\{\|\varphi_{\boldsymbol{\varepsilon}_{\mathbf{x}}}^{-1}(\mathbf{t})\|^2\omega_1(\mathbf{B_x}\mathbf{t})\ |\mathbf{B_x}|\ \right\}$, and $\widetilde{\omega}_2(\mathbf{B_y}\mathbf{s}) = \left\{\|\varphi_{\boldsymbol{\varepsilon}_{\mathbf{y}}}^{-1}(\mathbf{s})\|^2\omega_2(\mathbf{B_y}\mathbf{s})\ |\mathbf{B_y}|\ \right\}$. Therefore, the above display, or equivalently, (2), can be reduced to the form of

$$\iint \|\varphi_{\mathbf{u},\mathbf{v}}(\mathbf{t},\mathbf{s}) - \varphi_{\mathbf{u}}(\mathbf{t})\varphi_{\mathbf{v}}(\mathbf{s})\|^2\widetilde{\omega}_1(\mathbf{t})\widetilde{\omega}_2(\mathbf{s})d\mathbf{t}d\mathbf{s},$$

Following [33], we set $\widetilde{\omega}_1(\mathbf{t}) = \Gamma\{(1+p)/2\}/\{\pi^{(1+p)/2}\|\mathbf{t}\|^{1+p}\}$, $\widetilde{\omega}_2(\mathbf{s}) = \Gamma\{(1+q)/2\}/\{\pi^{(1+q)/2}\|\mathbf{s}\|^{1+q}\}$, where $\Gamma(\cdot)$ is the gamma function. The above display has an explicit form of

$$\mathcal{V}^2(\mathbf{u},\mathbf{v}) = \mathrm{E}(\|\mathbf{u}_1-\mathbf{u}_2\|\|\mathbf{v}_1-\mathbf{v}_2\| - 2\|\mathbf{u}_1-\mathbf{u}_2\|\|\mathbf{v}_1-\mathbf{v}_3\| + \|\mathbf{u}_1-\mathbf{u}_2\|\|\mathbf{v}_3-\mathbf{v}_4\|),$$

where $\{(\mathbf{u}_i,\mathbf{v}_i), i\in\{1,\ldots,4\}\}$, are four independent copies of $(\mathbf{u},\mathbf{v})$. It is important to remark here that $\mathcal{V}^2(\mathbf{u},\mathbf{v})$ is indeed the squared distance covariance $\mathrm{dcov}^2(\mathbf{u},\mathbf{v})$ between $\mathbf{u}$ and $\mathbf{v}$, which indicates that we can simply use $\mathrm{dcov}^2(\mathbf{u},\mathbf{v})$ to measure the departure from independence between $\mathbf{x}$ and $\mathbf{y}$.

If we had set $\omega_1(\mathbf{t}) = \Gamma\{(1+p)/2\}/\{\pi^{(1+p)/2}\|\mathbf{t}\|^{1+p}\}$ and $\omega_2(\mathbf{s}) = \Gamma\{(1+q)/2\}/\{\pi^{(1+q)/2}\|\mathbf{s}\|^{1+q}\}$ in (2), then (2) would be $\mathcal{V}^2(\mathbf{x},\mathbf{y})$ [33], which replaces $\mathbf{u}$ and $\mathbf{v}$ in $\mathcal{V}^2(\mathbf{u},\mathbf{v})$ with $\mathbf{x}$ and $\mathbf{y}$ respectively. We propose to modify the weight functions by letting $\widetilde{\omega}_1(\mathbf{t}) = \Gamma\{(1+p)/2\}/\{\pi^{(1+p)/2}\|\mathbf{t}\|^{1+p}\}$, and $\widetilde{\omega}_2(\mathbf{s}) = \Gamma\{(1+q)/2\}/\{\pi^{(1+q)/2}\|\mathbf{s}\|^{1+q}\}$, which leads (2) to have the form of $\mathcal{V}^2(\mathbf{u},\mathbf{v})$. Both $\mathcal{V}^2(\mathbf{u},\mathbf{v}) = \mathrm{dcov}^2(\mathbf{u},\mathbf{v})$ and $\mathcal{V}^2(\mathbf{x},\mathbf{y}) = \mathrm{dcov}^2(\mathbf{x},\mathbf{y})$ have the same form, although the arguments are changed. We refer to this phenomenon as the invariance law throughout the present context.

Theorem 1. *Assume* $\mathrm{E}(\|\mathbf{x}\|+\|\mathbf{y}\|+\|\boldsymbol{\varepsilon}_{\mathbf{x}}\|+\|\boldsymbol{\varepsilon}_{\mathbf{y}}\|)<\infty$ *in Model (1). Then* $\mathbf{x}$ *is independent of* $\mathbf{y}$ *if and only if* $\mathcal{V}^2(\mathbf{u},\mathbf{v})=0$.

Proof of Theorem 1. It suffices to show that $\mathbf{x}$ is independent of $\mathbf{y}$ if and only if $\mathbf{u}$ is independent of $\mathbf{v}$. In Model (1), $\boldsymbol{\varepsilon}_{\mathbf{x}}$, $\boldsymbol{\varepsilon}_{\mathbf{y}}$ and $(\mathbf{x},\mathbf{y})$ are mutually independent. On one hand, if $\mathbf{x}$ is independent of $\mathbf{y}$, it then follows that $\boldsymbol{\varepsilon}_{\mathbf{x}}$, $\boldsymbol{\varepsilon}_{\mathbf{y}}$, $\mathbf{x}$ and $\mathbf{y}$ are mutually independent. Therefore, by the form of Model (1), $\mathbf{u}$ is independent of $\mathbf{v}$.

One the other hand, if $\mathbf{u}$ is independent of $\mathbf{v}$, it follows that $\varphi_{\mathbf{u},\mathbf{v}}(\mathbf{t},\mathbf{s}) = \varphi_{\mathbf{u}}(\mathbf{t})\varphi_{\mathbf{v}}(\mathbf{s})$. By Model (1), we have

$$\varphi_{\mathbf{u}}(\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{t}) = \exp(i\boldsymbol{\alpha}_{\mathbf{x}}{}^{\top}\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{t})\varphi_{\mathbf{x}}(\mathbf{t})\varphi_{\boldsymbol{\varepsilon}_{\mathbf{x}}}(\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{t}), \quad \varphi_{\mathbf{v}}(\mathbf{B}_{\mathbf{y}}^{-1}\mathbf{s}) = \exp(i\boldsymbol{\alpha}_{\mathbf{y}}{}^{\top}\mathbf{B}_{\mathbf{y}}^{-1}\mathbf{s})\varphi_{\mathbf{y}}(\mathbf{s})\varphi_{\boldsymbol{\varepsilon}_{\mathbf{y}}}(\mathbf{B}_{\mathbf{y}}^{-1}\mathbf{s}),$$
$$\varphi_{\mathbf{u},\mathbf{v}}(\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{t},\mathbf{B}_{\mathbf{y}}^{-1}\mathbf{s}) = \exp(i\boldsymbol{\alpha}_{\mathbf{x}}{}^{\top}\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{t} + i\boldsymbol{\alpha}_{\mathbf{y}}{}^{\top}\mathbf{B}_{\mathbf{y}}^{-1}\mathbf{s})\varphi_{\mathbf{x},\mathbf{y}}(\mathbf{t},\mathbf{s})\varphi_{\boldsymbol{\varepsilon}_{\mathbf{x}}}(\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{t})\varphi_{\boldsymbol{\varepsilon}_{\mathbf{y}}}(\mathbf{B}_{\mathbf{y}}^{-1}\mathbf{s}).$$

Therefore, $\varphi_{\mathbf{x},\mathbf{y}}(\mathbf{t},\mathbf{s})\varphi_{\boldsymbol{\varepsilon}_{\mathbf{x}}}(\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{t})\varphi_{\boldsymbol{\varepsilon}_{\mathbf{y}}}(\mathbf{B}_{\mathbf{y}}^{-1}\mathbf{s}) = \varphi_{\mathbf{x}}(\mathbf{t})\varphi_{\mathbf{y}}(\mathbf{s})\varphi_{\boldsymbol{\varepsilon}_{\mathbf{x}}}(\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{t})\varphi_{\boldsymbol{\varepsilon}_{\mathbf{y}}}(\mathbf{B}_{\mathbf{y}}^{-1}\mathbf{s})$. This implies that $\mathbf{x}$ is independent of $\mathbf{y}$. The proof is now completed. □

Theorem 1 reveals that, it suffices to work with $\mathcal{V}^2(\mathbf{u},\mathbf{v}) = \mathrm{dcov}^2(\mathbf{u},\mathbf{v})$ to measure the departure from independence between $\mathbf{x}$ and $\mathbf{y}$ in the presence of measurement errors.

To quantify the degree of dependence between $\mathbf{x}$ and $\mathbf{y}$ in the presence of measurement errors, we desire to obtain a number which ranges from zero to one. By the Cauchy–Schwartz inequality, $\mathcal{V}^2(\mathbf{u},\mathbf{v}) \le \mathcal{V}(\mathbf{u},\mathbf{u})\mathcal{V}(\mathbf{v},\mathbf{v})$. It is thus natural to normalize $\mathcal{V}^2(\mathbf{u},\mathbf{v})$ with $\mathcal{V}^2(\mathbf{u},\mathbf{v}) \big/ \{\mathcal{V}(\mathbf{u},\mathbf{u})\mathcal{V}(\mathbf{v},\mathbf{v})\}$, which ranges from zero to one. However, this normalized quantity, $\mathcal{V}^2(\mathbf{u},\mathbf{v}) \big/ \{\mathcal{V}(\mathbf{u},\mathbf{u})\mathcal{V}(\mathbf{v},\mathbf{v})\}$, attains one if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent [33]. In the measurement error model (1), we are interested in measuring the degree of dependence between $\mathbf{x}$ and $\mathbf{y}$, rather than $\mathbf{u}$ and $\mathbf{v}$. This normalized quantity does not attain one even if $\mathbf{x}$ and $\mathbf{y}$ are exactly linearly dependent. This phenomenon is described through the toy example in Section 1. Throughout the present context our goal is to characterize the dependence between $\mathbf{x}$ and $\mathbf{y}$. The invariance law ensures that we can use $\mathcal{V}^2(\mathbf{u},\mathbf{v})$ to characterize the dependence between $\mathbf{x}$ and $\mathbf{y}$. However, $\mathcal{V}(\mathbf{u},\mathbf{u})$ and $\mathcal{V}(\mathbf{v},\mathbf{v})$ characterize the respective variabilities of $\mathbf{u}$ and $\mathbf{v}$. They cannot be used to describe the variabilities of $\mathbf{x}$ and $\mathbf{y}$. The variabilities of $\mathbf{x}$ and $\mathbf{y}$ are quite different from those of $\mathbf{u}$ and $\mathbf{v}$.

It is important to describe the variabilities of $\mathbf{x}$ and $\mathbf{y}$ precisely. Suppose a repeated measurement of $(\mathbf{u},\mathbf{v})$, denoted as $(\widetilde{\mathbf{u}},\widetilde{\mathbf{v}})$, is available. It also admits the measurement error model (1) in the sense that $\widetilde{\mathbf{u}} = \boldsymbol{\alpha}_{\mathbf{x}} + \mathbf{B}_{\mathbf{x}}{}^{\top}\mathbf{x} + \widetilde{\boldsymbol{\varepsilon}}_{\mathbf{x}}$, and $\widetilde{\mathbf{v}} = \boldsymbol{\alpha}_{\mathbf{y}} + \mathbf{B}_{\mathbf{y}}{}^{\top}\mathbf{y} + \widetilde{\boldsymbol{\varepsilon}}_{\mathbf{y}}$, where $(\widetilde{\boldsymbol{\varepsilon}}_{\mathbf{x}},\widetilde{\boldsymbol{\varepsilon}}_{\mathbf{y}})$ is an independent copy of $(\boldsymbol{\varepsilon}_{\mathbf{x}},\boldsymbol{\varepsilon}_{\mathbf{y}})$. We define

$$\rho^2(\mathbf{x},\mathbf{y}) = \frac{\mathcal{V}^2(\mathbf{u},\mathbf{v})}{\mathcal{V}(\mathbf{u},\widetilde{\mathbf{u}})\mathcal{V}(\mathbf{v},\widetilde{\mathbf{v}})}.$$

In other words, we use $\mathcal{V}(\mathbf{u},\widetilde{\mathbf{u}})$ and $\mathcal{V}(\mathbf{v},\widetilde{\mathbf{v}})$, rather than $\mathcal{V}(\mathbf{u},\mathbf{u})$ and $\mathcal{V}(\mathbf{v},\mathbf{v})$, to characterize the variabilities of $\mathbf{x}$ and $\mathbf{y}$, respectively. The squared distance covariance, $\mathcal{V}^2(\mathbf{u},\mathbf{v})$, characterizes the dependence between $\mathbf{x}$ and $\mathbf{y}$. In this way, $\rho^2(\mathbf{x},\mathbf{y})$ has been normalized to be from zero to one, which is summarized below.

Theorem 2. *Assume* $\mathrm{E}(\|\mathbf{x}\| + \|\mathbf{y}\| + \|\boldsymbol{\varepsilon}_{\mathbf{x}}\| + \|\boldsymbol{\varepsilon}_{\mathbf{y}}\|) < \infty$, *then we have* $\rho^2(\mathbf{x},\mathbf{y}) \le 1$, *and the equality holds if and only if* $\mathrm{E}\{d(\mathbf{u}_1,\mathbf{u}_2) \mid (\mathbf{x}_1,\mathbf{x}_2)\}$ *is linear in* $\mathrm{E}\{d(\mathbf{v}_1,\mathbf{v}_2) \mid (\mathbf{y}_1,\mathbf{y}_2)\}$, *where*

$$d(\mathbf{z}_1,\mathbf{z}_2) = \|\mathbf{z}_1 - \mathbf{z}_2\| - \mathrm{E}(\|\mathbf{z}_1 - \mathbf{z}_2\| \mid \mathbf{z}_1) - \mathrm{E}(\|\mathbf{z}_1 - \mathbf{z}_2\| \mid \mathbf{z}_2) + \mathrm{E}(\|\mathbf{z}_1 - \mathbf{z}_2\|).$$

Proof of Theorem 2. According to [25], we can write $\mathcal{V}^2(\mathbf{u},\mathbf{v})$ as $\mathrm{cov}\{d(\mathbf{u}_1,\mathbf{u}_2), d(\mathbf{v}_1,\mathbf{v}_2)\}$. By the law of total covariance, $\mathrm{cov}\{d(\mathbf{u}_1,\mathbf{u}_2), d(\mathbf{v}_1,\mathbf{v}_2)\}$ is equal to

$$\mathrm{cov}[\mathrm{E}\{d(\mathbf{u}_1,\mathbf{u}_2) \mid (\mathbf{x}_1,\mathbf{x}_2)\}, \mathrm{E}\{d(\mathbf{v}_1,\mathbf{v}_2) \mid (\mathbf{y}_1,\mathbf{y}_2)\}] + \mathrm{E}[\mathrm{cov}\{d(\mathbf{u}_1,\mathbf{u}_2), d(\mathbf{v}_1,\mathbf{v}_2) \mid (\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)\}].$$

It is straightforward to verify that, conditional on $(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)$, $d(\mathbf{u}_1,\mathbf{u}_2)$ is independent of $d(\mathbf{v}_1,\mathbf{v}_2)$. This is because $\boldsymbol{\varepsilon}_{\mathbf{x}}$ is independent of $\boldsymbol{\varepsilon}_{\mathbf{y}}$. Therefore, $\mathrm{cov}\{d(\mathbf{u}_1,\mathbf{u}_2), d(\mathbf{v}_1,\mathbf{v}_2)\} = \mathrm{cov}[\mathrm{E}\{d(\mathbf{u}_1,\mathbf{u}_2) \mid (\mathbf{x}_1,\mathbf{x}_2)\}, \mathrm{E}\{d(\mathbf{v}_1,\mathbf{v}_2) \mid (\mathbf{y}_1,\mathbf{y}_2)\}]$. By the Cauchy–Schwarz inequality, $\{\mathcal{V}^2(\mathbf{u},\mathbf{v})\}^2 \le \mathrm{var}[\mathrm{E}\{d(\mathbf{u}_1,\mathbf{u}_2) \mid (\mathbf{x}_1,\mathbf{x}_2)\}]\mathrm{var}[\mathrm{E}\{d(\mathbf{v}_1,\mathbf{v}_2) \mid (\mathbf{y}_1,\mathbf{y}_2)\}]$, and the equality holds true if and only if $\mathrm{E}\{d(\mathbf{u}_1,\mathbf{u}_2) \mid (\mathbf{x}_1,\mathbf{x}_2)\}$ is linear in $\mathrm{E}\{d(\mathbf{v}_1,\mathbf{v}_2) \mid (\mathbf{y}_1,\mathbf{y}_2)\}$.

Next it suffices to show that $\mathcal{V}^2(\mathbf{u},\widetilde{\mathbf{u}}) = \mathrm{var}[\mathrm{E}\{d(\mathbf{u}_1,\mathbf{u}_2) \mid (\mathbf{x}_1,\mathbf{x}_2)\}]$. Following similar arguments, we have

$$\mathcal{V}^2(\mathbf{u},\widetilde{\mathbf{u}}) = \mathrm{cov}[\mathrm{E}\{d(\mathbf{u}_1,\mathbf{u}_2) \mid (\mathbf{x}_1,\mathbf{x}_2)\}, \mathrm{E}\{d(\widetilde{\mathbf{u}}_1,\widetilde{\mathbf{u}}_2) \mid (\mathbf{x}_1,\mathbf{x}_2)\}],$$

which equals $\mathrm{var}[\mathrm{E}\{d(\mathbf{u}_1,\mathbf{u}_2) \mid (\mathbf{x}_1,\mathbf{x}_2)\}]$ by noting that $\mathrm{E}\{d(\widetilde{\mathbf{u}}_1,\widetilde{\mathbf{u}}_2) \mid (\mathbf{x}_1,\mathbf{x}_2)\} = \mathrm{E}\{d(\mathbf{u}_1,\mathbf{u}_2) \mid (\mathbf{x}_1,\mathbf{x}_2)\}$. Therefore, we have shown that $\mathcal{V}^2(\mathbf{u},\mathbf{v}) \le \mathcal{V}(\mathbf{u},\widetilde{\mathbf{u}})\mathcal{V}(\widetilde{\mathbf{v}},\mathbf{v})$, and the equality holds true if and only if $\mathrm{E}\{d(\mathbf{u}_1,\mathbf{u}_2) \mid (\mathbf{x}_1,\mathbf{x}_2)\}$ is linear in $\mathrm{E}\{d(\mathbf{v}_1,\mathbf{v}_2) \mid (\mathbf{y}_1,\mathbf{y}_2)\}$. □

Theorem 2 ensures that $\rho^2(\mathbf{x},\mathbf{y})$ ranges from zero to one, with the value of one attainable. It is generally anticipated that using $\rho^2(\mathbf{x},\mathbf{y})$ to describe the dependence between $\mathbf{x}$ and $\mathbf{y}$ is more precise than using $\mathrm{dcorr}^2(\mathbf{u},\mathbf{v})$. To verify this from the finite-sample level, one may refer to Example 2 in Section 3 for the superior performance of $\rho^2(\mathbf{x},\mathbf{y})$ in terms of feature screening in the presence of measurement errors.

2.2. The sample level

Suppose a random sample of size n, $\{(\mathbf{u}_i,\mathbf{v}_i), i \in \{1,\ldots,n\}\}$, is available. At the sample level, we can estimate $\mathcal{V}^2(\mathbf{u},\mathbf{v})$ with $\mathcal{V}_n^2(\mathbf{u},\mathbf{v})$, where

$$\mathcal{V}_n^2(\mathbf{u},\mathbf{v}) \stackrel{\text{def}}{=} \{n(n-1)\}^{-1}\sum_{(i,j)}\|\mathbf{u}_i - \mathbf{u}_j\|\|\mathbf{v}_i - \mathbf{v}_j\| - 2\{n(n-1)(n-2)\}^{-1}\sum_{(i,j,k)}\|\mathbf{u}_i - \mathbf{u}_j\|\|\mathbf{v}_i - \mathbf{v}_k\|$$
$$+\{n(n-1)(n-2)(n-3)\}^{-1}\sum_{(i,j,k,l)}\|\mathbf{u}_i - \mathbf{u}_j\|\|\mathbf{v}_k - \mathbf{v}_l\|. \tag{3}$$

Throughout we use the notations (i, j), (i, j, k) and (i, j, k, l) to denote the summands which are all distinctive from each other. Apparently, $\mathcal{V}_n^2(\mathbf{u}, \mathbf{v})$ is trivially small if $\mathbf{x}$ and $\mathbf{y}$ are independent, and relatively large otherwise. The properties of $\mathcal{V}_n^2(\mathbf{u}, \mathbf{v})$, including the asymptotic behaviors [33] and fast computation algorithms [17,18], have been extensive investigated in literature. We summarize the asymptotic properties under the null hypothesis in what follows.

Proposition 1 (33)**.** *Assume* $\mathrm{E}(\|\mathbf{x}\| + \|\mathbf{y}\| + \|\boldsymbol{\varepsilon}_{\mathbf{x}}\| + \|\boldsymbol{\varepsilon}_{\mathbf{y}}\|) < \infty$. *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are independent,*

$$n\, \mathcal{V}_n^2(\mathbf{u}, \mathbf{v}) \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j \{\chi_j^2(1) - 1\},$$

where $\xrightarrow{d}$ *stands for "convergence in distribution",* $\{\chi_j^2(1), j \in \{1, 2, \ldots\}\}$, *are independent chi-square random variables with one degree of freedom, and* $\{\lambda_j, j \in \{1, 2, \ldots\}\}$, *are unknown weights which depend upon the joint distribution of* $\mathbf{u}$ *and* $\mathbf{v}$.

Proof of Proposition 1. We merely sketch the proof here because it is an existing result in [33]. According to Theorem 4.12 of [17], $\mathcal{V}_n^2(\mathbf{u}, \mathbf{v})$ can be written as

$$\{n(n-1)\}^{-1} \sum_{i \neq j} d(\mathbf{u}_i, \mathbf{u}_j) d(\mathbf{v}_i, \mathbf{v}_j) + o_p(n^{-1}).$$

The desired conclusion is an immediate consequence of Theorem 5.5.2 of [29]. We remark here that the weights $\lambda_j, j \in \{1, 2, \ldots\}$, are the solutions to the equation $\mathrm{E}\{d(\mathbf{u}_1, \mathbf{u}_2) d(\mathbf{v}_1, \mathbf{v}_2) g(\mathbf{u}_2, \mathbf{v}_2) \mid (\mathbf{u}, \mathbf{v})\} = \lambda g(\mathbf{u}, \mathbf{v})$. □

[17] showed that the weights λ_js satisfy that

$$\sum_{j=1}^{\infty} \lambda_j = \mathrm{E}\left(\|\mathbf{u}_1 - \mathbf{u}_2\|\right) \mathrm{E}\left(\|\mathbf{v}_1 - \mathbf{v}_2\|\right) \text{ and } \sum_{j=1}^{\infty} \lambda_j^2 = \mathcal{V}^2(\mathbf{u}, \mathbf{u}) \mathcal{V}^2(\mathbf{v}, \mathbf{v}).$$

This motivates us to approximate the asymptotic null distribution with Gamma distribution by matching the first two moments. However, this approximation is not sufficiently accurate when the sample size n is small. There are some other approximation methods in literature. See, for example, [5]. In this paper, we advocate using random permutations to approximate the asymptotic null distribution. The p-value is calculated as the fraction of replicated test statistics under random permutations. One may refer to [13] and [27] for a detailed description of permutation tests.

To measure the departure from independence between $\mathbf{x}$ and $\mathbf{y}$, the normalized quantity requires to estimate the distance variances $\mathcal{V}(\mathbf{u}, \widetilde{\mathbf{u}})$ and $\mathcal{V}(\mathbf{v}, \widetilde{\mathbf{v}})$. Towards this goal, we assume the repeated measurements are available. Let $(\mathbf{u}_i, \mathbf{v}_i, \widetilde{\mathbf{u}}_i, \widetilde{\mathbf{v}}_i), i \in \{1, \ldots, n\}$ be a random sample of size n. We estimate $\mathcal{V}(\mathbf{u}, \widetilde{\mathbf{u}})$ and $\mathcal{V}(\mathbf{v}, \widetilde{\mathbf{v}})$ in the same way as (3) to obtain $\mathcal{V}_n(\mathbf{u}, \widetilde{\mathbf{u}})$ and $\mathcal{V}_n(\mathbf{v}, \widetilde{\mathbf{v}})$, and estimate $\rho^2(\mathbf{x}, \mathbf{y})$ with

$$\widehat{\rho}^2(\mathbf{x}, \mathbf{y}) = \frac{\mathcal{V}_n^2(\mathbf{u}, \mathbf{v}) + \mathcal{V}_n^2(\mathbf{u}, \widetilde{\mathbf{v}}) + \mathcal{V}_n^2(\widetilde{\mathbf{u}}, \mathbf{v}) + \mathcal{V}_n^2(\widetilde{\mathbf{u}}, \widetilde{\mathbf{v}})}{4\mathcal{V}_n(\mathbf{u}, \widetilde{\mathbf{u}})\mathcal{V}_n(\mathbf{v}, \widetilde{\mathbf{v}})}.$$

Because $\mathcal{V}_n^2(\mathbf{u}, \mathbf{v})$, $\mathcal{V}_n^2(\mathbf{u}, \widetilde{\mathbf{v}})$, $\mathcal{V}_n^2(\widetilde{\mathbf{u}}, \mathbf{v})$, $\mathcal{V}_n^2(\widetilde{\mathbf{u}}, \widetilde{\mathbf{v}})$ are all consistent estimates of $\mathcal{V}^2(\mathbf{u}, \mathbf{v})$, we take average of these four estimates to improve estimation efficacy. The asymptotic distribution of $\widehat{\rho}^2(\mathbf{x}, \mathbf{y})$, when $\mathbf{x}$ and $\mathbf{y}$ are not necessarily independent, is summarized in the following.

Theorem 3. *Assume* $\mathrm{E}(\|\mathbf{x}\| + \|\mathbf{y}\| + \|\boldsymbol{\varepsilon}_{\mathbf{x}}\| + \|\boldsymbol{\varepsilon}_{\mathbf{y}}\|) < \infty$. *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are dependent, then*

$$n^{1/2}\{\widehat{\rho}^2(\mathbf{x}, \mathbf{y}) - \rho^2(\mathbf{x}, \mathbf{y})\} \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = var(Z)$ *and* Z *is defined in* (5).

Proof of Theorem 3. By using the Hoeffding decomposition of U-statistic theory [29, Lemma 5.1.5A], $\mathcal{V}_n^2(\mathbf{u}, \mathbf{v}) - \mathcal{V}^2(\mathbf{u}, \mathbf{v})$ can be written as

$$2n^{-1} \sum_{i=1}^{n} \{h(\mathbf{u}_i, \mathbf{v}_i) - \mathrm{E}h(\mathbf{u}, \mathbf{v})\} + o_p(n^{-1/2}). \tag{4}$$

In the above decomposition, we merely keep the leading term. In addition, $h(\mathbf{u}_i, \mathbf{v}_i)$, for $i \in \{1, \ldots, n\}$, are independent copies of $h(\mathbf{u}, \mathbf{v})$, which is defined by

$$\begin{aligned} h(\mathbf{u}, \mathbf{v}) \stackrel{\text{def}}{=} \mathrm{E}\Big\{ & \|\mathbf{u} - \mathbf{u}_1\| \|\mathbf{v} - \mathbf{v}_1\| - \|\mathbf{u} - \mathbf{u}_1\| \|\mathbf{v} - \mathbf{v}_2\| - \|\mathbf{u}_1 - \mathbf{u}_2\| \|\mathbf{v} - \mathbf{v}_2\| \\ & - \|\mathbf{u} - \mathbf{u}_1\| \|\mathbf{v}_2 - \mathbf{v}_1\| + \|\mathbf{u}_1 - \mathbf{u}_2\| \|\mathbf{v} - \mathbf{v}_3\| + \|\mathbf{u} - \mathbf{u}_1\| \|\mathbf{v}_2 - \mathbf{v}_3\| \Big| (\mathbf{u}, \mathbf{v}) \Big\}. \end{aligned}$$

We emphasize that $h(\mathbf{u},\mathbf{v})$ is a function of $(\mathbf{u},\mathbf{v})$. Similarly, we derive that

$$\mathcal{V}_n^2(\mathbf{u},\widetilde{\mathbf{v}})-\mathcal{V}^2(\mathbf{u},\widetilde{\mathbf{v}})=2n^{-1}\sum_{i=1}^n\{h(\mathbf{u}_i,\widetilde{\mathbf{v}}_i)-\mathrm{E}h(\mathbf{u},\widetilde{\mathbf{v}})\}+o_p(n^{-1/2}),$$
$$\mathcal{V}_n^2(\widetilde{\mathbf{u}},\mathbf{v})-\mathcal{V}^2(\widetilde{\mathbf{u}},\mathbf{v})=2n^{-1}\sum_{i=1}^n\{h(\widetilde{\mathbf{u}}_i,\mathbf{v}_i)-\mathrm{E}h(\widetilde{\mathbf{u}},\mathbf{v})\}+o_p(n^{-1/2}),$$
$$\mathcal{V}_n^2(\widetilde{\mathbf{u}},\widetilde{\mathbf{v}})-\mathcal{V}^2(\widetilde{\mathbf{u}},\widetilde{\mathbf{v}})=2n^{-1}\sum_{i=1}^n\{h(\widetilde{\mathbf{u}}_i,\widetilde{\mathbf{v}}_i)-\mathrm{E}h(\widetilde{\mathbf{u}},\widetilde{\mathbf{v}})\}+o_p(n^{-1/2}).$$

Apparently, $\mathcal{V}^2(\mathbf{u},\mathbf{v})=\mathcal{V}^2(\mathbf{u},\widetilde{\mathbf{v}})=\mathcal{V}^2(\widetilde{\mathbf{u}},\mathbf{v})=\mathcal{V}^2(\widetilde{\mathbf{u}},\widetilde{\mathbf{v}})$ and $\mathrm{E}h(\mathbf{u},\mathbf{v})=\mathrm{E}h(\mathbf{u},\widetilde{\mathbf{v}})=\mathrm{E}h(\widetilde{\mathbf{u}},\mathbf{v})=\mathrm{E}h(\widetilde{\mathbf{u}},\widetilde{\mathbf{v}})$. We have

$$\begin{aligned}&4^{-1}\left\{\mathcal{V}_n^2(\mathbf{u},\mathbf{v})+\mathcal{V}_n^2(\mathbf{u},\widetilde{\mathbf{v}})+\mathcal{V}_n^2(\widetilde{\mathbf{u}},\mathbf{v})+\mathcal{V}_n^2(\widetilde{\mathbf{u}},\widetilde{\mathbf{v}})\right\}-\mathcal{V}^2(\mathbf{u},\mathbf{v})\\&\quad=(2n)^{-1}\sum_{i=1}^n\{h(\mathbf{u}_i,\mathbf{v}_i)+h(\mathbf{u}_i,\widetilde{\mathbf{v}}_i)+h(\widetilde{\mathbf{u}}_i,\mathbf{v}_i)+h(\widetilde{\mathbf{u}}_i,\widetilde{\mathbf{v}}_i)-4\mathrm{E}h(\mathbf{u},\mathbf{v})\}+o_p(n^{-1/2}).\end{aligned}$$

By using the Hoeffding decomposition of U-statistic theory again, we have

$$\mathcal{V}_n^2(\mathbf{u},\widetilde{\mathbf{u}})-\mathcal{V}^2(\mathbf{u},\widetilde{\mathbf{u}})=2n^{-1}\sum_{i=1}^n\{h(\mathbf{u}_i,\widetilde{\mathbf{u}}_i)-\mathrm{E}h(\mathbf{u},\widetilde{\mathbf{u}})\}+o_p(n^{-1/2}),$$

which yields

$$\mathcal{V}_n(\mathbf{u},\widetilde{\mathbf{u}})-\mathcal{V}(\mathbf{u},\widetilde{\mathbf{u}})=\{n\mathcal{V}(\mathbf{u},\widetilde{\mathbf{u}})\}^{-1}\sum_{i=1}^n\{h(\mathbf{u}_i,\widetilde{\mathbf{u}}_i)-\mathrm{E}h(\mathbf{u},\widetilde{\mathbf{u}})\}+o_p(n^{-1/2}).$$

Similarly, we can obtain that

$$\mathcal{V}_n(\mathbf{v},\widetilde{\mathbf{v}})-\mathcal{V}(\mathbf{v},\widetilde{\mathbf{v}})=\{n\mathcal{V}(\mathbf{v},\widetilde{\mathbf{v}})\}^{-1}\sum_{i=1}^n\{h(\mathbf{v}_i,\widetilde{\mathbf{v}}_i)-\mathrm{E}h(\mathbf{v},\widetilde{\mathbf{v}})\}+o_p(n^{-1/2}).$$

It follows that

$$\begin{aligned}\mathcal{V}_n(\mathbf{u},\widetilde{\mathbf{u}})\mathcal{V}_n(\mathbf{v},\widetilde{\mathbf{v}})-\mathcal{V}(\mathbf{u},\widetilde{\mathbf{u}})\mathcal{V}(\mathbf{v},\widetilde{\mathbf{v}})=&\ \{n\mathcal{V}(\mathbf{u},\widetilde{\mathbf{u}})\}^{-1}\,\mathcal{V}(\mathbf{v},\widetilde{\mathbf{v}})\sum_{i=1}^n\{h(\mathbf{u}_i,\widetilde{\mathbf{u}}_i)-\mathrm{E}h(\mathbf{u},\widetilde{\mathbf{u}})\}\\&+\{n\mathcal{V}(\mathbf{v},\widetilde{\mathbf{v}})\}^{-1}\,\mathcal{V}(\mathbf{u},\widetilde{\mathbf{u}})\sum_{i=1}^n\{h(\mathbf{v}_i,\widetilde{\mathbf{v}}_i)-\mathrm{E}h(\mathbf{v},\widetilde{\mathbf{v}})\}+o_p(n^{-1/2}).\end{aligned}$$

With Taylor's expansion, we can write $\widehat{\rho}^2(\mathbf{x},\mathbf{y})-\rho^2(\mathbf{x},\mathbf{y})$ as

$$\begin{aligned}&\frac{\mathcal{V}_n^2(\mathbf{u},\mathbf{v})+\mathcal{V}_n^2(\mathbf{u},\widetilde{\mathbf{v}})+\mathcal{V}_n^2(\widetilde{\mathbf{u}},\mathbf{v})+\mathcal{V}_n^2(\widetilde{\mathbf{u}},\widetilde{\mathbf{v}})-4\mathcal{V}^2(\mathbf{u},\mathbf{v})}{4\mathcal{V}(\mathbf{u},\widetilde{\mathbf{u}})\mathcal{V}(\mathbf{v},\widetilde{\mathbf{v}})}\\&\qquad-\frac{\mathcal{V}^2(\mathbf{u},\mathbf{v})\{\mathcal{V}_n(\mathbf{u},\widetilde{\mathbf{u}})\mathcal{V}_n(\mathbf{v},\widetilde{\mathbf{v}})-\mathcal{V}(\mathbf{u},\widetilde{\mathbf{u}})\mathcal{V}(\mathbf{v},\widetilde{\mathbf{v}})\}}{\mathcal{V}^2(\mathbf{u},\widetilde{\mathbf{u}})\mathcal{V}^2(\mathbf{v},\widetilde{\mathbf{v}})}+o_p(n^{-1/2}).\end{aligned}$$

Combining the above results, we obtain that

$$\widehat{\rho}^2(\mathbf{x},\mathbf{y})-\rho^2(\mathbf{x},\mathbf{y})=n^{-1}\sum_{i=1}^n Z_i+o_p(n^{-1/2}),$$

where Z_i, $i\in\{1,\ldots,n\}$, are independent copies of Z defined as

$$\begin{aligned}Z\overset{\text{def}}{=}&\ \{2\mathcal{V}(\mathbf{u},\widetilde{\mathbf{u}})\mathcal{V}(\mathbf{v},\widetilde{\mathbf{v}})\}^{-1}\{h(\mathbf{u},\mathbf{v})+h(\mathbf{u},\widetilde{\mathbf{v}})+h(\widetilde{\mathbf{u}},\mathbf{v})+h(\widetilde{\mathbf{u}},\widetilde{\mathbf{v}})-4\mathrm{E}h(\mathbf{u},\mathbf{v})\}\\&-\mathcal{V}^2(\mathbf{u},\mathbf{v})\mathcal{V}^{-3}(\mathbf{u},\widetilde{\mathbf{u}})\mathcal{V}^{-1}(\mathbf{v},\widetilde{\mathbf{v}})\{h(\mathbf{u},\widetilde{\mathbf{u}})-\mathrm{E}h(\mathbf{u},\widetilde{\mathbf{u}})\}-\mathcal{V}^2(\mathbf{u},\mathbf{v})\mathcal{V}^{-1}(\mathbf{u},\widetilde{\mathbf{u}})\mathcal{V}^{-3}(\mathbf{v},\widetilde{\mathbf{v}})\{h(\mathbf{v},\widetilde{\mathbf{v}})-\mathrm{E}h(\mathbf{v},\widetilde{\mathbf{v}})\}.\end{aligned}\tag{5}$$

Then the proof is completed by using the central limit theory and Cramér–Slutsky's theorem. □

3. Numerical studies

We conduct numerical studies to evaluate the finite-sample performance of $\mathcal{V}^2(\mathbf{u},\mathbf{v})$ and $\rho^2(\mathbf{x},\mathbf{y})$ in the presence of measurement errors. We remark here that, to test independence between $\mathbf{x}$ and $\mathbf{y}$, we merely use $\mathcal{V}^2(\mathbf{u},\mathbf{v})$ together

Table 1
Empirical sizes for Model I and powers for Models II and III for different dimensions and different levels of measurement errors. The significance level is fixed at $\alpha = 0.05$.

Model	d	c					
		0	0.2	0.4	0.6	0.8	1
	5	0.060	0.046	0.054	0.060	0.044	0.054
I	10	0.033	0.045	0.054	0.057	0.041	0.039
	20	0.049	0.054	0.041	0.066	0.047	0.049
	5	1.000	1.000	1.000	1.000	1.000	1.000
II	10	1.000	1.000	1.000	1.000	1.000	1.000
	20	1.000	1.000	1.000	1.000	1.000	1.000
	5	1.000	1.000	1.000	1.000	0.999	0.972
III	10	1.000	1.000	1.000	1.000	0.994	0.960
	20	1.000	0.998	0.999	0.992	0.947	0.871

with random permutations, and do not require repeated measurements be available. However, to measure nonlinear dependence, we shall have to use $\rho^2(\mathbf{x}, \mathbf{y})$ and require the repeated measurements be available. We shall also demonstrate how to use $\rho^2(\mathbf{x}, \mathbf{y})$ to perform feature screening in high dimensions.

Example 1. We use $\mathcal{V}^2(\mathbf{u}, \mathbf{v})$ to test independence between $\mathbf{x} = (X_1, \ldots, X_p)^\top$ and $\mathbf{y} = (Y_1, \ldots, Y_q)^\top$ in the presence of measurement errors. In this example, $p = q = d$. We consider three models, for $k \in \{1, \ldots, d\}$,

$$\text{Model I}: Y_k = \varepsilon_k, \quad \text{Model II}: Y_k = X_k + \varepsilon_k, \quad \text{Model III}: Y_k = \exp(X_k)\varepsilon_k.$$

In the above models, $\mathbf{x}$ is multivariate normal with mean zero and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{k\ell})_{d\times d}$, where $\sigma_{k\ell} = 0.5^{|k-\ell|}$, $k, \ell \in \{1, \ldots, d\}$. In addition, ε_ks are all independent standard normal. In Model I, $\mathbf{x}$ and $\mathbf{y}$ are independent, whereas in Models II and III, $\mathbf{x}$ and $\mathbf{y}$ are dependent. Instead of using $(\mathbf{x}, \mathbf{y})$, we merely use $(\mathbf{u}, \mathbf{v})$, where $\mathbf{u} = \mathbf{x} + c\boldsymbol{\varepsilon}_\mathbf{x}$ and $\mathbf{v} = \mathbf{y} + c\boldsymbol{\varepsilon}_\mathbf{y}$, $\boldsymbol{\varepsilon}_\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ and $\boldsymbol{\varepsilon}_\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ are independent measurement errors. The intensity parameter c is used to control the magnitude of measurement errors. We set the sample size $n = 100$, vary the dimension d from $\{5, 10, 20\}$. We increase the intensity parameter c from 0 to 1, step by 0.2. We fix the significance level $\alpha = 0.05$.

The empirical sizes for Model I and powers for Models II and III are charted in Table 1. The results for Model I imply that, we can maintain the type-I error rates pretty well. For Models II and III, our proposed test is powerful to detect both linear and nonlinear dependence in the presence of measurement errors. When $p = q = 20$ and $c = 1$, the empirical power of our proposed test is still as high as 0.871 in Model III.

Example 2. We demonstrate how to use $\rho^2(X_k, \mathbf{y})$, for $k \in \{1, \ldots, p\}$, to perform feature screening in regressions with high dimensional covariates $\mathbf{x} = (X_1, \ldots, X_p)^\top$. Let $\mathbf{y} = (Y_1, \ldots, Y_q)^\top$. We consider three models,

$$\begin{aligned} \text{Model IV}: \ \mathbf{y} &= X_1 + X_2X_3 + X_4 + \varepsilon, \\ \text{Model V}: \ \mathbf{y} &= X_1^2 + X_2^2 + 2\sin(X_3) + \exp(X_4)\varepsilon, \\ \text{Model VI}: \ Y_1 &= \sin(X_1) + X_2^2 + \varepsilon, \ Y_2 = X_3^2 + \exp(X_4)\widetilde{\varepsilon}. \end{aligned}$$

In the above models, $\mathbf{x}$ is multivariate normal with mean zero and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{k\ell})_{p\times p}$, where $\sigma_{k\ell} = 0.5^{|k-\ell|}$, $k, \ell \in \{1, \ldots, p\}$. In addition, ε and $\widetilde{\varepsilon}$ are independent standard normal. The response $\mathbf{y}$ is univariate in Models IV and V, and multivariate in Model VI. Let $\mathbf{u} = (U_1, \ldots, U_p)^\top$, where $U_k = X_k + \varepsilon_{x_k}$, $k \in \{1, \ldots, 100\}$ and $U_k = X_k$, $k \in \{101, \ldots, p\}$. Let $\mathbf{v} = (V_1, \ldots, V_q)^\top$, and $V_k = Y_k + \varepsilon_{y_k}$. All measurement errors are independent standard normal. In these experiments, $n = 200$ and $p = 2000$. In addition, we assume the repeated measurements $(\mathbf{u}, \mathbf{v}, \widetilde{\mathbf{u}}, \widetilde{\mathbf{v}})$ are available to measure the degree of dependence between $\mathbf{x}$ and $\mathbf{y}$.

Existing screening procedures proceed as follows. A particular marginal utility is used to quantify the degree of dependence between X_k and $\mathbf{y}$, for $k \in \{1, \ldots, p\}$. The importance of each covariate X_k is measured by the magnitude of this marginal utility. The larger the magnitude is, the more important this covariate is. We retain the top $\lfloor n/\ln n \rfloor$ covariates which have the largest utilities, where $\lfloor x \rfloor$ denotes the integer part of x.

We evaluate the performance of three independent screening procedures, "ρ-SIS", "Naive" DC-SIS and "Oracle" DC-SIS. The corresponding marginal utilities are $\rho^2(X_k, \mathbf{y})$, $\mathrm{dcorr}(U_k, \mathbf{v})$ and $\mathrm{dcorr}(X_k, \mathbf{y})$, respectively. We remark here that we estimate $\rho^2(X_k, \mathbf{y})$ using $(\mathbf{u}_i, \mathbf{v}_i)$, $i \in \{1, \ldots, n\}$ instead of using $(\mathbf{x}_i, \mathbf{y}_i)$, $i \in \{1, \ldots, n\}$.

We replicate each experiment 1000 times, and use three criteria to compare the performance of screening procedures: (1) the proportion that each single important covariate (denoted by $\mathcal{P}_s$) are selected out of 1000 replications, (2) the proportion that all important covariates (denoted by $\mathcal{P}_a$) are selected, and (3) the 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size required to include all important covariates.

The simulation results for Models IV, V and VI are reported in Table 2. The "Oracle" DC-SIS procedure is not surprisingly the best, which serves as a benchmark for comparison. The "ρ-SIS" follows, which is in line with our anticipations. However, the "Naive" DC-SIS is the worst across all scenarios. This phenomenon echoes the observation we made in the toy example in Section 1.

Table 2
The 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size, the proportion that each single important covariate and all important covariates are selected for Example 2.

Model	Method	Minimum Model Size					$\mathcal{P}_s$				$\mathcal{P}_a$
		5%	25%	50%	75%	95%	X_1	X_2	X_3	X_4	
IV	ρ-SIS	4	4	5	7	18	1.000	0.996	0.993	1.000	0.990
	Naive	4	5	11	38	170	1.000	0.863	0.848	0.999	0.748
	Oracle	4	4	4	4	6	1.000	0.999	0.999	1.000	0.998
V	ρ-SIS	4	7	13	23	54	0.941	0.981	1.000	0.976	0.903
	Naive	16	65	135	286	666	0.317	0.608	0.982	0.607	0.140
	Oracle	4	4	5	8	27	0.996	0.999	1.000	0.965	0.961
VI	ρ-SIS	6	11	18	30	55	0.953	0.951	0.964	0.961	0.841
	Naive	18	52	128	252	582	0.692	0.556	0.574	0.627	0.159
	Oracle	4	4	4	4	10	0.997	1.000	1.000	0.998	0.995

Table 3
The 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size, the proportion that each single important covariate and all important covariates are selected for Example 3.

Method	Minimum Model Size					$\mathcal{P}_s$		$\mathcal{P}_a$
	5%	25%	50%	75%	95%	Msa.2877.0	Msa.2134.0	
ρ-SIS	2	4	6	8	36	0.882	0.898	0.790
Naive	47	154	246	419	748	0.054	0.038	0.005

Example 3. We apply the screening procedures in Example 2 to the Cardiomyopathy microarray dataset. This dataset was analyzed by [15,28], and [22], etc. It contains $n = 30$ samples. The response $\mathbf{y}$ is the Ro1 expression level, and the covariates $\mathbf{x}$ is $p = 6319$ gene expression levels. [22] identified two genes, labeled Msa.2134.0 and Msa.2877.0, as the two most important covariates. They fitted an additive model using these two covariates, yielding the adjusted R^2 of 96.8% and the deviance explained of 98.3%. It is thus natural to treat these two genes as the two most important covariates.

We standardize all covariates marginally. To increase the difficulty of identifying these two genes, we add standard normal noises to the top 10 genes selected by [22]. We treat these noise-added data as surrogate observations. We apply the"Naive" DC-SIS procedure to these surrogate observations, and re-select the important covariates. We replicate this procedure 1000 times. The corresponding 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size that includes both Msa.2134.0 and Msa.2877.0 are reported in Table 3. We also include $\mathcal{P}_s$ and $\mathcal{P}_a$, the proportions that Msa.2134.0 and Msa.2877.0 are selected individually and simultaneously, if we select the top $\lfloor n/\ln n \rfloor = 8$ genes. The results are charted in Table 3. The proportion of either Msa.2134.0 or Msa.2877.0 is selected is around 0.05. Moreover, the 5% quantile of the minimum model size is 47, which means that, if we select the top 47 genes ranked by the "Naive" DC-SIS procedure, we only have a 5% chance to include both Msa.2134.0 and Msa.2877.0.

For the purposes of comparison, we apply the ρ-SIS procedure when repeated measurements are available in the surrogate observations. These results are also charted in Table 3. The proportion of selecting Msa.2134.0 and Msa.2877.0 simultaneously when we only select the top 8 genes is 0.79, which is much larger than that of the "Naive" DC-SIS procedure. The 95% quantile of the minimum model size is 36, much smaller than that of the "Naive" DC-SIS procedure. This indicates that the ρ-SIS procedure is much more effective than the "Naive" DC-SIS procedure when the observations are subject to substantial measurement errors.

4. Conclusions

All observations are measured with random errors. In some situations these measurement errors are ignorable, but not in others. We consider testing statistical independence and measuring nonlinear dependence when the measurement errors are substantial. This issue has rarely been touched in literature. To test statistical independence, we observe an invariance law, which simply replaces the random vectors of primary interest with the surrogate ones. To measure nonlinear dependence, we introduce a novel estimation procedure for the distance variances when the repeated measurements are available. In addition, we propose an independence screening procedure when the random vectors are high dimensional.

There are several issues which deserve delicate investigations. For example, in the present context our observations are all based on distance correlation. It is natural to ask whether similar results carry over to other independence measures such as projection correlation. How to conduct conditional independence test in the presence of measurement errors is another important issue. Investigations along these directions are under way.

CRediT authorship contribution statement

Jinlin Fan: Performed the computations. **Yaowu Zhang:** Conceived of the presented idea, Performed the computations, Developed the theory, Writing – original draft. **Liping Zhu:** Conceived of the presented idea, Developed the theory, Writing – original draft.

Acknowledgments

All authors contributed equally to this work, and their names are listed in an alphabetical order. This work is supported by National Natural Science Foundation (NNSF) of China (12171477, 11801349, 11731011, 11931014) and NNSF of Beijing, China (Z19J00009), and in part by the Program for Innovative Research Team of Shanghai University of Finance and Economics, China. All authors discussed the results and contributed to the final manuscript.

References

[1] A. Belloni, V. Chernozhukov, A. Kaul, Confidence bands for coefficients in high dimensional linear models with error-in-variables, 2017, arXiv preprint arXiv:1703.00469.
[2] W. Bergsma, A. Dassios, A consistent test of independence based on a sign covariance related to Kendall's tau, Bernoulli 20 (2) (2014) 1006–1028.
[3] P. Bickel, Y. Ritov, Efficient estimation in the errors in variables model, Ann. Statist. 15 (2) (1987) 513–540.
[4] J.R. Blum, J. Kiefer, M. Rosenblatt, Distribution free tests of independence based on the sample distribution function, Ann. Math. Stat. (1961) 485–498.
[5] D.A. Bodenham, N.M. Adams, A comparison of efficient approximations for a weighted sum of chi-squared random variables, Stat. Comput. 26 (4) (2016) 917–928.
[6] R.J. Carroll, K.-C. Li, Measurement error regression with unknown link: dimension reduction and data visualization, J. Amer. Statist. Assoc. 87 (420) (1992) 1040–1050.
[7] R.J. Carroll, D. Ruppert, L.A. Stefanski, C.M. Crainiceanu, Measurement Error in Nonlinear Models: A Modern Perspective, Chapman and Hall/CRC, New York, 2006.
[8] S. Chatterjee, A new coefficient of correlation, 2020, ArXiv Preprint ArXiv:1909.10140v4, in press (JASA).
[9] L.-P. Chen, Feature screening based on distance correlation for ultrahigh-dimensional censored data with covariate measurement error, Comput. Statist. 36 (2) (2021) 857–884.
[10] C.-L. Cheng, C.-L. Tsai, The invariance of some score tests in the linear model with classical measurement error, J. Amer. Statist. Assoc. 99 (467) (2004) 805–809.
[11] R.D. Cook, Regression Graphics: Ideas for Studying Regressions Through Graphics, vol. 482, John Wiley & Sons, New York, 2009.
[12] W.A. Fuller, Measurement Error Models, vol. 305, John Wiley & Sons, New York, 2009.
[13] P.I. Good, Permutation, Parametric, and Bootstrap Tests of Hypotheses, Springer Science & Business Media, New York, 2006.
[14] A. Gretton, K. Fukumizu, C.H. Teo, L. Song, B. Schölkopf, A.J. Smola, A kernel statistical test of independence, in: Advances in Neural Information Processing Systems, 2008, pp. 585–592.
[15] P. Hall, H. Miller, Using generalized correlation to effect variable selection in very high dimensional problems, J. Comput. Graph. Statist. 18 (3) (2009) 533–550.
[16] W. Hoeffding, A non-parametric test of independence, Ann. Math. Stat. (1948) 546–557.
[17] C. Huang, X. Huo, A statistically and numerically efficient independence test based on random projections and distance covariance, 2017, arXiv preprint arXiv:1701.06054.
[18] X. Huo, G.J. Székely, Fast computing for distance covariance, Technometrics 58 (4) (2016) 435–447.
[19] K.-C. Li, Sliced inverse regression for dimension reduction, J. Amer. Statist. Assoc. 86 (414) (1991) 316–327.
[20] M. Li, R. Li, Y. Ma, Inference in high dimensional linear measurement error models, J. Multivariate Anal. 184 (2021) 104759.
[21] B. Li, X. Yin, On surrogate dimension reduction for measurement error regression: an invariance law, Ann. Statist. 35 (5) (2007) 2143–2172.
[22] R. Li, W. Zhong, L. Zhu, Feature screening via distance correlation learning, J. Amer. Statist. Assoc. 107 (499) (2012) 1129–1139.
[23] H. Liang, W. Härdle, R.J. Carroll, Estimation in a semiparametric partially linear errors-in-variables model, Ann. Statist. 27 (5) (1999) 1519–1535.
[24] H. Liang, R. Li, Variable selection for partially linear models with measurement errors, J. Amer. Statist. Assoc. 104 (485) (2009) 234–248.
[25] R. Lyons, Distance covariance in metric spaces, Ann. Probab. 41 (5) (2013) 3284–3305.
[26] Y. Ma, R. Li, Variable selection in measurement error models, Bernoulli 16 (1) (2010) 274.
[27] F. Pesarin, L. Salmaso, Permutation Tests for Complex Data: Theory, Applications and Software, John Wiley & Sons, New York, 2010.
[28] M.R. Segal, K.D. Dahlquist, B.R. Conklin, Regression approaches for microarray data analysis, J. Comput. Biol. 10 (6) (2003) 961–980.
[29] R.J. Serfling, Approximation Theorems of Mathematical Statistics, vol. 162, John Wiley & Sons, New York, 2009.
[30] H. Shi, M. Drton, F. Han, Distribution-free consistent independence tests via center-outward ranks and signs, J. Amer. Statist. Assoc. (2020a) 1–16.
[31] H. Shi, M. Drton, F. Han, On the power of chatterjee rank correlation, 2020b, arXiv preprint arXiv:2008.11619, in press (Biometrika).
[32] L.A. Stefanski, R.J. Carroll, Conditional scores and optimal scores for generalized linear measurement-error models, Biometrika 74 (4) (1987) 703–716.
[33] G.J. Székely, M.L. Rizzo, N.K. Bakirov, Measuring and testing dependence by correlation of distances, Ann. Statist. 35 (6) (2007) 2769–2794.
[34] Y. Wang, J. Wang, S. Balakrishnan, A. Singh, Rate optimal estimation and confidence intervals for high-dimensional regression with missing covariates, J. Multivariate Anal. 174 (2019).
[35] X. Zeng, Y. Xia, H. Tong, Jackknife approach to the estimation of mutual information, Proc. Natl. Acad. Sci. 115 (40) (2018) 9956–9961.
[36] L. Zhu, K. Xu, R. Li, W. Zhong, Projection correlation between two random vectors, Biometrika 104 (4) (2017) 829–843.