Computational Statistics and Data Analysis 160 (2021) 107246

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

A kernel-based measure for conditional mean dependence

Tingyu Lai [a], Zhongzhan Zhang [a,*], Yafei Wang [b]

[a] Beijing University of Technology, Beijing, 100124, China
[b] University of Alberta, Edmonton, AB, T6G 2R3, Canada

ARTICLE INFO

Article history:
Received 11 October 2020
Received in revised form 24 January 2021
Accepted 2 April 2021
Available online 9 April 2021

Keywords:
Functional data
Conditional mean dependence
Hilbert-Schmidt norm
U-statistics

ABSTRACT

A novel metric, called kernel-based conditional mean dependence (KCMD), is proposed to measure and test the departure from conditional mean independence between a response variable Y and a predictor variable X, based on the reproducing kernel embedding and the Hilbert-Schmidt norm of a tensor operator. The KCMD has several appealing merits. It equals zero if and only if the conditional mean of Y given X is independent of X, i.e. $E(Y|X) = E(Y)$ almost surely, provided that the employed kernel is characteristic; it can be used to detect all kinds of conditional mean dependence with an appropriate choice of kernel; it has a simple expectation form and allows an unbiased empirical estimator. A class of test statistics based on the estimated KCMD is constructed, and a wild bootstrap test procedure to the conditional mean independence is presented. The limit distributions of the test statistics and the bootstrapped statistics under null hypothesis, fixed alternative hypothesis and local alternative hypothesis are given respectively, and a data-driven procedure to choose a suitable kernel is suggested. Simulation studies indicate that the tests based on the KCMD have close powers to the tests based on martingale difference divergence in monotone dependence, but excel in the cases of nonlinear relationships or the moment restriction on X is violated. Two real data examples are presented for the illustration of the proposed method.

1. Introduction

Suppose that X and Y are random elements in separable Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. We aim to test the hypothesis

$$H_0 : E(Y|X) = E(Y) \text{ almost surely, versus } H_1 : \mathrm{pr}\{E(Y|X) = E(Y)\} < 1, \tag{1}$$

by which we assess whether the predictor X has a contribution to the mean of response Y. Measuring and testing conditional mean dependence plays a vital role in statistics. When H_0 holds, pursuing a regression model for the conditional mean of Y is needless. As mentioned in Cook et al. (2002) and Park et al. (2015), many regression analysis problems are concerned with the model for the conditional mean of response given predictors. Furthermore, the theory of conditional mean dependence has rich uses in various statistical problems, such as variable screening (Shao and Zhang, 2014), model checking (Su and Zheng, 2017), and the applications in Lee and Shao (2018), Liu et al. (2019), and Zhang et al. (2018).

* Corresponding author.
E-mail address: zzhang@bjut.edu.cn (Z. Zhang).

https://doi.org/10.1016/j.csda.2021.107246
0167-9473/

The preceding test problem has been investigated by Shao and Zhang (2014) for $Y \in \mathbb{R}^1$ and $X \in \mathbb{R}^q$. Motivated by the distance covariance (Székely et al., 2007), they proposed the so-called martingale difference divergence (MDD) to measure the conditional mean dependence of Y given X. Park et al. (2015) generalized MDD to the setting $Y \in \mathbb{R}^p$ and $X \in \mathbb{R}^q$ and proposed partial martingale difference correlation to deal with the problem concerning $E(Y|X, Z) = E(Y|Z)$ for some additional predictor Z. Lee et al. (2020) proposed functional martingale difference divergence (FMDD), which extended MDD to the functional setting where X and Y can be functional variables. Zhang et al. (2018) employed the idea of MDD to a high dimension scenario. The MDD method is shown powerful in many cases, and has many applications, see also Su and Zheng (2017); Lee and Shao (2018); Liu et al. (2019).

As shown in Zhang et al. (2018), for $X = (x_1, \ldots, x_p)^{\mathrm{T}} \in \mathbb{R}^p$ and $Y \in \mathbb{R}$, the martingale difference divergence can be approximated using pairwise covariances when p is large,

$$\mathrm{MDD}(Y|X)^2 \approx \frac{1}{\sqrt{\tau}} \sum_{j=1}^{p} \mathrm{cov}^2(Y, x_j),$$

where $\tau = E|X - X'|^2$ and X' is an independent copy of X. Since the covariance only captures the linear dependence, the martingale difference divergence may have less power when it is employed to detect nonlinear relationships, especially in the cases of high dimensions. In fact, we may observe the phenomenon in the numerical results in Section 4. This property is similar to the distance covariance (Székely et al., 2007). As discussed in Shen et al. (2020), for testing dependence between two random vectors, the distance covariance has wonderful performance for monotone relationships, but may lose power for nonlinear dependencies; while the Hilbert-Schmidt independence criterion (HSIC) has opposite performance, that is, HSIC is slightly inferior to the distance covariance for monotone relationships but excel for various nonlinear dependencies. Noting that HSIC is a kernel-based method, this inspires us to seek a kernel-based method for conditional mean dependence.

We propose using the Hilbert-Schmidt norm of some operator to characterize the conditional mean dependence of Y given X with the following steps. First, the condition of H_0 is converted equivalently to a condition expressed with a family of finite signed measures on $\mathcal{X}$. Then, the finite signed measures are embedded into a reproducing kernel Hilbert space $\mathcal{H}$ consisting of functionals on $\mathcal{X}$, with a chosen kernel. Finally, a tensor operator from $\mathcal{H}$ to $\mathcal{Y}$ is constructed and its Hilbert-Schmidt norm is used as a measure for the conditional mean dependence. We call it kernel-based conditional mean dependence (KCMD). Then, we construct a test based on KCMD to the hypothesis (1). It is noteworthy that the proposed measure enjoys the following properties. First, the kernel-based conditional mean dependence has a simple expectation form, which enables us to construct an unbiased empirical estimator easily. Second, when the chosen kernel is characteristic, the KCMD is nonnegative and equals to zero if and only if the null hypothesis holds. Third, since one can choose different kernels for different data, our method has wider applications, compared to the martingale different divergence. Fourth, the estimated KCMD is feasible and has the form of computationally attractive U-statistic, which inherits the classical asymptotic properties of U-statistics. Last but not least, it is worth to point out that the functional martingale difference divergence proposed by Lee et al. (2020) is a special case of KCMD corresponding to the distance kernel described in Remark 1. Furthermore, this tool can be applied to variable selection and model checking, as shown in Section 5, as well as other applications mentioned in Section 6. Through simulations, we find that the tests based on the KCMD measure have close performance for monotone relationships as that based on FMDD, but are superior for nonlinear relationships and relationships that the moment restriction $E(|X|) < \infty$ is violated.

The rest of the paper is organized as follows. In Section 2, we define the kernel-based conditional mean dependence by using kernel embedding method and derive its properties. Then we construct an unbiased estimator of KCMD and give its asymptotic distributions under null hypothesis and fixed alternatives respectively. In Section 3, we construct a test statistic based on the estimated KCMD for the conditional mean independence of Y given X and derive its asymptotic properties under null hypothesis, fixed alternatives, and local alternatives. Then we provide a wild bootstrap procedure to carry out the test in practice and give some criterion to choose a suitable kernel. In Section 4, we explore the finite sample performance of the proposed tests by comparing with existing competing methods through simulations. In Section 5 we illustrate the applications of the proposed method in two real data examples. A discussion is given in Section 6. All the technical proof details are deferred to the appendix.

2. Kernel-based conditional mean dependence

2.1. Notations and formulation

Let $\mathcal{X}, \mathcal{Y}$ be separable Hilbert spaces with inner products $\langle \cdot, \cdot \rangle_{\mathcal{Z}}$ and norms $\| \cdot \|_{\mathcal{Z}}$ respectively, and $\mathcal{B}(\mathcal{Z})$ be the σ-field generated by all open subsets of $\mathcal{Z}$, where $\mathcal{Z} = \mathcal{X}$ or $\mathcal{Y}$. Suppose X and Y are two nondegenerate random elements taking values from $\mathcal{X}$ and $\mathcal{Y}$, respectively; their joint probability measure is P_{XY}, which is defined on $\mathcal{X} \times \mathcal{Y}$, and the marginal distributions of X and Y are P_X and P_Y respectively. We always assume $E(Y)$ exists.

Suppose that $\mathcal{H}$ is the reproducing kernel Hilbert space (RKHS) of functions from $\mathcal{X}$ to $\mathbb{R}$ associated with a characteristic kernel $k(\cdot, \cdot)$. The inner product on $\mathcal{H}$ is denoted by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. Then $k(\cdot, t) \in \mathcal{H}$ and $f(t) = \langle f, k(\cdot, t) \rangle_{\mathcal{H}}$ for every $t \in \mathcal{X}$, $f \in \mathcal{H}$. For $y \in \mathcal{Y}$ and $h \in \mathcal{H}$, let $y \otimes h$ be the tensor product defined by $y \otimes h(y') = \langle y, y' \rangle h$ for $y' \in \mathcal{Y}$, which is a linear operator from $\mathcal{Y}$ to $\mathcal{H}$. Then the Hilbert-Schmidt norm of $y \otimes h$ is the square root of

$$\|y \otimes h\|_{HS}^2 = \langle y \otimes h, y \otimes h \rangle_{HS} = \langle y, y \rangle_{\mathcal{Y}} \langle h, h \rangle_{\mathcal{H}}.$$

To see more details on the Hilbert-Schmidt norm, one can refer to Gretton et al. (2005).

2.2. Representation of the conditional mean dependence

Now we introduce the measure of conditional mean dependence. The null hypothesis $H_0 : E(Y|X) = E(Y)$, almost surely, is equivalent to

$$E(\langle Y - E(Y), g \rangle_{\mathcal{Y}} I_B(X)) = 0 \text{ for all } B \in \mathcal{B}(\mathcal{X}), g \in \mathcal{Y},$$

where $I_B(\cdot)$ is the indicator function of B. Denote $\mu = E(Y)$. For any $g \in \mathcal{Y}$, define the signed measure

$$\lambda_g(B) = \int \langle y - \mu, g \rangle_{\mathcal{Y}} I_B(x) dP_{XY}(x, y) = E[\langle Y - \mu, g \rangle_{\mathcal{Y}} I_B(X)], \quad B \in \mathcal{B}(\mathcal{X}),$$

then the measure λ_g is finite for fixed g. Denote by $\mathcal{M}$ the set of all finite signed Borel measures on $\mathcal{B}(\mathcal{X})$, and define an embedding map $\Pi : \mathcal{M} \mapsto \mathcal{H}$,

$$\Pi(\nu) = \int k(\cdot, x) d\nu(x), \ \nu \in \mathcal{M}.$$

Then, we have

$$\Pi(\lambda_g) = \int \langle y - \mu, g \rangle_{\mathcal{Y}} k(\cdot, x) dP_{XY}(x, y) := \int (y - \mu) \otimes k(\cdot, x) dP_{XY}(x, y)(g), \tag{2}$$

here $\int (y - \mu) \otimes k(\cdot, x) dP_{XY}(x, y)$ (Bochner integral) defines a linear operator from $\mathcal{Y}$ to $\mathcal{H}$. The equation (2) can be verified along the routine beginning from indicator functions, to simple functions, and then integrable functions. Since the kernel is characteristic, Π is an injective map, and thus it follows that

$$\lambda_g(B) = E[\langle Y - \mu, g \rangle_{\mathcal{Y}} I_B(X)] = 0, \text{ for all } B \in \mathcal{B}(\mathcal{X}), g \in \mathcal{Y}$$

if and only if

$$\int (y - \mu) \otimes k(\cdot, x) dP_{XY}(x, y) = \mathbf{0},$$

where $\mathbf{0}$ denotes the zero linear operator. This relationship indicates that we can use quantity

$$\| \int (y - \mu) \otimes k(\cdot, x) dP_{XY}(x, y) \|_{HS}^2$$

to measure the conditional mean dependence. By straightforward calculations,

$$\begin{aligned}
&\| \int (y - \mu) \otimes k(\cdot, x) dP_{XY}(x, y) \|_{HS}^2 \\
=&E\{\langle Y - \mu, Y' - \mu \rangle_{\mathcal{Y}} \langle k(\cdot, X), k(\cdot, X') \rangle_{\mathcal{H}}\} \\
=&E\{k(X, X') \langle Y - \mu, Y' - \mu \rangle_{\mathcal{Y}}\},
\end{aligned}$$

where (X', Y') is an independent copy of (X, Y) and the last equation follows from that k is the kernel of $\mathcal{H}$. For simplicity, we denote $\langle \cdot, \cdot \rangle_{\mathcal{Y}} = \langle \cdot, \cdot \rangle$ and $\| \cdot \|_{\mathcal{Y}} = \| \cdot \|$.

Definition 1 *(Kernel-based conditional mean dependence).* For the two random elements X and Y and a given positive definite kernel $k(\cdot, \cdot)$ on $\mathcal{X} \times \mathcal{X}$, we define the kernel-based conditional mean dependence of Y on X as

$$\text{KCMD}(Y, X) = E\{k(X, X') \langle Y - \mu, Y' - \mu \rangle\}, \tag{3}$$

(assuming the expectation exists), where $\mu = E(Y)$, (X', Y') is an independent copy of (X, Y).

Remark 1. The FMDD in Lee et al. (2020) is the special case of KCMD corresponding to the distance-induced kernel

$$k_d(x, x') = \|x - x_0\|_{\mathcal{X}} + \|x' - x_0\|_{\mathcal{X}} - \|x - x'\|_{\mathcal{X}},$$

where x_0 is a fixed point in $\mathcal{X}$. To see this, replacing the kernel k in (3) with k_d, we have

$$\begin{aligned}
\mathrm{KCMD}(Y,X) &= E\{\|X-x_0\|_{\mathcal{X}}\langle Y-\mu, Y'-\mu\rangle\} + E\{\|X'-x_0\|_{\mathcal{X}}\langle Y-\mu, Y'-\mu\rangle\} \\
&\quad - E\{\|X-X'\|_{\mathcal{X}}\langle Y-\mu, Y'-\mu\rangle\} \\
&= E\{\|X-x_0\|_{\mathcal{X}}\langle Y-\mu, EY'-\mu\rangle\} + E\{\|X'-x_0\|_{\mathcal{X}}\langle EY-\mu, Y'-\mu\rangle\} \\
&\quad - E\{\|X-X'\|_{\mathcal{X}}\langle Y-\mu, Y'-\mu\rangle\} \\
&= -E\{\|X-X'\|_{\mathcal{X}}\langle Y-\mu, Y'-\mu\rangle\},
\end{aligned}$$

which is just the FMDD. The positive definite property of this kernel is verified in Sejdinovic et al. (2013).

By the preceding arguments, we have the following proposition, which shows that the KCMD fully characterizes the conditional mean independence.

Proposition 1. *Suppose that $k(\cdot,\cdot)$ is a bounded, positive definite and characteristic kernel. Then* $\mathrm{KCMD}(Y,X)$ *is well defined, and*

(i) $\mathrm{KCMD}(Y,X) \geq 0$*;*
(ii) $\mathrm{KCMD}(Y,X) = 0$ *if and only if $E(Y|X) = E(Y)$, almost surely.*

Remark 2. We do not require the kernel in Definition 1 is characteristic. However, by Proposition 1(ii), we should choose a characteristic kernel in general. There are many characteristic kernels that have been studied, such as the Gaussian kernel $k(x,y) = \exp(-\sigma^{-1}\|x-y\|^2)$, the Laplace kernel $k(x,y) = \exp(-\sigma^{-1}\|x-y\|)$, $\sigma > 0$. For more details of characteristic kernels, see Fukumizu et al. (2009), Sriperumbudur et al. (2008) and Sriperumbudur et al. (2010). We will use the Gaussian and Laplace kernels in our simulation studies.

Denote

$$\begin{aligned}
\varphi(x,x') &= k(x,x') - E\{k(x,X')\} - E\{k(X,x')\} + E\{k(X,X')\}, \\
\psi(y,y') &= \langle y-E(Y), y'-E(Y')\rangle = \langle y,y'\rangle - \langle y, E(Y')\rangle - \langle y', E(Y)\rangle + \langle E(Y), E(Y')\rangle,
\end{aligned}$$

then we have another expression of the KCMD, which is useful to construct its estimators.

Proposition 2. *Suppose that (X',Y') is an independent copy of (X,Y), then under the conditions of Proposition 1, it holds that*

(i) $\mathrm{KCMD}(Y,X) = E[\varphi(X,X')\psi(Y,Y')]$*;*
(ii) $\mathrm{KCMD}(Y,X) \leq \sqrt{E[\varphi^2(X,X')]E[\psi^2(Y,Y')]}$.

2.3. Empirical estimators and asymptotic properties

We now give an empirical estimator of $\mathrm{KCMD}(Y,X)$. Given an independent identically distributed (i.i.d.) sample $(X_i,Y_i)_{i=1}^n$ from the joint distribution of (X,Y), a simple and straightforward estimator for $\mathrm{KCMD}(Y,X)$ is $\frac{1}{n^2}\sum_{i,j=1}^n k(X_i, X_j)\langle Y_i-\overline{Y}, Y_j-\overline{Y}\rangle$, where $\overline{Y} = n^{-1}\sum_{i=1}^n Y_i$, but it is biased and doesn't perform well compared to the following unbiased estimator when testing conditional mean independence in simulations. Inspired by Park et al. (2015) and Lee et al. (2020), we build an unbiased estimator of $\mathrm{KCMD}(Y,X)$ by using the $\mathcal{U}$-centering approach. See also Székely et al. (2014).

Denote

$$\begin{aligned}
c_{ij} &= \begin{cases} k(X_i,X_j), & i\neq j \\ 0, & i=j \end{cases}, \quad d_{ij} = \begin{cases} \langle Y_i,Y_j\rangle, & i\neq j \\ 0, & i=j \end{cases}, \\
c_{i\cdot} &= \frac{1}{n-2}\sum_{s=1}^n c_{is}, \quad c_{\cdot j} = \frac{1}{n-2}\sum_{t=1}^n c_{tj}, \quad c_{\cdot\cdot} = \frac{1}{(n-1)(n-2)}\sum_{s,t=1}^n c_{st}, \\
d_{i\cdot} &= \frac{1}{n-2}\sum_{s=1}^n d_{is}, \quad d_{\cdot j} = \frac{1}{n-2}\sum_{t=1}^n d_{tj}, \quad d_{\cdot\cdot} = \frac{1}{(n-1)(n-2)}\sum_{s,t=1}^n d_{st}, \\
C_{ij} &= \begin{cases} c_{ij}-c_{i\cdot}-c_{\cdot j}+c_{\cdot\cdot}, & i\neq j \\ 0, & i=j \end{cases}, \quad D_{ij} = \begin{cases} d_{ij}-d_{i\cdot}-d_{\cdot j}+d_{\cdot\cdot}, & i\neq j \\ 0, & i=j \end{cases}.
\end{aligned}$$

C_{ij} is the $\mathcal{U}$-centering version of c_{ij}. Then an estimator of $\mathrm{KCMD}(Y,X)$ is defined as

$$\mathrm{KCMD}_n(Y,X)=\frac{1}{n(n-3)}\sum_{i\neq j}C_{ij}D_{ij}. \tag{4}$$

With Proposition 2, it can be shown that the estimator (4) is unbiased and admits a U-statistic expression

$$\mathrm{KCMD}_n(Y,X)=\frac{1}{\binom{n}{4}}\sum_{i<j<s<t}h(Z_i,Z_i,Z_s,Z_t), \tag{5}$$

with the kernel

$$h(Z_i,Z_j,Z_s,Z_t)=\frac{1}{4!}\sum_{(u,v,q,r)}^{(i,j,s,t)}(c_{uv}d_{uv}-c_{uv}d_{uq}-c_{uv}d_{vr}+c_{uv}d_{qr}), \tag{6}$$

where $Z_i=(X_i,Y_i)$ and the sum is over all 4! permutations of (i,j,s,t).

Proposition 3. $E[\mathrm{KCMD}_n(Y,X)]=\mathrm{KCMD}(Y,X)$.

By the theory of U-statistics, we have the following theorems.

Theorem 1. *Suppose that $k(\cdot,\cdot)$ is a bounded, positive definite and characteristic kernel, and $E\|Y\|<\infty$. Then*

$$\mathrm{KCMD}_n(Y,X)\xrightarrow{a.s}\mathrm{KCMD}(Y,X).$$

The following Theorem 2 and Theorem 3 reveal the asymptotic distributions of $\mathrm{KCMD}_n(Y,X)$ under the null and alternative hypothesis respectively.

Theorem 2. *Suppose that $k(\cdot,\cdot)$ is a bounded, positive definite and characteristic kernel, and $E\|Y\|^2<\infty$. Then, under H_0, it holds*

$$n\mathrm{KCMD}_n(Y,X)\xrightarrow{d}\sum_{i=1}^{\infty}\gamma_i(N_i^2-1),$$

where N_is are i.i.d. standard normal distributed random variables, and $(\gamma_i)_{i=1}^{\infty}$ is the sequence of eigenvalues of the operator Γ with the kernel $\Gamma(z,z')=\varphi(x,x')\psi(y,y')$, $z=(x,y)$, $z'=(x',y')$, that is,

$$\int\Gamma(z,z')\phi_i(z')dP_{XY}(z')=\gamma_i\phi_i(z),$$

where $\phi_i(\cdot)$s are the corresponding orthonormal eigenfunctions of the operator.

Theorem 3. *Suppose that $k(\cdot,\cdot)$ is a bounded, positive definite and characteristic kernel, and $E\|Y\|^2<\infty$. Then, under H_1, it holds*

$$\sqrt{n}\{\mathrm{KCMD}_n(Y,X)-\mathrm{KCMD}(Y,X)\}\xrightarrow{d}N(0,4\sigma^2),$$

where $\sigma^2=\mathrm{var}\{\Lambda(Z)\}$ with $Z=(X,Y)$ and $\Lambda(z)=E\{\Gamma(z,Z)\}$.

3. Test of conditional mean independence

3.1. Test statistic and local asymptotic properties

Theorems 2 and 3 suggest that an appropriate statistic for the test of hypothesis (1) is

$$S_n=n\mathrm{KCMD}_n(Y,X),$$

and H_0 will be rejected for its large values. This test is obviously consistent since S_n converges in distribution under H_0 and goes in probability to infinity under fixed alternatives.

In order to understand the power of the test furthermore, we consider the behavior of the test statistic S_n under the local alternative $H_{1,n}:Y=E(Y)+n^{-\beta}m(X)+\varepsilon$, $\beta>0$, where $m(\cdot)$ is a regression function defined on $\mathcal{X}$ satisfying $E\{m(X)\}=0$, $\mathrm{KCMD}(m(X),X)>0$ and ε is a nondegenerate random error with $E(\varepsilon|X)=0$. Given these conditions, we have the following theorem.

Theorem 4. *Suppose that $k(\cdot,\cdot)$ is a bounded, positive definite and characteristic kernel, and $E(|m(X)|^2+|\varepsilon|^2)<\infty$. Then, the following facts hold under $H_{1,n}$.*

(i) *If* $\beta > 1/2$,

$$S_n \xrightarrow{d} \sum_{i=1}^{\infty} \gamma_i'(N_i^2 - 1),$$

where γ_i' *is defined similar to* γ_i *in Theorem 2 except that* Y *is replaced with* ε.

(ii) *If* $\beta = 1/2$,

$$S_n \xrightarrow{d} c + V + \sum_{i=1}^{\infty} \gamma_i'(N_i^2 - 1),$$

where $c = \mathrm{KCMD}(m(X), X) > 0$ *and* V *is distributed as* $N(0, \tau^2)$ *with* $\tau^2 = 4\mathrm{var}\{\Delta(\mathcal{Z})\}$, $\mathcal{Z} = (X, \varepsilon)$, $\Delta(z) = E\{\varphi(x, X)\psi(y, m(X)\}$, *and* $z = (x, y)$.

(iii) *If* $0 < \beta < 1/2$,

$$S_n \xrightarrow{p} \infty.$$

Remark 3. Theorem 2 in Lee et al. (2020) reveals similar limiting behaviors for their U-estimator of FMDD. Our test statistic has the same convergence rate as the FMDD statistic. This suggests that, in the sense of limiting behaviors, the test based on the KCMD can be as powerful as that based on the FMDD for the discrepancy rate slower than $n^{-1/2}$, but its power for the discrepancy rate of $n^{-1/2}$ is different, since its asymptotic distribution is not the same as that of the FMDD test statistic.

Theorem 4 indicates that when $0 < \beta < 1/2$ the test has asymptotic power 1; when $\beta > 1/2$ the test has asymptotic power α, the nominal significance level; when $\beta = 1/2$ the test has a nontrivial asymptotic power between α and 1. Therefore, a reasonable method for comparing two sequences of tests asymptotically is to consider the local limiting power when $\beta = 1/2$. Based on this, we develop a skill to choose kernel for tests based on the KCMD, which is discussed in Section 3.3.

Since the γ_is in Theorem 2 are unknown, the asymptotic distribution can not be used to compute the critical value of the proposed test. In practice, we can use a bootstrap procedure to approximate the asymptotic null distribution of S_n. We introduce the procedure in the next subsection.

3.2. Wild bootstrap test

To give an approximated critical value of the test, we suggest the following wild bootstrap method.

Algorithm 1 *(Wild bootstrap).*

(1) Generate an i.i.d. sequence $(\zeta_i^{(b)})_{i=1}^n$ of random variables with mean zero and variance 1. Compute the bootstrap statistic

$$\mathrm{KCMD}_n^*(Y, X)^{(b)} = \frac{1}{n(n-3)} \sum_{i \neq j} \zeta_i^{(b)} C_{ij} D_{ij} \zeta_j^{(b)}.$$

(2) Repeat step (1) for B times and collect data $(S_{n,b}^*)_{b=1}^B$, where $S_{n,b}^* = n\mathrm{KCMD}_n^*(Y, X)^{(b)}$.
(3) For significance level α, compute the $(1-\alpha)$th quantile of $(S_{n,b}^*)_{b=1}^B$, denoted by $Q_{(1-\alpha),n}^*$.
(4) If $S_n > Q_{(1-\alpha),n}^*$, reject H_0.

The wild bootstrap method has been employed extensively, see Lee et al. (2020) as a recent reference. The following theorems illustrate the asymptotic behaviors of the bootstrap test statistic under the null hypothesis, fixed alternatives and local alternatives, which show that the bootstrap test is consistent for any fixed alternative and local alternatives with the discrepancy rate $n^{-\beta}$, $0 < \beta < 1/2$.

Theorem 5. *Suppose that* $k(\cdot, \cdot)$ *is a bounded, positive definite and characteristic kernel,* $E\|Y\|^4 < \infty$ *and* $E|\zeta|^4 < \infty$. *Then, under* H_0, *we have*

$$n\mathrm{KCMD}_n^*(Y, X) \xrightarrow{d^*} \sum_{i=1}^{\infty} \gamma_i(N_i^2 - 1) \; a.s.,$$

where γ_i, N_i *are defined in Theorem 2 and the notation* $S_n^* \xrightarrow{d^*} S$ *a.s. means* $(S_n^*|Z_1, Z_2, \ldots)$ *converges in distribution to* $(S|Z_1, Z_2, \ldots)$ *for almost every sequence* $(Z_1, Z_2, \ldots)$.

Theorem 6. *Suppose that $k(\cdot,\cdot)$ is a bounded, positive definite and characteristic kernel, $E\|Y\|^4<\infty$ and $E|\zeta|^4<\infty$. Then, under H_1, we have*

$$\mathrm{pr}(S_n > Q^*_{1-\alpha}) \to 1.$$

Theorem 7. *Suppose that $k(\cdot,\cdot)$ is a bounded, positive definite and characteristic kernel, $E(|m(X)|^4+|\varepsilon|^4)<\infty$ and $E|\zeta|^4<\infty$. Then, under $H_{1,n}$, we have the following conclusions.*

(i) *If $\beta>1/2$,*

$$\mathrm{pr}(S_n > Q^*_{1-\alpha}) \to \alpha.$$

(ii) *If $\beta=1/2$,*

$$\mathrm{pr}(S_n > Q^*_{1-\alpha}) \to 1-G(q),$$

where G is the distribution function of $c+V+\sum_{i=1}^{\infty}\gamma_i'(N_i^2-1)$, which is the limiting distribution in Theorem 4 (ii), and q is the $(1-\alpha)$th quantile of null limiting distribution.

(iii) *If $0<\beta<1/2$,*

$$\mathrm{pr}(S_n > Q^*_{1-\alpha}) \to 1.$$

Remark 4. Theorem 5, 6 and 7 are paralleled to Theorem 5 and 6 in Lee et al. (2020). It is worth noting that the conditions assumed in our theorems are less strict. For example, in Theorem 5 in Lee et al. (2020), the conditions include $E(\|X\|^4+\|Y\|^8)<\infty$, while the corresponding part in our Theorem 5 only requires that $E(\|Y\|^4)<\infty$. This indicates that the proposed test has a greater scope of applications. For instance, consider the true regression model $Y=5\cos(X)+\varepsilon$, where X has a Cauchy distribution and ε is from the standard normal distribution. In this case, the condition $E\|X\|^4<\infty$ is not satisfied, the FMDD test may suffer from the loss of power, whereas the conditions in above theorems still hold when k is bounded. See the simulation results of Example 3 (iv) and (v) in Section 4.1.

3.3. The choice of kernel

The choice of the kernel influences the practical performance of the KCMD tests because it determines how well certain type of regression relationship can be detected. From the argument of Section 2.2, one should first consider characteristic kernels, for example, Gaussian kernel $k(x,y)=\exp(-\sigma^{-1}\|x-y\|^2)$, or Laplace kernel $k(x,y)=\exp(-\sigma^{-1}\|x-y\|)$, $\sigma>0$. For the choice of bandwidth σ of the kernels, we can use median heuristic (Pfister et al., 2018), taking the median of $\{\|X_i-X_j\|^2 : i<j\}$ as σ for Gaussian kernel, or the median of $\{\|X_i-X_j\| : i<j\}$ as σ for Laplace kernel.

To make a choice from two kernels, such as Gaussian kernel k_1 and Laplace kernel k_2, we use the following strategy: (i) for each kernel, compute the critical value by Algorithm 1 and the test statistic, denoted as $Q^{*(i)}_{(1-\alpha),n}$ and $S_n^{(i)}$, $i=1,2$, respectively; (ii) compute quantity $\Omega_i=(S_n^{(i)}-Q^{*(i)}_{(1-\alpha),n})/S_n^{(i)}$, $i=1,2$, which is the relative distance between the test statistic and the critical value; (iii) choose the kernel which has the maximum value of Ω_i. The idea behind this thumb rule is that, under alternative hypothesis, if a kernel can detect the regression relationship well, it should reveal a large test statistic comparing to the critical value. Usually, this procedure will involve additional computation. One can combine this rule with the median heuristic. For example, first use the median heuristic to choose the bandwidth for Gaussian and Laplace kernels, then use this rule to choose one between Gaussian kernel and Laplace kernel.

In practice, we do not always have to choose a characteristic kernel. It may be possible and potentially beneficial to consider some non-characteristic kernels when they are particularly powerful in detecting certain types of regression relationships.

4. Simulations

In this section, we explore the finite sample performance of tests based on KCMD. To examine the performance of different kernels, we use Gaussian kernel $k(x,y)=\exp(-\sigma^{-1}\|x-y\|^2)$ and Laplace kernel $k(x,y)=\exp(-\sigma^{-1}\|x-y\|)$ $(\sigma>0)$ in the simulations. Let $\mathrm{KCMD_{G,a}}$ be the test based on the KCMD corresponding to the Gaussian kernel with bandwidth $\sigma=a$, $\mathrm{KCMD_{G,m}}$ the test based on the KCMD corresponding to the Gaussian kernel with bandwidth determined by median heuristic, and $\mathrm{KCMD_{L,a}}$, $\mathrm{KCMD_{L,m}}$ be the quantities defined similarly for Laplace kernel. We also implement the kernel choosing rule described in Section 3.3 to make a choice between two given kernels, and denote the resulting test as $\mathrm{KCMD_c}$. Since the FMDD is a special case of KCMD, we compare our method to the FMDD test (FMDD), which is implemented by the wild bootstrap method described in Lee et al. (2020). We set sample size from 50 to 100, repeat each setting 2000

Table 1
Empirical sizes of the tests based on KCMD and FMDD in Example 1 with $n = 50, 100$.

Scenario	Test	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
		$n = 50$	$n = 100$	$n = 50$	$n = 100$	$n = 50$	$n = 100$
(i)	$\text{KCMD}_{\text{G,m}}$	0.012	0.013	0.060	0.051	0.104	0.104
	$\text{KCMD}_{\text{L,m}}$	0.011	0.011	0.064	0.056	0.106	0.104
	FMDD	0.011	0.011	0.060	0.054	0.104	0.103
(ii)	$\text{KCMD}_{\text{G,m}}$	0.015	0.011	0.064	0.050	0.116	0.097
	$\text{KCMD}_{\text{L,m}}$	0.013	0.014	0.065	0.050	0.118	0.098
	FMDD	0.013	0.017	0.061	0.053	0.117	0.097
(iii)	$\text{KCMD}_{\text{G,m}}$	0.009	0.015	0.053	0.051	0.112	0.093
	$\text{KCMD}_{\text{L,m}}$	0.010	0.015	0.056	0.051	0.114	0.095
	FMDD	0.009	0.013	0.058	0.057	0.108	0.099
(iv)	$\text{KCMD}_{\text{G,m}}$	0.013	0.008	0.063	0.049	0.115	0.101
	$\text{KCMD}_{\text{L,m}}$	0.013	0.006	0.062	0.052	0.121	0.105
	FMDD	0.011	0.011	0.060	0.051	0.118	0.108

times, and report the size and power performance of the respective tests at significance levels $\alpha = 0.01$, 0.05 and 0.10. For each example, the bootstrap sample size is equal to $B = 500$ and the $(\zeta_i)_{i=1}^n$ are generated from the distribution

$$\text{pr}\left(\zeta_i = \frac{-(\sqrt{5}-1)}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}}, \quad \text{pr}\left(\zeta_i = \frac{(\sqrt{5}+1)}{2}\right) = 1 - \frac{\sqrt{5}+1}{2\sqrt{5}}.$$

This distribution is also used in Patilea et al. (2016) and Lee et al. (2020). Both KCMD and FMDD can be used for vector data and functional data. We present multivariate examples in the following, and put functional data examples and comparisons with other test methods in the next subsection.

4.1. Multivariate data

Example 1. Generate the i.i.d. sample $(X_i, Y_i)_{i=1}^n$ of (X, Y) from the following models. In the following scenarios, $\mathbf{I}_p$ denotes the $p \times p$ identity matrix.

(i) X and Y are independent random vectors from the multivariate standard normal distribution $N(\mathbf{0}, \mathbf{I}_5)$.
(ii) Let $\mathbf{\Sigma} = (\sigma_{ij}) \in \mathbb{R}^{5\times5}$ with $\sigma_{ij} = 0.7^{|i-j|}$, and

$$\begin{aligned} &X = (Z_1, \dots, Z_5)^{\mathrm{T}} \sim N(\mathbf{0}, \mathbf{\Sigma}), \\ &\epsilon = (\epsilon_1, \dots, \epsilon_5)^{\mathrm{T}} \sim N(\mathbf{0}, \mathbf{I}_5), \\ &V_k \sim U(0, 1), k = 1, \dots, 5, \\ &Y = (V_1, \dots, V_5)^{\mathrm{T}} + (Z_1\epsilon_1, \dots, Z_5\epsilon_5)^{\mathrm{T}}, \end{aligned}$$

where $U(0, 1)$ is the uniform distribution on interval $[0, 1]$.
(iii) $Z_1, \dots, Z_5$ are independent variables from the binomial distribution $B(10, 0.5)$, and $X = Z_1 + Z_2 + Z_3^2$, $Y = Z_4 + Z_5^2$.
(iv) Let $\mathbf{\Sigma} = (\sigma_{ij}) \in \mathbb{R}^{10\times10}$ with $\sigma_{ij} = 0.7^{|i-j|}$, and

$$\begin{aligned} &X = (Z_1, \dots, Z_{10})^{\mathrm{T}} \sim N(\mathbf{0}, \mathbf{\Sigma}), \\ &\epsilon = (\epsilon_1, \dots, \epsilon_{10})^{\mathrm{T}} \sim N(\mathbf{0}, \mathbf{I}_{10}), \\ &Y = (Z_1\epsilon_1, \dots, Z_{10}\epsilon_{10})^{\mathrm{T}}. \end{aligned}$$

In this example, the null hypothesis holds in all scenarios. In scenarios (i) and (iii) of Example 1, X and Y are independent while in other scenarios X and Y are not independent. Table 1 lists the empirical sizes of the tests at significance levels $\alpha = 0.01$, 0.05 and 0.10. From Table 1, the empirical sizes of all tests are close to the nominal significance levels in all scenarios.

Example 2. Generate the i.i.d. sample $(X_i, Y_i)_{i=1}^n$ of (X, Y) from the following models.
(i)

$$\begin{aligned} &X = (Z_1, \dots, Z_p)^{\mathrm{T}} \sim N(\mathbf{0}, \mathbf{I}_p), \\ &\epsilon \sim N(0, 1), \end{aligned}$$

Table 2
Empirical sizes and powers of the tests based on KCMD and FMDD for different rs in Example 2 with $p = 5$ and $n = 50, 100$.

Scenario	α	Test	$n = 50$	$n = 100$	$n = 50$	$n = 100$	$n = 50$	$n = 100$	$n = 50$	$n = 100$
(i)			$r = 0$		$r = 0.10$		$r = 0.20$		$r = 0.30$	
	0.01	$\mathrm{KCMD_{G,m}}$	0.015	0.012	0.034	0.049	0.124	0.290	0.338	0.701
		$\mathrm{KCMD_{L,m}}$	0.012	0.011	0.031	0.050	0.123	0.291	0.329	0.697
		FMDD	0.011	0.013	0.033	0.053	0.132	0.290	0.340	0.675
	0.05	$\mathrm{KCMD_{G,m}}$	0.058	0.062	0.117	0.154	0.289	0.526	0.572	0.856
		$\mathrm{KCMD_{L,m}}$	0.056	0.056	0.115	0.157	0.288	0.524	0.573	0.858
		FMDD	0.059	0.058	0.126	0.159	0.287	0.506	0.566	0.850
	0.10	$\mathrm{KCMD_{G,m}}$	0.129	0.106	0.195	0.252	0.417	0.637	0.697	0.928
		$\mathrm{KCMD_{L,m}}$	0.125	0.111	0.198	0.245	0.415	0.633	0.704	0.926
		FMDD	0.125	0.115	0.198	0.253	0.409	0.624	0.683	0.921
(ii)			$r = 0$		$r = 0.05$		$r = 0.10$		$r = 0.15$	
	0.01	$\mathrm{KCMD_{G,m}}$	0.017	0.013	0.042	0.093	0.200	0.576	0.483	0.917
		$\mathrm{KCMD_{L,m}}$	0.017	0.014	0.040	0.112	0.249	0.675	0.582	0.966
		FMDD	0.016	0.013	0.062	0.192	0.355	0.833	0.721	0.991
	0.05	$\mathrm{KCMD_{G,m}}$	0.055	0.058	0.141	0.236	0.421	0.796	0.747	0.982
		$\mathrm{KCMD_{L,m}}$	0.058	0.056	0.165	0.281	0.491	0.851	0.827	0.995
		FMDD	0.061	0.059	0.225	0.383	0.624	0.935	0.910	0.999
	0.10	$\mathrm{KCMD_{G,m}}$	0.117	0.120	0.235	0.346	0.573	0.865	0.855	0.994
		$\mathrm{KCMD_{L,m}}$	0.117	0.119	0.257	0.400	0.632	0.919	0.901	0.998
		FMDD	0.116	0.121	0.332	0.521	0.749	0.965	0.951	1.000
(iii)			$r = 0$		$r = 0.20$		$r = 0.40$		$r = 0.50$	
	0.01	$\mathrm{KCMD_{G,m}}$	0.009	0.011	0.025	0.056	0.108	0.465	0.183	0.768
		$\mathrm{KCMD_{L,m}}$	0.009	0.011	0.021	0.050	0.075	0.375	0.141	0.672
		FMDD	0.011	0.009	0.013	0.022	0.021	0.065	0.041	0.120
	0.05	$\mathrm{KCMD_{G,m}}$	0.055	0.052	0.120	0.206	0.344	0.781	0.526	0.947
		$\mathrm{KCMD_{L,m}}$	0.059	0.059	0.117	0.181	0.303	0.721	0.467	0.920
		FMDD	0.058	0.058	0.077	0.102	0.134	0.266	0.167	0.377
	0.10	$\mathrm{KCMD_{G,m}}$	0.113	0.110	0.229	0.324	0.487	0.878	0.697	0.981
		$\mathrm{KCMD_{L,m}}$	0.111	0.114	0.217	0.301	0.455	0.851	0.658	0.965
		FMDD	0.117	0.112	0.159	0.186	0.232	0.420	0.302	0.583
(iv)			$r = 0$		$r = 0.50$		$r = 1.00$		$r = 1.50$	
	0.01	$\mathrm{KCMD_{G,m}}$	0.016	0.013	0.093	0.430	0.497	0.992	0.770	1.000
		$\mathrm{KCMD_{L,m}}$	0.012	0.011	0.078	0.358	0.381	0.984	0.647	1.000
		FMDD	0.011	0.013	0.031	0.083	0.067	0.311	0.110	0.482
	0.05	$\mathrm{KCMD_{G,m}}$	0.056	0.051	0.316	0.721	0.839	0.999	0.976	1.000
		$\mathrm{KCMD_{L,m}}$	0.058	0.051	0.279	0.669	0.775	0.999	0.946	1.000
		FMDD	0.054	0.049	0.134	0.273	0.291	0.706	0.377	0.879
	0.10	$\mathrm{KCMD_{G,m}}$	0.118	0.103	0.459	0.836	0.912	1.000	0.993	1.000
		$\mathrm{KCMD_{L,m}}$	0.119	0.103	0.420	0.789	0.883	1.000	0.985	1.000
		FMDD	0.117	0.102	0.231	0.423	0.467	0.869	0.570	0.972

$$Y = r(Z_1 + \ldots + Z_p) + (Z_1 + Z_2)\epsilon,$$

where $r = 0, 0.10, 0.20, 0.30$.

(ii) Let $\boldsymbol{\Sigma} = (\sigma_{ij}) \in \mathbb{R}^{p\times p}$ with $\sigma_{ij} = 0.7^{|i-j|}$,

$$X = (Z_1, \ldots, Z_p)^{\mathrm{T}} \sim N(\mathbf{0}, \boldsymbol{\Sigma}),$$
$$\epsilon = (\epsilon_1, \ldots, \epsilon_p)^{\mathrm{T}} \sim N(\mathbf{0}, \mathbf{I}_p),$$
$$Y = rX^3 + \epsilon,$$

where $X^3 = (Z_1^3, \ldots, Z_p^3)^{\mathrm{T}}$ and $r = 0, 0.05, 0.10, 0.15$.

(iii) Denote $X^2 = (Z_1^2, \ldots, Z_p^2)^{\mathrm{T}}$ and $X \circ \epsilon = (Z_1\epsilon_1, \ldots, Z_p\epsilon_p)^{\mathrm{T}}$,

$$X = (Z_1 \ldots, Z_p)^{\mathrm{T}} \sim N(\mathbf{0}, \mathbf{I}_p),$$
$$\epsilon = (\epsilon_1, \ldots, \epsilon_p)^{\mathrm{T}} \sim N(\mathbf{0}, \mathbf{I}_p),$$
$$Y = rX^2 + X \circ \epsilon,$$

where $r = 0, 0.20, 0.40, 0.50$.

Table 3
Powers of the tests based on KCMD and FMDD for different ps in Example 2 with $n = 50, 100$.

Scenario	r	α	Test	$p=5$		$p=10$		$p=20$		$p=40$	
				$n=50$	$n=100$	$n=50$	$n=100$	$n=50$	$n=100$	$n=50$	$n=100$
(i)	0.2	0.01	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.141	0.274	0.228	0.554	0.336	0.785	0.424	0.913
			$\mathrm{KCMD}_{\mathrm{L,m}}$	0.128	0.282	0.228	0.555	0.344	0.779	0.421	0.916
			FMDD	0.141	0.278	0.223	0.539	0.333	0.768	0.412	0.917
		0.05	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.315	0.531	0.467	0.761	0.601	0.914	0.689	0.973
			$\mathrm{KCMD}_{\mathrm{L,m}}$	0.310	0.521	0.473	0.758	0.598	0.918	0.689	0.977
			FMDD	0.318	0.513	0.460	0.752	0.598	0.912	0.685	0.976
		0.10	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.429	0.659	0.592	0.861	0.721	0.962	0.799	0.989
			$\mathrm{KCMD}_{\mathrm{L,m}}$	0.427	0.650	0.593	0.863	0.720	0.963	0.803	0.990
			FMDD	0.430	0.643	0.601	0.854	0.714	0.960	0.799	0.989
(ii)	0.1	0.01	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.215	0.598	0.374	0.868	0.619	0.985	0.817	1.000
			$\mathrm{KCMD}_{\mathrm{L,m}}$	0.264	0.686	0.430	0.910	0.646	0.990	0.833	1.000
			FMDD	0.377	0.832	0.533	0.966	0.730	0.998	0.868	1.000
		0.05	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.423	0.797	0.630	0.951	0.834	0.999	0.942	1.000
			$\mathrm{KCMD}_{\mathrm{L,m}}$	0.483	0.860	0.682	0.968	0.859	1.000	0.947	1.000
			FMDD	0.619	0.943	0.766	0.989	0.903	1.000	0.958	1.000
		0.10	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.580	0.875	0.756	0.976	0.904	1.000	0.959	1.000
			$\mathrm{KCMD}_{\mathrm{L,m}}$	0.644	0.917	0.799	0.984	0.918	1.000	0.965	1.000
			FMDD	0.769	0.968	0.869	0.995	0.951	1.000	0.974	1.000
(iii)	0.4	0.01	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.739	1.000	0.321	0.967	0.117	0.569	0.047	0.195
			$\mathrm{KCMD}_{\mathrm{L,m}}$	0.645	1.000	0.279	0.947	0.112	0.526	0.051	0.185
			FMDD	0.100	0.512	0.061	0.246	0.038	0.118	0.025	0.054
		0.05	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.971	1.000	0.706	0.999	0.349	0.828	0.180	0.428
			$\mathrm{KCMD}_{\mathrm{L,m}}$	0.946	1.000	0.660	0.998	0.342	0.807	0.178	0.414
			FMDD	0.373	0.896	0.233	0.568	0.137	0.297	0.105	0.171
		0.10	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.996	1.000	0.855	1.000	0.502	0.921	0.303	0.581
			$\mathrm{KCMD}_{\mathrm{L,m}}$	0.991	1.000	0.816	0.999	0.482	0.905	0.301	0.572
			FMDD	0.576	0.975	0.398	0.748	0.236	0.473	0.194	0.288
(iv)	1.0	0.01	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.494	0.992	0.200	0.826	0.085	0.354	0.035	0.106
			$\mathrm{KCMD}_{\mathrm{L,m}}$	0.384	0.984	0.174	0.760	0.077	0.327	0.032	0.100
			FMDD	0.079	0.321	0.047	0.159	0.037	0.073	0.022	0.043
		0.05	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.839	0.999	0.519	0.960	0.278	0.641	0.154	0.291
			$\mathrm{KCMD}_{\mathrm{L,m}}$	0.768	0.999	0.460	0.942	0.266	0.607	0.157	0.282
			FMDD	0.278	0.710	0.179	0.452	0.123	0.225	0.109	0.118
		0.10	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.925	1.000	0.674	0.987	0.401	0.784	0.239	0.446
			$\mathrm{KCMD}_{\mathrm{L,m}}$	0.889	1.000	0.631	0.975	0.384	0.758	0.232	0.442
			FMDD	0.469	0.874	0.301	0.612	0.223	0.367	0.160	0.252

(iv) Denote $\cos(X) = (\cos(Z_1), \dots, \cos(Z_p))^{\mathrm{T}}$,

$$X = (Z_1, \dots, Z_p)^{\mathrm{T}} \sim N(\mathbf{0}, \mathbf{I}_p),$$
$$\epsilon = (\epsilon_1, \dots, \epsilon_p)^{\mathrm{T}} \sim N(\mathbf{0}, \mathbf{I}_p),$$
$$Y = r\{\cos(X) + X^2\} + \epsilon,$$

where $r = 0, 0.50, 1.00, 1.50$.

In this example, we first set $p = 5$ and take different values of r for the scenarios, where $r = 0$ indicates that the null hypothesis holds. Table 2 summarizes the empirical sizes and powers of the KCMD-based tests with Gaussian kernel, the KCMD-based tests with Laplace kernel and the FMDD-based test. In all scenarios, the empirical sizes of all the tests approximate the significance levels well. In terms of the power, as r increases or the sample size gets larger, the powers of all the tests increase. In scenarios (i) and (ii), where the relationships are monotone, the powers of the KCMD-based tests are close or slightly inferior to those of the FMDD-based test; while in scenarios (iii) and (iv), where the relationships are nonlinear, the KCMD-based tests are much more powerful than the FMDD-based test.

We then vary the dimension p from 5 to 40 and fix r at the third value in each scenario (i.e., r takes $0.2, 0.1, 0.4$ and 1.0 in scenarios (i), (ii), (iii) and (iv) respectively). The results are summarized in Table 3. In scenarios (i) and (ii), the powers of the tests increase as the dimension p increases; in scenarios (iii) and (iv), the powers of the tests diminish quickly as p increases, but the KCMD-based tests have higher power than the FMDD-based test. It seems that, for the monotone relationships considered, the increasing of dimension gives more information to the relationships, while for the nonlinear relationships, larger dimension makes it more difficult to detect the relationships.

Table 4
Powers of the tests based on KCMD and FMDD in Example 3 with $n = 50, 100$.

Scenario	Test	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
		$n = 50$	$n = 100$	$n = 50$	$n = 100$	$n = 50$	$n = 100$
(i)	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.337	0.895	0.712	0.989	0.834	0.998
	$\mathrm{KCMD}_{\mathrm{L,m}}$	0.345	0.918	0.731	0.992	0.858	0.999
	FMDD	0.155	0.483	0.410	0.807	0.553	0.918
(ii)	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.069	0.345	0.280	0.740	0.446	0.909
	$\mathrm{KCMD}_{\mathrm{L,m}}$	0.076	0.380	0.301	0.781	0.473	0.921
	FMDD	0.031	0.077	0.139	0.307	0.238	0.472
(iii)	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.241	0.344	0.496	0.529	0.678	0.695
	$\mathrm{KCMD}_{\mathrm{L,m}}$	0.280	0.398	0.529	0.581	0.722	0.740
	FMDD	0.254	0.449	0.559	0.640	0.732	0.782
(iv)	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.088	0.291	0.257	0.527	0.384	0.649
	$\mathrm{KCMD}_{\mathrm{L,m}}$	0.193	0.603	0.432	0.838	0.568	0.918
	FMDD	0.035	0.086	0.153	0.263	0.297	0.417
(v)	$\mathrm{KCMD}_{\mathrm{G,m}}$	0.074	0.126	0.412	0.511	0.984	0.987
	$\mathrm{KCMD}_{\mathrm{L,m}}$	0.056	0.106	0.377	0.474	0.982	0.985
	FMDD	0.002	0.002	0.068	0.073	0.522	0.614

Example 3. Generate the i.i.d. sample $(X_i, Y_i)_{i=1}^n$ of (X, Y) from the following models.

(i) $Z_1, \ldots, Z_4$ are independent variables from the binomial distribution $B(10, 0.5)$, and $X = (Z_1, \ldots, Z_4)^{\mathrm{T}}$, $Y = 0.02\sin(Z_1)(Z_2 + Z_3)^2 + 0.2\log(Z_4 + 1) + \varepsilon$, where $\varepsilon \sim N(0, 1)$.
(ii) $Z_1, \ldots, Z_5$ are independent variables from the uniform distribution $U(-1, 1)$, and $X = (Z_1, \ldots, Z_5)^{\mathrm{T}}$, $Y_1 = (Z_1 + Z_2)^2 + \varepsilon$ and $Y_2 = Z_3Z_4 + Z_4Z_5 + Z_3Z_5 + Z_3\varepsilon$, $Y = (Y_1, Y_2)^{\mathrm{T}}$, where $\varepsilon \sim N(0, 1)$.
(iii) $Z_1, \ldots, Z_5$ are independent variables from the t-distribution with 5 degrees of freedom, and $X = (Z_1, \ldots, Z_5)^{\mathrm{T}}$, $Y_1 = (Z_1 + Z_2)^2 + Z_3^3 + \varepsilon$ and $Y_2 = \cos(Z_4 + Z_5) + Z_3Z_5 + e^{Z_1+Z_2}\varepsilon$, $Y = (Y_1, Y_2)^{\mathrm{T}}$, where $\varepsilon \sim N(0, 1)$.
(iv) $Z_1, \ldots, Z_5$ are independent variables from the Cauchy distribution with location parameter 0 and scale parameter 1, and $X = (Z_1, \ldots, Z_5)^{\mathrm{T}}$, $Y = 5\cos(X) + \varepsilon$, where $\varepsilon \sim N(\mathbf{0}, \mathbf{I}_5)$.
(v) It is identical to (iv), except that $Y = 0.3X^2 + \varepsilon$.

In this example, we consider different kinds of distributions and relationships, and the alternative hypothesis holds in all scenarios. The simulation results are summarized in Table 4. In scenario (iii), the KCMD-tests are slightly worse than the FMDD-based test; while they are more powerful than the FMDD-based test in all other scenarios, especially in scenarios (ii), (iv) and (v), in which the relationships are complex and nonlinear or the moment restriction $E(|X|) < \infty$ required in FMDD is violated.

Example 4. This example is designed to examine whether the proposed kernel choosing rule can catch the higher power kernel from the given two kernels in the following three scenarios.

(i) The sample $(X_i, Y_i)_{i=1}^n$ is generated as that in Example 3(i). The kernels are taken as Gaussian kernels with bandwidths $a_1 = 1$ and $a_2 = 1/2$, respectively.
(ii) The sample $(X_i, Y_i)_{i=1}^n$ is generated as that in Example 3(ii). The kernels are taken as Laplace kernels with bandwidths $a_1 = 1$ and $a_2 = 1/2$, respectively.
(iii) The sample $(X_i, Y_i)_{i=1}^n$ is generated as that in Example 3(iii). The kernels are taken as Gaussian and Laplace kernels with bandwidth $a_1 = 1$ and $a_2 = 1/2$, respectively.

The simulation results are presented in Table 5. It appears that the thumb rule of choosing kernel can catch the higher power kernel in all the scenarios.

4.2. Functional data

In this subsection, we evaluate the performance of the proposed tests through three simulation examples of functional data. We compare the proposed tests with the FMDD test and the test developed in Patilea et al. (2016), denoted as PSS. The PSS test is implemented by the R package *fdapss* available at http://webspersoais.usc.es/persoais/cesar.sanchez/. In the simulations all the functional variables are defined on interval $[0, 1]$ and observed at 201 equally spaced points, and the Gaussian processes are generated by *rproc2fdata* function in *R* package *fda.usc*. For sake of space, we only report the results of the test corresponding to Laplace kernel. The performance of the test related to Gaussian kernel is similar.

Table 5
Powers of the tests based on KCMD with different kernels in Example 4 with $n = 50, 100$.

Scenario	Test	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
		$n = 50$	$n = 100$	$n = 50$	$n = 100$	$n = 50$	$n = 100$
(i)	$\mathrm{KCMD}_{\mathrm{G,a1}}$	0.397	0.969	0.856	0.998	0.924	1.000
	$\mathrm{KCMD}_{\mathrm{G,a2}}$	0.060	0.460	0.499	0.910	0.778	0.971
	$\mathrm{KCMD}_{\mathrm{C}}$	0.314	0.957	0.766	0.993	0.908	0.997
(ii)	$\mathrm{KCMD}_{\mathrm{L,a1}}$	0.716	0.998	0.953	1.000	0.980	1.000
	$\mathrm{KCMD}_{\mathrm{L,a2}}$	0.420	0.887	0.833	0.995	0.940	0.999
	$\mathrm{KCMD}_{\mathrm{C}}$	0.664	0.998	0.941	1.000	0.977	1.000
(iii)	$\mathrm{KCMD}_{\mathrm{G,a1}}$	0.397	0.972	0.840	0.999	0.930	1.000
	$\mathrm{KCMD}_{\mathrm{L,a2}}$	0.436	0.895	0.832	0.995	0.922	0.999
	$\mathrm{KCMD}_{\mathrm{C}}$	0.428	0.962	0.822	0.997	0.944	1.000

Example 5. We consider three scenarios in this example. Scenario (i) was used in Lei (2014) and Lee et al. (2020) and scenario (ii) comes from Lee et al. (2020).

(i) Let $\theta_j = r\bar{\theta}_j/||\bar{\theta}||_2$, with $\bar{\theta}_1 = 0.3$, $\bar{\theta}_j = 4(-1)^j/j^2$ for $j \geq 2$, and set $\Phi_1(t) = 1$, $\Phi_j(t) = \sqrt{2}\cos((j-1)\pi t)$. The data is generated by

$$X_i(t) = \sum_{j=1}^{100} j^{-1.1/2} X_{ij}\Phi_j(t), \quad \theta(t) = \sum_{j=1}^{100} \theta_j \Phi_j(t), \quad t \in [0, 1],$$
$$Y_i = \langle X_i, \theta\rangle + \varepsilon_i, \quad i = 1, \ldots, n,$$

with X_{ij}, ε_i being independent standard Gaussian. The parameter r^2 corresponds to the strength of the signal and is set to $r^2 = 0, 0.1, 0.2, 0.5$, as used in Lee et al. (2020).

(ii) The data is drawn from the quadratic regression model

$$Y_i = r\int_0^1 X_i^2(t)dt + \varepsilon_i,$$

where X_i and ε_i are defined in scenario (i), and $r = 0, 0.5, 1.0, 2.0$.

(iii) The sample is drawn from the model

$$Y_i(t) = r\cos(X_i(t)) + \varepsilon_i(t),$$

where X_i is defined in scenario (i), ε_is are generated independently from a Gaussian process with mean zero by $rproc2fdata$ function using the default setting, and $r = 0, 0.5, 1.0, 2.0$.

Example 6. Consider two scenarios in this example in which the responses are functional and the predictors are scalar. These scenarios were also adopted by Patilea et al. (2016) and Lee et al. (2020).

(i) The data $(X_i, Y_i)_{i=1}^n$ is generated as follows: X_i is drawn from the log-normal distribution with mean 3 and standard deviation 0.5, and

$$Y_i(t) = \mu(t) + \varepsilon_i(t),$$
$$\mu(t) = 0.01e^{-4(t-0.3)^2},$$

where ε_i is a Brownian bridge and independent of X_i. Obviously Y_i is independent of X_i.

(ii) The data $(X_i, Y_i)_{i=1}^n$ is generated by

$$Y_i(t) = \mu(t)X_i + \varepsilon_i(t),$$

here X_i and ε_i are generated as in scenario (i).

Example 7. Consider two scenarios in which both the responses and the predictors are functional variables. Scenario (i) was used in Patilea et al. (2016) and Lee et al. (2020).

Table 6
Empirical sizes and powers of the tests based on KCMD and FMDD for different rs in Example 5 with $n = 50, 100$.

Scenario	α	Test	$n=50$	$n=100$	$n=50$	$n=100$	$n=50$	$n=100$	$n=50$	$n=100$
(i)			$r^2=0$		$r^2=0.1$		$r^2=0.2$		$r^2=0.5$	
	0.01	$\text{KCMD}_{\text{L,m}}$	0.012	0.011	0.039	0.113	0.119	0.295	0.363	0.772
		FMDD	0.012	0.013	0.041	0.126	0.113	0.295	0.369	0.783
	0.05	$\text{KCMD}_{\text{L,m}}$	0.059	0.051	0.168	0.294	0.298	0.548	0.608	0.930
		FMDD	0.066	0.052	0.176	0.293	0.301	0.547	0.602	0.927
	0.10	$\text{KCMD}_{\text{L,m}}$	0.116	0.101	0.265	0.420	0.422	0.696	0.736	0.965
		FMDD	0.125	0.095	0.264	0.426	0.422	0.699	0.737	0.967
(ii)			$r=0$		$r=0.5$		$r=1.0$		$r=2.0$	
	0.01	$\text{KCMD}_{\text{L,m}}$	0.020	0.012	0.133	0.607	0.327	0.973	0.481	1.000
		FMDD	0.024	0.010	0.032	0.084	0.052	0.162	0.074	0.208
	0.05	$\text{KCMD}_{\text{L,m}}$	0.061	0.057	0.481	0.926	0.863	1.000	0.967	1.000
		FMDD	0.059	0.060	0.154	0.371	0.238	0.689	0.306	0.816
	0.10	$\text{KCMD}_{\text{L,m}}$	0.117	0.099	0.710	0.980	0.965	1.000	0.996	1.000
		FMDD	0.109	0.092	0.284	0.620	0.462	0.947	0.519	0.980
(iii)			$r=0$		$r=0.5$		$r=1.0$		$r=2.0$	
	0.01	$\text{KCMD}_{\text{L,m}}$	0.011	0.011	0.032	0.063	0.106	0.403	0.243	0.853
		FMDD	0.011	0.014	0.016	0.025	0.039	0.099	0.054	0.221
	0.05	$\text{KCMD}_{\text{L,m}}$	0.050	0.054	0.128	0.200	0.305	0.680	0.555	0.964
		FMDD	0.047	0.047	0.087	0.111	0.139	0.272	0.210	0.521
	0.10	$\text{KCMD}_{\text{L,m}}$	0.125	0.117	0.216	0.321	0.450	0.795	0.679	0.991
		FMDD	0.128	0.125	0.155	0.193	0.244	0.405	0.355	0.699

Table 7
The empirical sizes and powers of the tests based on KCMD, FMDD and PSS in Example 6 with $n = 50, 100$.

Scenario	Test	$\alpha=0.01$		$\alpha=0.05$		$\alpha=0.10$	
		$n=50$	$n=100$	$n=50$	$n=100$	$n=50$	$n=100$
(i)	$\text{KCMD}_{\text{L,m}}$	0.010	0.006	0.050	0.055	0.098	0.094
	FMDD	0.010	0.008	0.057	0.048	0.101	0.106
	PSS	0.006	0.007	0.055	0.049	0.108	0.091
(ii)	$\text{KCMD}_{\text{L,m}}$	0.099	0.256	0.276	0.528	0.416	0.641
	FMDD	0.147	0.405	0.418	0.715	0.575	0.816
	PSS	0.105	0.283	0.311	0.552	0.463	0.670

(i) The data $(X_i, Y_i)_{i=1}^n$ is generated by the functional linear model

$$Y_i(t) = \int_0^1 \nu(s,t)X_i(s)ds + \varepsilon_i(t), \quad t \in [0,1],$$

where X_i and ε_i are generated by independent Brownian bridges, $\nu(s,t) = r \cdot \exp(t^2/2 + s^2/2)$, and $r = 0, 0.20, 0.40, 0.75$.

(ii) The data $(X_i, Y_i)_{i=1}^n$ is generated by the model

$$Y_i(t) = \int_0^1 \nu(s,t)X_i^2(s)ds + \varepsilon_i(t), \quad t \in [0,1],$$

where all the variables are the same as in scenario (i), $r = 0, 0.5, 1.0, 2.0$.

Tables 6, 7, 8 summarize the empirical sizes and powers of the considered tests for Examples 5, 6, 7 respectively. The results show that all the tests give close empirical sizes to the nominal significance levels in the scenarios where H_0 holds (such as in the cases of $r = 0$ in Examples 5, 7, and in scenario (i) of Example 6), and the powers of them increase as the sample size gets larger or the discrepancy from the null hypothesis becomes greater (when r goes larger).

When the regression model is linear, the powers of the KCMD-based tests are close to those of the FMDD-based test and the PSS test, or slightly lower than those of the FMDD-based test. However, similar to the findings in the multivariate examples, the KCMD-based tests considered are often more powerful when the regression model is nonlinear, and their powers may increase faster than others with the samples size or the discrepancy from the null hypothesis, as shown, for example, in scenario (ii) of Table 6.

Table 8
Empirical sizes and powers of the tests based on KCMD, FMDD and PSS for different rs in Example 7 with $n = 50, 100$.

Scenario	α	Test	$n = 50$	$n = 100$	$n = 50$	$n = 100$	$n = 50$	$n = 100$	$n = 50$	$n = 100$
(i)			$r = 0$		$r = 0.20$		$r = 0.40$		$r = 0.75$	
	0.01	$KCMD_{L,m}$	0.018	0.008	0.113	0.321	0.639	0.976	0.995	1.000
		FMDD	0.012	0.008	0.148	0.380	0.734	0.990	0.997	1.000
		PSS	0.016	0.012	0.048	0.089	0.257	0.645	0.854	0.999
	0.05	$KCMD_{L,m}$	0.049	0.058	0.316	0.621	0.901	0.997	1.000	1.000
		FMDD	0.058	0.058	0.372	0.690	0.951	1.000	1.000	1.000
		PSS	0.070	0.070	0.144	0.266	0.491	0.843	0.962	1.000
	0.10	$KCMD_{L,m}$	0.099	0.100	0.441	0.763	0.951	1.000	1.000	1.000
		FMDD	0.112	0.099	0.485	0.838	0.970	1.000	1.000	1.000
		PSS	0.103	0.128	0.234	0.360	0.615	0.897	0.973	1.000
(ii)			$r = 0$		$r = 0.5$		$r = 1.0$		$r = 2.0$	
	0.01	$KCMD_{L,m}$	0.013	0.008	0.043	0.136	0.233	0.751	0.710	0.998
		FMDD	0.011	0.014	0.011	0.029	0.039	0.136	0.071	0.388
		PSS	0.022	0.011	0.015	0.027	0.022	0.202	0.127	0.880
	0.05	$KCMD_{L,m}$	0.055	0.048	0.181	0.405	0.635	0.961	0.980	1.000
		FMDD	0.061	0.054	0.091	0.149	0.209	0.601	0.410	0.970
		PSS	0.064	0.073	0.057	0.080	0.117	0.452	0.442	0.971
	0.10	$KCMD_{L,m}$	0.104	0.107	0.336	0.594	0.784	0.982	0.995	1.000
		FMDD	0.110	0.108	0.195	0.329	0.405	0.846	0.758	0.998
		PSS	0.124	0.119	0.131	0.193	0.223	0.567	0.628	0.987

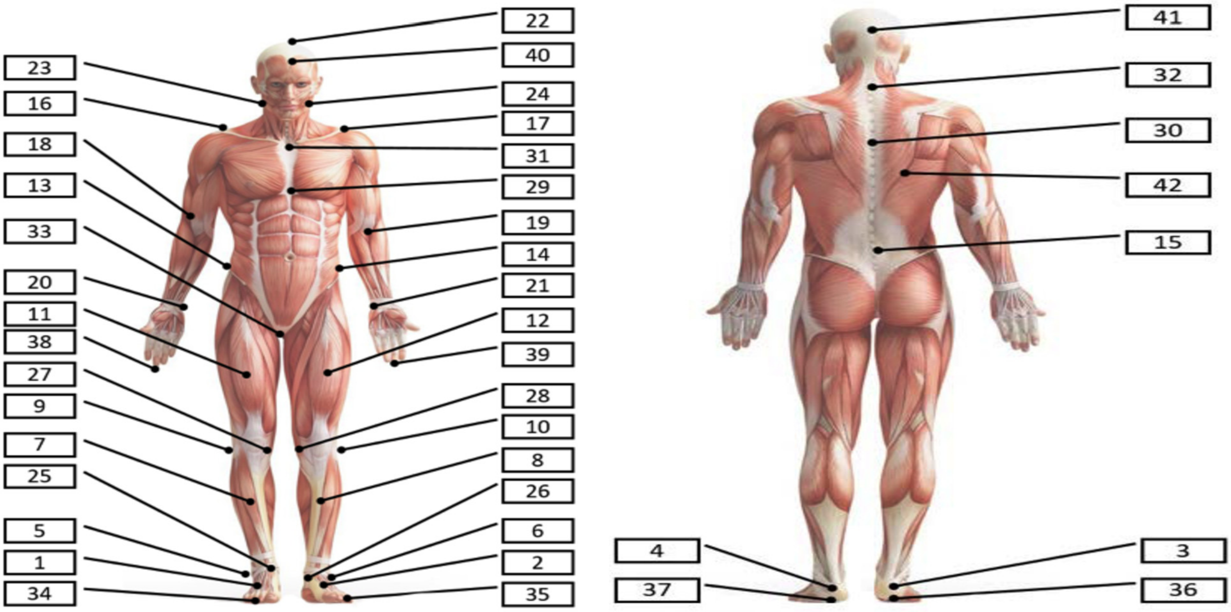

Fig. 1. The layout of the monitoring points. The pictures are from 2018 APMCM Contest Problem A.

5. Real data analysis

We present two real data examples in this section. The first real data example studies the elderly people balance ability, and the dataset is available in https://www.saikr.com/c/nd/5834, which is adopted as a raw data of a contest (2018 APMCM Contest Problem A). The dataset contains the basic data of the elderly subjects and the calibrated raw data of each subject in free walk state. There are 80 subjects, but four of them miss the raw data of free walk. Hence we consider the data of 76 subjects. The response Y of interest is fall times in one year and the predictors are curves of 42 monitoring points on the body of the elderly subjects, see the layout of the points in Fig. 1. The goal is to extract important body balance features from the 42 monitoring points in order for a comprehensive body balance assessment for elderly people. We employ the $KCMD_{L,m}$ test with Laplace kernel to check the conditional dependence of Y on the signal curve of each monitoring point at significance level 0.1. The bootstrap sample size is taken as 500. Then 25 monitoring points, which are listed Table 9, are detected to have a significant contribution to the conditional mean of the response. We also implement the FMDD-based test and present the results in Table 9. Both methods choose the same monitoring points. It appears that most of the

Table 9
The monitoring points selected by KCMD and FMDD.

	selected monitoring points															
KCMD	1	3	5	7	9	11	13	14	16	18	20	21	23	25	27	28
	29	30	31	32	36	38	39	40	41							
FMDD	1	3	5	7	9	11	13	14	16	18	20	21	23	25	27	28
	29	30	31	32	36	38	39	40	41							

selected monitoring points are located at the right side of body. After selecting the key monitoring points, further research, such as building a balance risk assessment system based on the 25 indicators, can be implemented.

The second example is used to illustrate how to apply our method to model checking. The real data we explore is the Tecator dataset contained in R package $fda.usc$. This dataset includes values of a 100-channel spectrum (of wavelength $850-1050$ mm) of absorbance, water content, fat content and protein content for 215 meat samples. Lee et al. (2020) have explored this dataset using the FMDD method. We compare our method with the FMDD and do some further analysis. The relationship between the fat content Y and the spectrometric curves X has been studied in Yao and Müller (2010). They proposed the functional quadratic regression model

$$E(Y|X)=\mu+\int_{\mathcal{T}}\beta(t)X^c(t)dt+\int_{\mathcal{T}}\int_{\mathcal{T}}\gamma(s,t)X^c(s)X^c(t)dsdt,$$

where $X^c(t)=X(t)-E\{X(t)\}$. Our goal is to test whether this model is suitable for the relationship between the fat content and spectrometric curve. We give the procedure in the following.

Firstly, use $\mathrm{KCMD_{L,m}}$ and the FMDD methods to test the conditional mean independence. With 500 bootstrap sample size, the p-values of $\mathrm{KCMD_{L,m}}$ and FMDD tests are zero. This indicates that X has a significant contribution to the conditional mean of Y.

Secondly, remove the linear effect of X on Y and explore the conditional independence of $Y-\mu-\langle X,\beta\rangle$ given X. We first fit (Y,X) to a linear model by the function $fregre.pc$ in the $fda.usc$ package, where the number of principal components used is taken as 3 or 5. Then we use $\mathrm{KCMD_{L,m}}$ and the FMDD test to the obtained residual and X. When the number of principal components is 3, the p-value of $\mathrm{KCMD_{L,m}}$ test is 0 and that of the FMDD test is 0.034, both tests reject the null hypothesis provided the given significance level $\alpha=0.05$. When the number of principal components is 5, the p-values of the tests are 0.028 and 0.116 respectively, and the $\mathrm{KCMD_{L,m}}$ test rejects the null hypothesis at level $\alpha=0.05$, while the FMDD test does not. This indicates that the KCMD-based test is more sensitive to the nonlinear relationship.

Thirdly, we fit the functional quadratic model using the method described in Yao and Müller (2010), where the number of principal components used in the estimation is 5. Then we use the KCMD test to the residuals and X. The p-value obtained is 0.76. Thus, it has no significant evidence to deny the functional quadratic model. This conclusion is consistent with the result in Horváth et al. (2013).

6. Discussion

In this paper, we propose a conditional mean dependence measure KCMD by employing the reproducing kernel embedding operator, and a test method to the conditional mean dependence of a response based on the KCMD measure. Both theoretical and numerical results show the competitive performance of the proposed tests.

KCMD may be employed for other purpose besides the conditional mean dependence test. An immediate extension is to test the conditional quantile dependence. To be precise, let $Q_\tau(Y)$ be the τth quantile of Y and $Q_\tau(Y|X)$ be the τth quantile of Y given X. The hypothesis to be tested is

$$H_0: Q_\tau(Y|X)=Q_\tau(Y) \text{ almost surely}, \quad \text{versus} \quad H_1: \mathrm{pr}\{Q_\tau(Y|X)=Q_\tau(Y)\}<1,$$

for $\tau\in(0,1)$. Note that $Q_\tau(Y|X)=Q_\tau(Y)$ is equivalent to $E(V|X)=E(V)=0$ (see Shao and Zhang (2014)), where $V=\tau-\mathbf{1}(Y-Q_\tau(Y)\le 0)$ and $\mathbf{1}$ is the indicator function. Therefore, the proposed measure or test procedure can be applied directly by replacing Y with V.

Another example is the estimation in regression models. For instance, for the linear regression model

$$Y=\langle X,\beta\rangle+\epsilon,$$

where X and Y could be vectors or functions, similar to the least squares method, one can estimate β by

$$\hat{\beta}=\arg\min_{\beta}\mathrm{KCMD}(Y-\langle X,\beta\rangle,X).$$

However, it needs further efforts to explore the theoretical results.

Some issues deserve further consideration: (i) With technology development, more and more data are collected as complex objects, hence it is imperative to extend the method to a more general space such as Banach space. (ii) The KCMD-based tests can be modified for testing the goodness-of-fit of models, as is shown in Section 5, but further theoretical investigations are needed. (iii) The computation amount of the estimated KCMD is $O(n^2)$, which might be reduced to $O(n\log n)$ using some similar skills for distance covariance (Huo and Székely, 2016).

Acknowledgement

The authors wish to thank the Associate Editor and two anonymous referees for their constructive comments on earlier drafts of this article. This work is partly supported by the National Natural Science Foundation of China (No. 11771032).

Appendix A

The proofs of Propositions 1, 2 and 3 are of routine works, therefore we omit here; the proof of Theorem 6 can be done using the same paradigm as the proof of Theorem 7, thus it is also omitted. Note that S_n has a similar representation to the FMDD statistic in Lee et al. (2020), the proofs of the asymptotic results can be carried out along the same lines as the proofs of the theorems in Lee et al. (2020), but no strong moment conditions are needed here due to the boundedness of the kernel $k(\cdot,\cdot)$. For easy reading of the paper, we write down the details of the proofs in the following. They are deduced based on the limit theory of U-statistics.

Proof of Theorem 1. By the assumption that $k(\cdot,\cdot)$ is bounded, there exists a positive constant M such that $|c_{uv}|<M$ for all $1\le u,v\le n$. With the combination of $E\|Y\|<\infty$, we have

$$\begin{aligned}
E|h(Z_1,Z_2,Z_3,Z_4)| &\le \frac{1}{4!}\sum_{(u,v,q,r)}^{(1,2,3,4)} E|c_{uv}d_{uv}-c_{uv}d_{uq}-c_{uv}d_{vr}+c_{uv}d_{qr}| \\
&\le M\frac{1}{4!}\sum_{(u,v,q,r)}^{(1,2,3,4)} E|d_{uv}-d_{uq}-d_{vr}+d_{qr}| \\
&\le 4ME\|Y\|E\|Y'\| \\
&<\infty.
\end{aligned}$$

By the strong law of large number for U-statistics,

$$\mathrm{KCMD}_n(Y,X)\xrightarrow{\text{a.s}}\mathrm{KCMD}(Y,X).$$

This completes the proof. □

Proof of Theorem 2. Define

$$h_c(z_1,\ldots,z_c)=E[h(z_1,\ldots,z_c,Z_{c+1},\ldots,Z_4)],\quad c=1,2,$$

where $z_i=(x_i,y_i)$, $1\le i\le 4$. It can be shown that $h_1(z)=0$ under H_0, where $z=(x,y)$. To see this, the sum in $h_1(z)$ can be decomposed as four parts according to which of the indexes u,v,q,r takes value 1, that is,

$$\begin{aligned}
h_1(z)=&\frac{1}{4!}\Bigg(E\{\sum_{(v,q,r)}^{(2,3,4)}[k(x,X_v)\langle y,Y_v\rangle-k(x,X_v)\langle y,Y_q\rangle-k(x,X_v)\langle Y_v,Y_r\rangle+k(x,X_v)\langle Y_q,Y_r\rangle]\} \\
&+E\{\sum_{(u,q,r)}^{(2,3,4)}[k(X_u,x)\langle Y_u,y\rangle-k(X_u,x)\langle Y_u,Y_q\rangle-k(X_u,x)\langle y,Y_r\rangle+k(X_u,x)\langle Y_q,Y_r\rangle]\} \\
&+E\{\sum_{(u,v,r)}^{(2,3,4)}[k(X_u,X_v)\langle Y_u,Y_v\rangle-k(X_u,X_v)\langle Y_u,y\rangle-k(X_u,X_v)\langle Y_v,Y_r\rangle+k(X_u,X_v)\langle y,Y_r\rangle]\} \\
&+E\{\sum_{(u,v,q)}^{(2,3,4)}[k(X_u,X_v)\langle Y_u,Y_v\rangle-k(X_u,X_v)\langle Y_u,Y_q\rangle-k(X_u,X_v)\langle Y_v,y\rangle+k(X_u,X_v)\langle Y_q,y\rangle]\}\Bigg) \\
:=&\frac{1}{4!}(J_1+J_2+J_3+J_4).
\end{aligned}$$

By condition $E(Y|X)=E(Y)$,

$$
\begin{aligned}
J_1 =&6E\{k(x,X)\langle y,Y\rangle\}-6E\{k(x,X)\langle y,\mu\rangle\}\\
&-6E\{k(x,X)\langle Y,\mu\rangle\}+6E\{k(x,X)\langle \mu,\mu\rangle\}\\
=&6E\{k(x,X)\langle y,E(Y|X)\rangle\}-6E\{k(x,X)\langle y,\mu\rangle\}\\
&-6E\{k(x,X)\langle E(Y|X),\mu\rangle\}+6E\{k(x,X)\langle \mu,\mu\rangle\}\\
=&6E\{k(x,X)\langle y,\mu\rangle\}-6E\{k(x,X)\langle y,\mu\rangle\}\\
&-6E\{k(x,X)\langle \mu,\mu\rangle\}+6E\{k(x,X)\langle \mu,\mu\rangle\}\\
=&0.
\end{aligned}
$$

Similarly, we have

$$J_2=J_3=J_4=0.$$

Hence, $h_1(z)=0$, which implies that $\mathrm{var}(h_1(Z))=0$. By a similar computation, under H_0, we have

$$h_2(z,z')=\frac{1}{6}\left\{k(x,x')-Ek(x,X')-Ek(X,x')+Ek(X,X')\right\}\langle y-\mu,y'-\mu\rangle.$$

Since X and Y are not degenerate, we have $\mathrm{var}(h_2(Z,Z'))>0$. Thus, $\mathrm{KCMD}_n(Y,X)$ is a degenerate U-statistic of order 1. Under the assumption that $k(\cdot,\cdot)$ is bounded and $E\|Y\|^2<\infty$, we have $Eh^2(Z_1,Z_2,Z_3,Z_4)<\infty$. Then, by the Theorem 5.52 in Serfling (1980), we have $n\mathrm{KCMD}_n(Y,X)\xrightarrow{\mathrm{d}}\sum_{i=1}^{\infty}\gamma_i(N_i^2-1)$. $\square$

Proof of Theorem 3. With direct calculations, we have $h_1(z)=\frac{1}{2}\{\Lambda(z)+\mathrm{KCMD}(Y,X)\}$, where $h_1(z)$ is defined in the proof of Theorem 2 and $\Lambda(z)$ is defined in Theorem 3. Therefore $\mathrm{var}(h_1(Z))=\frac{1}{4}\mathrm{var}(\Lambda(Z))=\frac{1}{4}\sigma^2>0$ due to the fact that X and Y are not degenerate variables. Under the assumption that $k(\cdot,\cdot)$ is bounded and $E\|Y\|^2<\infty$, we have $Eh^2(Z_1,Z_2,Z_3,Z_4)<\infty$. Then, by the Theorem 5.51A in Serfling (1980), we have

$$\sqrt{n}\{\mathrm{KCMD}_n(Y,X)-\mathrm{KCMD}(Y,X)\}\xrightarrow{\mathrm{d}}N(0,4\sigma^2).$$

This completes the proof. $\square$

Proof of Theorem 4. It is easy to see that

$$\mathrm{KCMD}_n(Y,X)=\mathrm{KCMD}_n(Y-E(Y),X).$$

Therefore, under the local alternative $H_{1,n}:Y=E(Y)+n^{-\beta}m(X)+\varepsilon$, we have

$$\mathrm{KCMD}_n(Y,X)=\mathrm{KCMD}_n(n^{-\beta}m(X)+\varepsilon,X).$$

Noting that

$$
\begin{aligned}
\langle n^{-\beta}m(X_i)+\varepsilon_i,n^{-\beta}m(X_j)+\varepsilon_j\rangle =&n^{-2\beta}\langle m(X_i),m(X_j)\rangle+\langle\varepsilon_i,\varepsilon_j\rangle\\
&+n^{-\beta}\langle m(X_i),\varepsilon_j\rangle+n^{-\beta}\langle\varepsilon_i,m(X_j)\rangle,
\end{aligned}
$$

$\mathrm{KCMD}_n(Y,X)$ can be decomposed into three terms,

$$\mathrm{KCMD}_n(Y,X)=n^{-2\beta}\mathrm{KCMD}_n(m(X),X)+\mathrm{KCMD}_n(\varepsilon,X)+n^{-\beta}V_n, \tag{A.1}$$

where V_n is a U-statistic,

$$V_n=\frac{1}{\binom{n}{4}}\sum_{i<j<s<t}H(\mathcal{Z}_i,\mathcal{Z}_j,\mathcal{Z}_s,\mathcal{Z}_t)$$

for the symmetrical kernel H defined as

$$H(\mathcal{Z}_i,\mathcal{Z}_j,\mathcal{Z}_s,\mathcal{Z}_t)=\frac{1}{4!}\sum_{(u,v,q,r)}^{(i,j,s,t)}(c_{uv}l_{uv}-c_{uv}l_{uq}-c_{uv}l_{vr}+c_{uv}l_{qr}),$$

with $l_{ij}=\langle m(X_i),\varepsilon_j\rangle+\langle\varepsilon_i,m(X_j)\rangle$ and $\mathcal{Z}_i=(X_i,\varepsilon_i)$. Similar to the proofs in Theorem 2, it can be shown that $E(H)=0$ and V_n is a nondegenerate U-statistic. Therefore, by the standard result of U-statistics, $n^{1/2}V_n=O_p(1)$. Since $\mathrm{KCMD}(m(X),X)>$

0 and $\mathrm{KCMD}(\varepsilon, X) = 0$, by Theorem 2 and 3, we have $\mathrm{KCMD}_n(m(X), X) = \mathrm{KCMD}(m(X), X) + O_p(n^{-1/2})$ and $\mathrm{KCMD}_n(\varepsilon, X) = O_p(n^{-1})$. Combining with (A.1), it implies that

$$\begin{aligned} n\mathrm{KCMD}_n(Y, X) =& n^{1-2\beta}\mathrm{KCMD}(m(X), X) + O_p(n^{1/2-2\beta}) \\ &+ n\mathrm{KCMD}_n(\varepsilon, X) + O_p(n^{1/2-\beta}). \end{aligned} \tag{A.2}$$

Now, we consider the following three scenarios:

(i) If $\beta > 1/2$, from (A.2) and Theorem 2, we have

$$n\mathrm{KCMD}_n(Y, X) = n\mathrm{KCMD}_n(\varepsilon, X) + o_p(1) \xrightarrow{\mathrm{d}} \sum_{i=1}^{\infty} \gamma_i'(N_i^2 - 1),$$

which completes the proof of Theorem 4 (i).

(iii) If $0 < \beta < 1/2$, then the first term on the right hand side of (A.2) is dominant, and thus it is shown that $n\mathrm{KCMD}_n(Y, X) \xrightarrow{\mathrm{p}} \infty$. This completes the proof of Theorem 4 (iii).

(ii) If $\beta = 1/2$, by equation (A.1), it follows that

$$n\mathrm{KCMD}_n(Y, X) = \mathrm{KCMD}_n(m(X), X) + n\mathrm{KCMD}_n(\varepsilon, X) + n^{1/2} V_n \tag{A.3}$$

Notice that $\mathrm{KCMD}_n(\varepsilon, X)$ is U-statistic

$$\frac{1}{\binom{n}{4}} \sum_{i<j<s<t} h(\mathcal{Z}_i, \mathcal{Z}_j, \mathcal{Z}_s, \mathcal{Z}_t),$$

with

$$h(\mathcal{Z}_i, \mathcal{Z}_j, \mathcal{Z}_k, \mathcal{Z}_t) = \frac{1}{4!} \sum_{(u,v,q,r)}^{(i,j,s,t)} (c_{uv}\mathcal{E}_{uv} - c_{uv}\mathcal{E}_{uq} - c_{uv}\mathcal{E}_{vr} + c_{uv}\mathcal{E}_{qr}),$$

where $\mathcal{E}_{ij} = \langle \varepsilon_i, \varepsilon_j \rangle$. Define

$$h_c(z_1, \ldots, z_c) = E[h(z_1, \ldots, z_c, \mathcal{Z}_{c+1}, \ldots, \mathcal{Z}_4)], \quad c = 1, 2,$$

and

$$H_1(z) = E[H(z, \mathcal{Z}_2, \mathcal{Z}_3, \mathcal{Z}_4)].$$

Similar to the arguments in the proof of Theorem 2, we have $h_1(z) = 0$ and

$$h_2(z, z') = \frac{1}{6} \left\{ k(x, x') - Ek(x, X') - Ek(X, x') + Ek(X, X') \right\} \langle \varepsilon - E(\varepsilon), \varepsilon' - E(\varepsilon) \rangle.$$

Thus, $\mathrm{KCMD}_n(\varepsilon, X)$ is degenerate. By the Lemma 5.1.5A in Serfling (1980), we have

$$\begin{aligned} n\mathrm{KCMD}_n(\varepsilon, X) &= \frac{6}{n-1} \sum_{i \neq j} h_2(\mathcal{Z}_i, \mathcal{Z}_j) + o_p(1) \\ &= \frac{6}{n-1} \sum_{i,j=1}^{n} h_2(\mathcal{Z}_i, \mathcal{Z}_j) - \frac{6}{n-1} \sum_{i=1}^{n} h_2(\mathcal{Z}_i, \mathcal{Z}_i) + o_p(1). \end{aligned} \tag{A.4}$$

Applying Mercer's theorem (see Ferreira and Menegatto (2009), Theorem 1.1), we have

$$6h_2(\mathcal{Z}, \mathcal{Z}') = \sum_{i=1}^{\infty} \gamma_i' \phi_i(\mathcal{Z}) \phi_i(\mathcal{Z}'), \tag{A.5}$$

where $(\gamma_i', \phi_i(\cdot))_{i=1}^{\infty}$ are eigenvalues and eigenfunctions of $6h_2$, and the eigenfunctions are orthonormal, that is, $E[\phi_i(\mathcal{Z})\phi_j(\mathcal{Z})] = \delta_{ij}$, $\delta_{ij} = 0$ when $i \neq j$; $\delta_{ij} = 1$ when $i = j$. It can be shown that $E\phi_i(\mathcal{Z}) = 0$ for $i = 1, 2, \ldots$ (refer to page 5 in the supplementary material of Lee et al. (2020)). Replacing $n-1$ with n in (A.4) does not change the limiting distribution. Therefore, $n\mathrm{KCMD}_n(\varepsilon, X)$ can be rewritten as

$$n\mathrm{KCMD}_n(\varepsilon, X) = \sum_{i=1}^{\infty} \gamma_i' \left[\left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \phi_i(\mathcal{Z}_j) \right\}^2 - \frac{1}{n} \sum_{i=1}^{n} \phi_i^2(\mathcal{Z}_i) \right] + o_p(1).$$

Recall that V_n is a nondegenerate U-statistic. Thus, similarly we have

$$n^{1/2}V_n = \frac{4}{\sqrt{n}}\sum_{i=1}^{n} H_1(\mathcal{Z}_i) + o_p(1).$$

Consequently, we can write $n\text{KCMD}_n(Y, X)$ as

$$\begin{aligned} n\text{KCMD}_n(Y, X) =& \text{KCMD}_n(m(X), X) + \frac{4}{\sqrt{n}}\sum_{i=1}^{n} H_1(\mathcal{Z}_i) + \\ & \sum_{i=1}^{\infty}\gamma_i'\left[\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\phi_i(\mathcal{Z}_j)\right\}^2 - \frac{1}{n}\sum_{j=1}^{n}\phi_i^2(\mathcal{Z}_j)\right] + o_p(1). \end{aligned}$$

For any fixed positive integer L, using the multivariate central limit theorem, we have

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} H_1(\mathcal{Z}_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \phi_1(\mathcal{Z}_i) \\ \vdots \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \phi_L(\mathcal{Z}_i) \end{pmatrix} \xrightarrow{\text{d}} N(\mathbf{0}, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} \text{var}(H_1(\mathcal{Z})) & E[H_1(\mathcal{Z})\phi_1(\mathcal{Z})] & \dots & E[H_1(\mathcal{Z})\phi_L(\mathcal{Z})] \\ E[\phi_1(\mathcal{Z})H_1(\mathcal{Z})] & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ E[\phi_L(\mathcal{Z})H_1(\mathcal{Z})] & 0 & \dots & 1 \end{pmatrix}.$$

Using this result and Theorem 1.4.8 in van der Vaart and Wellner (1996), the countable sequence $(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} H_1(\mathcal{Z}_i)$, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_1(\mathcal{Z}_i)$, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_2(\mathcal{Z}_i), \dots)$ converges in distribution to a random sequence $(K, N_1, N_2, \dots)$, where K and N_is are normally distributed variables with zero mean and covariances

$$\begin{aligned} &E(K^2) = \text{var}(H_1(\mathcal{Z})), \\ &E(KN_i) = E[H_1(\mathcal{Z})\phi_i(\mathcal{Z})], \\ &E(N_iN_j) = \delta_{ij}. \end{aligned}$$

In addition, by the law of large number, it follows that

$$\frac{1}{n}\sum_{j=1}^{n}\phi_i^2(\mathcal{Z}_j) \xrightarrow{\text{p}} E\phi_i^2(\mathcal{Z}) = 1.$$

Therefore, noting that $\sum_{i=1}^{L}(\gamma_i')^2\text{var}(N_i^2 - 1) \to 2\sum_{i=1}^{\infty}(\gamma_i')^2 = 72E[h_2^2(\mathcal{Z}, \mathcal{Z}')] < \infty$ as $L \to \infty$ and applying the continuous mapping theorem, we have

$$\frac{4}{\sqrt{n}}\sum_{i=1}^{n} H_1(\mathcal{Z}_i) + \sum_{i=1}^{\infty}\gamma_i'\left[\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\phi_i(\mathcal{Z}_j)\right\}^2 - \frac{1}{n}\sum_{j=1}^{n}\phi_i^2(\mathcal{Z}_j)\right] \xrightarrow{\text{d}} 4K + \sum_{i=1}^{\infty}\gamma_i'(N_i^2 - 1).$$

By Theorem 1, $\text{KCMD}_n(m(X), X) \xrightarrow{\text{p}} \text{KCMD}(m(X), X)$. Then, by Slutsky lemma,

$$\begin{aligned} n\text{KCMD}_n(Y, X) =& \text{KCMD}_n(m(X), X) + \frac{4}{\sqrt{n}}\sum_{i=1}^{n} H_1(\mathcal{Z}_i) \\ & + \sum_{i=1}^{\infty}\gamma_i'\left\{\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\phi_i(\mathcal{Z}_j)\right)^2 - \frac{1}{n}\sum_{j=1}^{n}\phi_i^2(\mathcal{Z}_j)\right\} + o_p(1) \\ & \xrightarrow{\text{d}} \text{KCMD}(m(X), X) + 4K + \sum_{i=1}^{\infty}\gamma_i'(N_i^2 - 1). \end{aligned}$$

Let $V = 4K$, then the proof is complete. □

We need some notations and Theorem 4 in Lee et al. (2020) to finish the following proofs. Let S_n^* be a bootstrap statistic depending on $(Z_i)_{i=1}^n$. Define $S_n^* = o_p^*(1)$ $a.s.$ if

$$\mathrm{pr}^*(|S_n^*| > \epsilon) \to 0 \, a.s.$$

for any $\epsilon > 0$, where pr^* denotes the conditional probability given $(Z_i)_{i=1}^n$. Similarly, we can define $\mathrm{var}^*(S_n^*)$ as the conditional variance of S_n^* given $(Z_i)_{i=1}^n$. For convenience of reference, we list Theorem 4 in Lee et al. (2020) as a lemma.

Lemma A.1. *(Lee et al. (2020)) Suppose h is a symmetric kernel satisfying $E\{h(Z, Z')^4\} < \infty$, $U_n = \frac{1}{n(n-1)} \sum_{i \neq j} h(Z_i, Z_j)$, and $(W_i)_{i=1}^n$ are i.i.d. copies of random variable W with $E(W) = 0$, $E(W^2) = 1$ and $E(W^4) < \infty$. Then the bootstrap statistic given by $nU_n^* = \frac{1}{n(n-1)} \sum_{i \neq j} h(Z_i, Z_j) W_i W_j$ has the following limiting distribution,*

$$nU_n^* \xrightarrow{d^*} \sum_{i=1}^{\infty} \gamma_i (N_i^2 - 1) \quad a.s.,$$

where $(N_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. $N(0, 1)$ random variables and $(\gamma_i)_{i \in \mathbb{N}}$ is a sequence of eigenvalues associated with the kernel h.

Proof of Theorem 5. By Lemma A.1, we only need to show that

$$\frac{1}{n-3} \sum_{i \neq j} \zeta_i C_{ij} D_{ij} \zeta_j = \frac{1}{n-3} \sum_{i \neq j} \Gamma(Z_i, Z_j) \zeta_i \zeta_j + o_p^*(1) \quad a.s. \tag{A.6}$$

And this holds if

$$\mathrm{var}^* \left(\frac{1}{n-3} \sum_{i \neq j} (C_{ij} D_{ij} - \Gamma(Z_i, Z_j)) \zeta_i \zeta_j \right) = \frac{1}{(n-3)^2} \sum_{i \neq j} (C_{ij} D_{ij} - \Gamma(Z_i, Z_j))^2 \xrightarrow{\text{a.s.}} 0. \tag{A.7}$$

Next we show (A.7) holds. For the ease of notation, write $\varphi_{ij} = \varphi(X_i, X_j)$ and $\psi_{ij} = \psi(Y_i, Y_j)$. Then $\Gamma(Z_i, Z_j) = \varphi_{ij}\psi_{ij}$. Since $k(\cdot, \cdot)$ is bounded, there exists a constant M such that $\sup_{i,j}(|C_{ij}|) < M$. Notice that

$$\begin{aligned}
\sum_{i \neq j} (C_{ij} D_{ij} - \Gamma(Z_i, Z_j))^2 &= \sum_{i \neq j} (C_{ij} D_{ij} - \varphi_{ij}\psi_{ij})^2 \\
&= \sum_{i \neq j} (C_{ij} D_{ij} - C_{ij}\psi_{ij} + C_{ij}\psi_{ij} - \varphi_{ij}\psi_{ij})^2 \\
&\leq 2 \sum_{i \neq j} (D_{ij} - \psi_{ij})^2 C_{ij}^2 + 2 \sum_{i \neq j} \psi_{ij}^2 (C_{ij} - \varphi_{ij})^2 \\
&\leq 2M^2 \sum_{i \neq j} (D_{ij} - \psi_{ij})^2 + 2 \left(\sum_{i \neq j} \psi_{ij}^4 \right)^{1/2} \left(\sum_{i \neq j} (C_{ij} - \varphi_{ij})^4 \right)^{1/2}.
\end{aligned}$$

Under the condition $E\|Y\|^4 < \infty$, we have

$$\begin{aligned}
E\psi_{12}^4 &= E\langle Y_1 - \mu, Y_2 - \mu \rangle^4 \\
&\leq E\|Y_1 - \mu\|^4 \|Y_2 - \mu\|^4 \\
&= E\|Y_1 - \mu\|^4 E\|Y_2 - \mu\|^4 \\
&< \infty.
\end{aligned}$$

Therefore, by the law of large number for U-statistic

$$\frac{1}{n^2} \sum_{i \neq j} \psi_{ij}^4 \xrightarrow{\text{a.s.}} E\psi_{12}^4.$$

Thus, we only need to show that

$$\frac{1}{n^2} \sum_{i \neq j} (D_{ij} - \psi_{ij})^2 \xrightarrow{\text{a.s.}} 0, \tag{A.8}$$

$$\frac{1}{n^2} \sum_{i \neq j} (C_{ij} - \varphi_{ij})^4 \xrightarrow{\text{a.s.}} 0. \tag{A.9}$$

We only prove (A.8), (A.9) can be shown by the similar arguments. By straightforward calculations, we have

$$\frac{1}{n^2}\sum_{i\neq j}(D_{ij}-\psi_{ij})^2 = \frac{1}{n^2}\sum_{i\neq j}\Big(\frac{1}{n-2}\sum_{l=1}^{n}\langle Y_i, Y_l\rangle + \frac{1}{n-2}\sum_{l=1}^{n}\langle Y_j, Y_l\rangle - \frac{1}{(n-1)(n-2)}\sum_{q,l=1}^{n}\langle Y_q, Y_l\rangle - \langle Y_i,\mu\rangle - \langle Y_j,\mu\rangle + \langle \mu,\mu\rangle\Big)^2.$$

For simplicity, we replace $n-2$ and $n-1$ on the right hand side of the preceding formula by n since it does not change the result. Denote the obtained quantity as

$$T_n = \frac{1}{n^2}\sum_{i\neq j}\left(\frac{1}{n}\sum_{l=1}^{n}\langle Y_i, Y_l\rangle + \frac{1}{n}\sum_{l=1}^{n}\langle Y_j, Y_l\rangle - \frac{1}{n^2}\sum_{q,l=1}^{n}\langle Y_q, Y_l\rangle - \langle Y_i,\mu\rangle - \langle Y_j,\mu\rangle + \langle \mu,\mu\rangle\right)^2,$$

then

$$\begin{aligned} T_n &= \frac{1}{n^2}\sum_{i\neq j}\left(\frac{1}{n}\sum_{l=1}^{n}\langle Y_i, Y_l-\mu\rangle + \frac{1}{n}\sum_{l=1}^{n}\langle Y_j, Y_l-\mu\rangle - \frac{1}{n^2}\sum_{q,l=1}^{n}\big(\langle Y_q, Y_l\rangle - \langle\mu,\mu\rangle\big)\right)^2 \\ &\leq \frac{3}{n^2}\sum_{i\neq j}\left\{\left(\frac{1}{n}\sum_{l=1}^{n}\langle Y_i, Y_l-\mu\rangle\right)^2 + \left(\frac{1}{n}\sum_{l=1}^{n}\langle Y_j, Y_l-\mu\rangle\right)^2 + \left(\frac{1}{n^2}\sum_{q,l=1}^{n}\big(\langle Y_q, Y_l\rangle - \langle\mu,\mu\rangle\big)\right)^2\right\}. \end{aligned} \tag{A.10}$$

Under the condition $E\|Y\|^4<\infty$, we have $E|\langle Y_s, Y_t-\mu\rangle\langle Y_s, Y_r-\mu\rangle|<\infty$ for $1\leq s,t,r\leq 3$ and $E|\langle Y_1, Y_2-\mu\rangle\langle Y_1, Y_3-\mu\rangle| = 0$. Then, by the strong law of large number for V-statistics, we have

$$\begin{aligned} \frac{1}{n^2}\sum_{i\neq j}\left(\frac{1}{n}\sum_{l=1}^{n}\langle Y_i, Y_l-\mu\rangle\right)^2 &\leq \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{l=1}^{n}\langle Y_i, Y_l-\mu\rangle\right)^2 \\ &= \frac{1}{n^3}\sum_{i,l,q=1}^{n}\langle Y_i, Y_l-\mu\rangle\langle Y_i, Y_q-\mu\rangle \xrightarrow{\text{a.s.}} 0. \end{aligned}$$

Similar arguments show that the second and third terms in (A.10) converge almost surely to zero. Hence $T_n \xrightarrow{\text{a.s.}} 0$. □

Proof of Theorem 7. Under the local alternative hypothesis and the conditions in Theorem 7, (A.6) still holds, that is,

$$\frac{1}{n-3}\sum_{i\neq j}\zeta_i C_{ij} D_{ij}\zeta_j = \frac{1}{n-3}\sum_{i\neq j}\Gamma(\mathcal{Z}_i,\mathcal{Z}_j)\zeta_i\zeta_j + o_p^*(1) \quad a.s.$$

Further, it can be shown that

$$\frac{1}{n-3}\sum_{i\neq j}\Gamma(\mathcal{Z}_i,\mathcal{Z}_j)\zeta_i\zeta_j = \frac{1}{n-3}\sum_{i\neq j}\varphi(X_i,X_j)\psi(\varepsilon_i,\varepsilon_j)\zeta_i\zeta_j + o_p^*(1) \quad a.s.$$

To see this, similar to the proof of Theorem 5,

$$\begin{aligned} &\operatorname{var}^*\left(\frac{1}{n-3}\sum_{i\neq j}\left\{\Gamma(\mathcal{Z}_i,\mathcal{Z}_j) - \varphi(X_i,X_j)\psi(\varepsilon_i,\varepsilon_j)\right\}\zeta_i\zeta_j\right) \\ &= \frac{1}{(n-3)^2}\sum_{i\neq j}\varphi^2(X_i,X_j)\left\{\frac{1}{n^{2\beta}}\langle m(X_i), m(X_j)\rangle + \frac{1}{n^{\beta}}\Big(\langle m(X_i),\varepsilon_j\rangle + \langle m(X_j),\varepsilon_i\rangle\Big)\right\}^2 \\ &\leq O(n^{-2-4\beta})\left\{\sum_{i\neq j}\varphi^4(X_i,X_j)\right\}^{1/2}\left\{\sum_{i\neq j}\langle m(X_i), m(X_j)\rangle^4\right\}^{1/2} \\ &\quad + O(n^{-2-2\beta})\left\{\sum_{i\neq j}\varphi^4(X_i,X_j)\right\}^{1/2}\left\{\sum_{i\neq j}\big(\langle m(X_i),\varepsilon_j\rangle + \langle m(X_j),\varepsilon_i\rangle\big)^4\right\}^{1/2} \xrightarrow{\text{a.s.}} 0 \end{aligned}$$

The last formula comes from the following facts. Since kernel $k(\cdot,\cdot)$ is bounded and $E(|m(X)|^4 + |\varepsilon|^4) < \infty$,

$$\frac{1}{n^2}\sum_{i\neq j}\varphi^4(X_i, X_j) \xrightarrow{\text{a.s.}} E\varphi^4(X, X'),$$

$$\frac{1}{n^2}\sum_{i\neq j}\langle m(X_i), m(X_j)\rangle^4 \xrightarrow{\text{a.s.}} E\left[\langle m(X), m(X')\rangle^4\right],$$

$$\frac{1}{n^2}\sum_{i\neq j}\left(\langle m(X_i), \varepsilon_j\rangle + \langle m(X_j), \varepsilon_i\rangle\right)^4 \xrightarrow{\text{a.s.}} E\left[\left(\langle m(X), \varepsilon'\rangle + \langle m(X'), \varepsilon\rangle\right)^4\right].$$

Therefore,

$$\frac{1}{n-3}\sum_{i\neq j}\zeta_i C_{ij} D_{ij}\zeta_j = \frac{1}{n-3}\sum_{i\neq j}\varphi(X_i, X_j)\psi(\varepsilon_i, \varepsilon_j)\zeta_i\zeta_j + o_p^*(1) \quad a.s.$$

Applying Lemma A.1, we have

$$n\mathrm{KCMD}_n^*(Y, X) \xrightarrow{\mathrm{d}^*} \mathcal{G}_0 := \sum_{i=1}^{\infty}\gamma_i'(N_i^2 - 1)\, a.s.$$

Using Lemma 11.2.1 (ii) in Lehmann and Romano (2005) and the fact that $\mathcal{G}_0$ is a continuous random variable (using similar arguments on page 13 of supplementary material of Lee et al. (2020)), we have

$$Q^*_{(1-\alpha),n} \xrightarrow{\mathrm{p}} Q_{(1-\alpha),\mathcal{G}_0},$$

where $Q_{(1-\alpha),\mathcal{G}_0}$, is the $(1-\alpha)$th quantile of $\mathcal{G}_0$. Then, by Theorem 4, the conclusions follow. $\square$

References

Cook, R.D., Li, B., et al., 2002. Dimension reduction for conditional mean in regression. Ann. Stat. 30, 455–474.
Ferreira, J., Menegatto, V., 2009. Eigenvalues of integral operators defined by smooth positive definite kernels. Integral Equ. Oper. Theory 64, 61–81.
Fukumizu, K., Gretton, A., Schölkopf, B., Sriperumbudur, B.K., 2009. Characteristic kernels on groups and semigroups. In: Advances in Neural Information Processing Systems, vol. 21. Curran Associates, Red Hook, NY, pp. 473–480.
Gretton, A., Bousquet, O., Smola, A., Schölkopf, B., 2005. Measuring statistical dependence with Hilbert-Schmidt norms. In: International Conference on Algorithmic Learning Theory. Springer, Berlin, pp. 63–77.
Horváth, L., Reeder, R., et al., 2013. A test of significance in functional quadratic regression. Bernoulli 19, 2120–2151.
Huo, X., Székely, G.J., 2016. Fast computing for distance covariance. Technometrics 58, 435–447.
Lee, C., Zhang, X., Shao, X., 2020. Testing conditional mean independence for functional data. Biometrika 107, 331–346.
Lee, C.E., Shao, X., 2018. Martingale difference divergence matrix and its application to dimension reduction for stationary multivariate time series. J. Am. Stat. Assoc. 113, 216–229.
Lehmann, E.L., Romano, J.P., 2005. Testing Statistical Hypotheses, 3rd edition. Springer, New York.
Lei, J., 2014. Adaptive global testing for functional linear models. J. Am. Stat. Assoc. 109, 624–634.
Liu, J., Xu, P., Lian, H., 2019. Estimation for single-index models via martingale difference divergence. Comput. Stat. Data Anal. 137, 271–284.
Park, T., Shao, X., Yao, S., et al., 2015. Partial martingale difference correlation. Electron. J. Stat. 9, 1492–1517.
Patilea, V., Sánchez-Sellero, C., Saumard, M., 2016. Testing the predictor effect on a functional response. J. Am. Stat. Assoc. 111, 1684–1695.
Pfister, N., Bühlmann, P., Schölkopf, B., Peters, J., 2018. Kernel-based tests for joint independence. J. R. Stat. Soc., Ser. B, Stat. Methodol. 80, 5–31.
Sejdinovic, D., Sriperumbudur, B., Gretton, A., Fukumizu, K., et al., 2013. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. Ann. Stat. 41, 2263–2291.
Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. John Wiley & Sons.
Shao, X., Zhang, J., 2014. Martingale difference correlation and its use in high-dimensional variable screening. J. Am. Stat. Assoc. 109, 1302–1318.
Shen, C., Priebe, C.E., Vogelstein, J.T., 2020. From distance correlation to multiscale graph correlation. J. Am. Stat. Assoc. 115, 280–291.
Sriperumbudur, B.K., Gretton, A., Fukumizu, K., Lanckriet, G., Schölkopf, B., 2008. Injective Hilbert space embeddings of probability measures. In: 21st Annual Conference on Learning Theory (COLT 2008). Omnipress, pp. 111–122.
Sriperumbudur, B.K., Gretton, A., Fukumizu, K., Schölkopf, B., Lanckriet, G.R., 2010. Hilbert space embeddings and metrics on probability measures. J. Mach. Learn. Res. 11, 1517–1561.
Su, L., Zheng, X., 2017. A martingale-difference-divergence-based test for specification. Econ. Lett. 156, 162–167.
Székely, G.J., Rizzo, M.L., Bakirov, N.K.B., 2007. Measuring and testing dependence by correlation of distances. Ann. Stat. 35, 2769–2794.
Székely, G.J., Rizzo, M.L., et al., 2014. Partial distance correlation with methods for dissimilarities. Ann. Stat. 42, 2382–2412.
van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer, New York.
Yao, F., Müller, H.G., 2010. Functional quadratic regression. Biometrika 97, 49–64.
Zhang, X., Yao, S., Shao, X., et al., 2018. Conditional mean and quantile dependence testing in high dimension. Ann. Stat. 46, 219–246.