Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 34(1): 411–423.
DOI: 10.3934/era.2026019
Received: 15 November 2025
Revised: 22 December 2025
Accepted: 09 January 2026
Published: 14 January 2026

Research article

A new measure of partial conditional mean independence in Hilbert spaces

Jiangyuan Bian*and Zhongzhan Zhang

School of Mathematics, Statistics and Mechanics, Beijing University of Technology, Beijing 100124, China

* **Correspondence:** Email: jybian@emails.bjut.edu.cn.

Abstract: A novel metric, called partial martingale difference-angle divergence, is proposed to test and measure partial conditional mean (in)dependence for Hilbert elements. The partial martingale difference-angle divergence has some appealing properties. It is nonnegative and equals zero if and only if the partial conditional mean independence holds; it has a simple expectation form; it does not require the moment condition for the predictor variable. We construct an estimator for partial martingale difference-angle divergence and derive its asymptotic properties. Finite sample simulations show that the proposed test performs well and has strong testing power for nonlinear relationships. Two real data examples are introduced to illustrate the application of the proposed test.

Keywords: partial conditional mean independence; functional data; V-statistic; nonlinear relationship; permutation test

1. Introduction

Suppose that X, Y, and Z are random elements in $\mathbb{L}^2$ spaces $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$, respectively. We aim to test the hypothesis

$$H_0 : P\{\mathbb{E}(Y \mid X, Z) = \mathbb{E}(Y \mid Z)\} = 1 \quad \text{vs.} \quad H_1 : \text{others} \tag{1.1}$$

by which we assess whether X contributes to the conditional mean of Y after adjusting for the influence of Z. Measuring and testing partial conditional mean dependence plays a vital role in statistics. When H_0 holds, it is believed that X has no additional effect on Y given Z, and we can omit X from the regression model. This is, to some extent, consistent with the idea of significance test [1] or omitted variable test [2].

In particular, when Z is absent, the null hypothesis in (1.1) becomes $P\{\mathbb{E}(Y \mid X) = \mathbb{E}(Y)\} = 1$, which was studied extensively by many authors. [3] proposed the martingale difference divergence (MDD) to measure the dependence of the conditional mean of $Y \in \mathbb{R}^1$ on $X \in \mathbb{R}^q$. As an extension of MDD, [4] proposed functional martingale difference divergence (FMDD), which can be applied to vector-valued data and functional data. [5] proposed a kernel-based conditional mean dependence (KCMD) test, which has stronger power than FMDD when there is a nonlinear relationship between Y and X. Many other methods have been developed, see [6–9] among others.

However, in practice, there may be situations where Z is known to contribute to the variation of Y. In recent years, measuring and testing partial conditional mean dependence has not only attracted much attention in statistics but has also seen developments in machine learning. [10] extended MDD and the partial correlation in [11] by defining partial martingale difference divergence (pMDD) to test (1.1), where X, Y, and Z are random vectors in Euclidean spaces. [12] proposed a partial mean independence test (pMIT) which possesses zero equivalence (i.e., pMIT equals 0 if and only if H_0 holds), and constructed the estimator based on deep neural networks and sample splitting. [13] introduced the projected covariance measure (PCM) and its application to variable significance testing. [14] constructed the target parameter $\tau := \mathbb{E}[\{\mathbb{E}(Y \mid X, Z) - \mathbb{E}(Y \mid X)\}^2]$ and adopted the plug-in estimator of τ. However, the constructed test statistic exhibits a degenerate distribution. To avoid this issue, [15] (see also [16]) constructed a new target parameter $\tau^0 := \mathbb{E}[\{Y - \mathbb{E}(Y \mid Z)\}^2] - \mathbb{E}[\{Y - \mathbb{E}(Y \mid X, Z)\}^2]$ and derived the asymptotic normality of the estimator under the null hypothesis. The tests mentioned above were developed for random variables in Euclidean spaces and may not be directly applied to functional data except the test statistic pMDD. Since there has been insufficient research on (1.1) in Hilbert spaces, we establish a general framework for measuring and testing the partial conditional mean dependence within such spaces. Crucially, the framework's formulation and theoretical validity encompass both finite-dimensional Euclidean spaces and function spaces, enabling rigorous applications across dimensionality spectra.

The remainder of this article is organized as follows. Section 2 introduces some notions and a new metric, partial martingale difference-angle divergence (pMDAD). Section 3 provides the estimator of pMDAD and establishes its asymptotic properties. Section 4 details the practical implementation of the proposed testing framework. Section 5 examines the finite-sample performance of the proposed test through simulation studies. We demonstrate the application by two real data examples in Section 6. In Section 7, some conclusions are included. All relevant proofs are provided in the Appendix.

2. Partial martingale difference-angle divergence

Let (X, Y, Z) be a random element on the product space $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ with marginal probability measures μ_X, μ_Y, and μ_Z on $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, respectively. Denote the joint measure of (X, Y, Z) by μ_{XYZ}. For simplicity of notation, all the inner products and norms on $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ are denoted as $\langle \cdot, \cdot \rangle$ and $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$.

Denote $U = Y - \mathbb{E}(Y \mid Z)$. The null hypothesis H_0 in (1.1) holds if and only if

$$\mathbb{E}(Y - \mathbb{E}(Y \mid Z) \mid X, Z) = \mathbb{E}(U \mid X, Z) = 0, \quad \text{a.e.,} \tag{2.1}$$

which is equivalent to

$$\left\| \mathbb{E}\left[UI\left(\langle X, h \rangle + \langle Z, k \rangle \le t \right) \right] \right\|^2 = 0, \text{ a.e. for all } h \in \mathcal{X}, k \in \mathcal{Z} \text{ and } t \in \mathbb{R}, \tag{2.2}$$

where $I(\cdot)$ represents the indicator function.

Let (X', Y', Z') be an independent copy of (X, Y, Z) and $U' = Y' - \mathbb{E}(Y' \mid Z')$, and (2.2) is equivalent to

$$\mathbb{E}\Big[\langle U, U'\rangle \Big\{\int_{\mathcal{X}}\int_{\mathcal{Z}}\int_{\mathbb{R}} I(\langle X,h\rangle+\langle Z,k\rangle\le t)I(\langle X',h\rangle+\langle Z',k\rangle\le t)\phi(t)dtd\nu_Z(k)d\nu_X(h)\Big\}\Big]=0, \tag{2.3}$$

where we fix $\phi(t)$ to be the probability density function of t, which follows the standard normal distribution, and ν_X and ν_Z to be nondegenerate Gaussian measures with mean zero and covariance operators Q_X and Q_Z on $\mathcal{X}$ and $\mathcal{Z}$, respectively.

Let $(\lambda_X^j, e_X^j)_{j=1}^{\infty}$, $(\lambda_Z^j, e_Z^j)_{j=1}^{\infty}$ be the eigenvalues and the corresponding eigenfunctions of Q_X and Q_Z, respectively. For $x \in \mathcal{X}$, $z \in \mathcal{Z}$, define $Q_X^{1/2}x = \sum_{j=1}^{\infty} \sqrt{\lambda_X^j}\langle x, e_X^j\rangle e_X^j$, $Q_Z^{1/2}z = \sum_{j=1}^{\infty} \sqrt{\lambda_Z^j}\langle z, e_Z^j\rangle e_Z^j$. Then, based on the Lemma 2 and Theorem 1 in [17] and combined with the work of [18], there is an explicit formula of the curly brackets in (2.3)

$$\frac{1}{4}+\frac{1}{2\pi}\theta(X,X',Z,Z'),$$

with

$$\theta(X,X',Z,Z') = \arcsin\Big(\frac{1+\langle Q_X^{1/2}X, Q_X^{1/2}X'\rangle+\langle Q_Z^{1/2}Z, Q_Z^{1/2}Z'\rangle}{\sqrt{1+\|Q_X^{1/2}X\|^2+\|Q_Z^{1/2}Z\|^2}\sqrt{1+\|Q_X^{1/2}X'\|^2+\|Q_Z^{1/2}Z'\|^2}}\Big). \tag{2.4}$$

Therefore, (2.3) boils down to

$$\mathbb{E}[\langle U, U'\rangle\theta(X,X',Z,Z')]=0. \tag{2.5}$$

Thus we can define partial martingale difference-angle divergence as follows.

Definition 1. *(Partial martingale difference-angle divergence). The partial martingale difference-angle divergence of Y given X, after controlling for the effect of Z, that is,* $\mathrm{pMDAD}(Y \mid X, Z)$ *can be defined as*

$$\mathrm{pMDAD}(Y\mid X,Z)=\mathbb{E}[\langle U, U'\rangle\theta(X,X',Z,Z')]. \tag{2.6}$$

Remark 1. *Suppose $\mathcal{X}=\mathbb{R}^p$, $\mathcal{Z}=\mathbb{R}^r$ and replace the Gaussian measure with the standard normal distribution, i.e., $(h^T,k^T,t)^T\in\mathbb{R}^{p+r+1}$ follows the multivariate normal distribution with zero mean and identity covariance matrix. Then, we have*

$$\theta(X,X',Z,Z')=\arcsin\Big(\frac{1+X^TX'+Z^TZ'}{\sqrt{1+X^TX+Z^TZ}\sqrt{1+X'^TX'+Z'^TZ'}}\Big)$$

in (2.6).

Based on the preceding analysis, we establish the following proposition, which demonstrates that $\mathrm{pMDAD}(Y \mid X, Z)$ fully characterizes partial conditional mean independence and can thus be effectively used to distinguish between the null and alternative hypotheses.

Proposition 1. *The given* $\mathrm{pMDAD}(Y \mid X, Z)$ *satisfies the following properties:*

(i) $\mathrm{pMDAD}(Y \mid X, Z) \ge 0$*;*
(ii) $\mathrm{pMDAD}(Y \mid X, Z) = 0$ *if and only if* $\mathbb{E}(Y \mid X, Z) = \mathbb{E}(Y \mid Z)$ *almost surely.*

3. Empirical estimator of pMDAD and its asymptotic property

We construct an estimator of pMDAD$(Y \mid X, Z)$ and give its asymptotic theories in this section.

Suppose that $\{(X_i, Y_i, Z_i)\}_{i=1}^n$ are i.i.d. observations from μ_{XYZ} on $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$. Since U_i is unobservable, we may consider using the nonparametric estimator of $\mathbb{E}(Y \mid Z_i)$ [19, 20]

$$\hat{\mathbb{E}}(Y \mid Z_i) = \frac{\sum_{j=1}^n g(h^{-1}\|Z_i - Z_j\|)Y_j}{\sum_{l=1}^n g(h^{-1}\|Z_i - Z_l\|)}$$

to substitute $\mathbb{E}(Y \mid Z_i)$ in $U_i = Y - \mathbb{E}(Y \mid Z_i)$, where $g(\cdot)$ is a kernel function and h is a bandwidth.

Define

$$p_{ij} = \langle \hat{U}_i, \hat{U}_j \rangle, \qquad a_{ij} = \theta(X_i, X_j, Z_i, Z_j).$$

Then, an estimator of pMDAD$(Y \mid X, Z)$ is defined as

$$\text{pMDAD}_n(Y \mid X, Z) = \frac{1}{n^2}\sum_{i,j=1}^n p_{ij}a_{ij}. \tag{3.1}$$

We can use the estimator as a test statistic, and reject H_0 for its large value. Furthermore, to study the asymptotic properties of the test statistic, we need to introduce some notations and assumptions.

Denote $S_{\mathcal{Z}}$ as the support of Z, which is a bounded subset of $\mathcal{Z}$, and denote the $\epsilon(>0)$-neighborhood of $S_{\mathcal{Z}}$ by $S_{\mathcal{Z}}^{\epsilon} = \{z \in \mathcal{Z}, \exists z' \in S_{\mathcal{Z}}, \|z - z'\| \le \epsilon\}$. The small ball probability function of Z is defined by $\psi_z(h) = P(\|z - Z\| \le h)$. Let $N_\varepsilon(S_{\mathcal{Z}})$ be the minimal number of open balls in $\mathcal{Z}$ of radius $\varepsilon(>0)$ which are necessary to cover $S_{\mathcal{Z}}$ and denote the Kolmogorov's entropy of $S_{\mathcal{Z}}$ for any ε as $\Psi_{S_{\mathcal{Z}}}(\varepsilon) = \log(N_\varepsilon(S_{\mathcal{Z}}))$. Then, we need the following assumptions.

(A1) $\exists 0 < C_1 < \infty, b > 0, \epsilon_0 > 0$, such that for any $z, z' \in S_{\mathcal{Z}}^{\epsilon_0}, \|E(Y \mid Z = z) - E(Y \mid Z = z')\| \le C_1\|z - z'\|^b$.

(A2) $\exists C_2 > 0$, such that for any $r \ge 1, \mathbb{E}(\|Y\|^r \mid Z) < C_2 r! < \infty$, a.e., where $r! = r(r-1)\cdots(r-[r]+1)$, and $[r]$ is the largest integer smaller than r.

(A3) There exists a deterministic function $f(z)$ on $S_{\mathcal{Z}}$, such that $0 < \inf_{z\in S_{\mathcal{Z}}} f(z) \le \sup_{z\in S_{\mathcal{Z}}} f(z) < \infty$, and a non-decreasing function $\psi(h)$ with $\psi(h) \to 0$ as $h \to 0$, such that for every $z \in S_{\mathcal{Z}}$ and $h > 0$, $\psi_z(h) = f(z)\psi(h) + o(\psi(h))$.

(A4) The function $\psi(h)$ given in (A3) is differentiable. $\exists C_5 > 0, h_0 > 0$, such that for any $h < h_0$, $0 \le \psi'(h) < C_5$.

(A5) For n large enough, $\frac{(\log n)^2}{n\psi(h)} < \Psi_{S_{\mathcal{Z}}}\left(\frac{\log n}{n}\right) < \frac{n\psi(h)}{\log n}$.

(A6) Kolmogorov's entropy of $S_{\mathcal{Z}}$ satisfies $\sum_{n=1}^\infty \exp\left\{(1-\epsilon_1)\Psi_{S_{\mathcal{Z}}}\left(\frac{\log n}{n}\right)\right\} < \infty$ for some $\epsilon_1 > 1$.

(A7) The kernel function $g(\cdot)$ is bounded and Lipschitz continuous on the support $[0, 1]$, $g(1) > 0$, and it has a continuous bounded derivative on $(0, 1)$ such that $g'(u) \le 0$ for all $0 < u < 1$.

The assumptions (A1)–(A7) are adopted from [19, 20] to establish the asymptotic property of kernel regression estimators when the regressor is valued in a semi-metric space. Here, (A1)–(A3) and (A7) are standard conditions for obtaining pointwise convergence rates, where (A1) specifies the smoothness of the regression function, and (A2) is the sole assumption concerning the response variable, describing moment existence. (A4)–(A6) are designed to ensure uniform convergence rates over $S_{\mathcal{Z}}$.

Theorem 2. *Suppose assumptions (A1)–(A7) hold, $h \to 0$ as $n \to \infty$. Then, for an i.i.d. random sample $\{(X_i, Y_i, Z_i)\}_{i=1}^n$, we have*

$$\lim_{n\to\infty} \mathrm{pMDAD}_n(Y \mid X, Z) = \mathrm{pMDAD}(Y \mid X, Z),$$

almost surely.

Obviously, for hypothesis test (1.1), the corresponding rejection region is $R^+ = \{\mathrm{pMDAD}_n(Y \mid X, Z) > c_n\}$, where c_n is the critical value. In order to validate the test further, we provide the following theorem.

Theorem 3. *Suppose assumptions (A1)–(A7) hold and $h \to 0$ as $n \to \infty$. Then, under H_1,*

$$n \cdot \mathrm{pMDAD}_n(Y \mid X, Z) \stackrel{p}{\longrightarrow} \infty,$$

as $n \to \infty$, where $\stackrel{p}{\longrightarrow}$ denotes convergence in probability.

The theorems demonstrate the test based on $\mathrm{pMDAD}_n(Y \mid X, Z)$ will be consistent against all fixed alternatives.

4. Implementation

Define the test statistic for the hypothesis (1.1) as

$$T_n = n \cdot \mathrm{pMDAD}_n(Y \mid X, Z), \tag{4.1}$$

and reject H_0 for its large value. To address this issue and efficiently implement the test, we employ a permutation method to approximate the p-value. Algorithm 1 summarizes the computational procedure.

Algorithm 1 computing p-value for the permutation test

Require: An i.i.d. random sample $\{(X_i, Y_i, Z_i), i \in \{1, \cdots, n\}\}$, Number of permutations B.
Ensure: The approximated p-value: $\hat{p} = \frac{1+\sum_{b=1}^{B} I(T_n^{(b)} \geq T_n)}{B+1}$
for $\{(X_i, Y_i, Z_i), i \in \{1, \cdots, n\}\}$, compute the statistic T_n using (4.1).
for b=1:B **do**
  generate the permuted sample of $\mathbb{X}_n = (X_1, \cdots, X_n)$ as $\mathbb{X}_n^{(b)} = \{X_1^{(b)}, \cdots, X_n^{(b)}\}$.
  compute $T_n^{(b)}$ from $\{(X_i^{(b)}, Y_i, Z_i), i \in \{1, \cdots, n\}\}$ using (4.1).
end for

If $\hat{p}$ is less than the given significance level α, the null hypothesis H_0 is rejected. In our simulation experiments, the adaptive number of permutations is set to $B = \min(\lfloor 200 + 5000/n \rfloor, 400)$, where n is the sample size [21]. Such a setting can not only fulfill the requirement of adaptive permutation number but also ensure computational efficiency when the sample size is small.

5. Simulations

In this section, we will illustrate the finite sample performance through simulations.

The proposed test is denoted as pMDAD, and the test in [10] is denoted as pMDD. When pMDAD is applied to functional data, we choose the Fourier basis as the eigenfunctions of the covariance operator when the functional data exhibit periodic features; otherwise, we select the spline basis [17]. The eigenvalues are set as $\lambda_i = 1/i^a$, and we let $a = 3$ in our paper. The number of eigenvalues can be set differently in different situations.

As for the estimation of $\mathbb{E}(Y \mid z)$ in our proposed tests, we choose $g(h^{-1}\|z - z'\|)) = \exp(-h^{-1}\|z - z'\|^2)$ with the optimal bandwidth h selected by minimizing the generalized cross-validation (GCV) function [9]:

$$\mathrm{GCV}(h) = \frac{\sum_{i=1}^{n}[\hat{\mathbb{E}}(Y \mid Z_i) - Y_i]^2}{n[1 - \mathrm{tr}(G)/n]^2},$$

where $\mathrm{tr}(G)$ is the trace of the matrix $G = (g_{ij})_{n\times n}$ with $g_{ij} = g(h^{-1}\|Z_i - Z_j\|) \,/\, \sum_{l=1}^{n} g(h^{-1}\|Z_i - Z_l\|)$.

The following examples are adapted from Examples 5.1(1) and 5.3 in [10]. Under each setting with a significance level $\alpha = 0.05$, the empirical size or power (rejection rate) of the tests is recorded through 1000 repetitions.

Note that the proposed test can be used for both vector data and functional data. We first consider Euclidean space data, and put functional data examples in the following.

Example 1. *Let $X, Y, Z, e_1, e_2 \in \mathbb{R}^p$.*

(i) *Let X, Y, and Z be independent variables from $N(0, I_{p\times p})$.*
(ii) *Let $X_s = Z_s + e_{1,s}$ and $Y_s = 2Z_s + \cos(X_s) + e_{2,s}$, $s = 1, \cdots, p$, where $Z, e_1, e_2 \sim N(0, I_{p\times p})$. Here, the subscript s represents the s-th component of the vector.*
(iii) *Only replace Y_s in* (ii) *with $Y_s = \exp(Z_s) + \cos(X_s) + e_{2,s}$, $s = 1, \cdots, p$*
(iv) *Only replace Y_s in* (ii) *with $Y_s = 2Z_s + 0.8\cos(X_s) \times e_{2,s}$, $s = 1, \cdots, p$.*
(v) *Only replace Y_s in* (ii) *with $Y_s = 2Z_s + 0.6X_s^2 \times e_{2,s}$, $s = 1, \cdots, p$.*

Table 1 presents the sizes and powers of pMDAD and pMDD in all scenarios for different dimensions ($p \in \{5, 15, 25, 50, 100\}$). As can be seen, in scenario (i), the empirical sizes of both tests are close to the nominal significance level. As for power, both tests show an increase as the sample size increases. Totally, in scenarios (ii) to (v), when detecting nonlinear relationships between Y and X, the proposed test outperforms the pMDD test. Moreover, it can be observed that under the given settings, the power of the proposed test is not sensitive to variations in the dimensions of X and Z, while the pMDD test may lose some power when p becomes larger. This might be owing to the projection employed in the construction of pMDAD.

In the following examples, we consider the performance of the proposed tests for functional data. All the functional variables are defined on the interval [0, 1] and observed at 201 equally spaced points.

Example 2. *Let $X(t) = \sum_{s=1}^{50} \xi_s\phi_s(t)$, $Y(t) = \sum_{s=1}^{50} \gamma_s\phi_s(t)$, $Z(t) = \sum_{s=1}^{50} \eta_s\phi_s(t)$, where $\phi_s(t) = \sqrt{2}\cos(s\pi t)$. For $m > 0$, we consider the following scenarios.*

(i) *Let ξ_s, γ_s, and η_s be independent random variables distributed as standard normal distribution.*
(ii) *Let $\xi_s = \exp(\eta_s) + e_{1s}$ and $\gamma_s = \eta_s + \sin(\xi_s) + e_{2s}$, where e_{1s} and e_{2s} are independent random variables following the standard normal distribution, for $s = 1, \cdots, 50$.*
(iii) *The setting is the same as in* (ii), *except that $\gamma_s = \eta_s + \xi_s^2 + e_{2s}$, $s = 1, \cdots, 50$.*
(iv) *The setting is the same as in* (ii), *except that $\gamma_s = \eta_s + \cos(\xi_s) \times e_{2s}$, $s = 1, \cdots, 50$.*

Table 1. Empirical sizes and powers in Example 1 with $n = 30, 50$ at significance level 0.05.

Scenario	p	n = 30		n = 50	
		pMDAD	pMDD	pMDAD	pMDD
(i)	5	0.041	0.049	0.044	0.045
	15	0.060	0.053	0.053	0.047
	25	0.041	0.045	0.057	0.060
	50	0.056	0.058	0.050	0.049
	100	0.050	0.050	0.056	0.056
(ii)	5	0.979	0.152	1.000	0.220
	15	0.999	0.074	1.000	0.094
	25	1.000	0.056	1.000	0.072
	50	1.000	0.047	1.000	0.049
	100	0.999	0.026	1.000	0.043
(iii)	5	0.769	0.105	0.986	0.136
	15	0.855	0.055	0.998	0.094
	25	0.847	0.052	1.000	0.057
	50	0.779	0.040	0.998	0.045
	100	0.634	0.022	1.000	0.046
(iv)	5	0.770	0.097	1.000	0.099
	15	1.000	0.068	1.000	0.082
	25	1.000	0.050	1.000	0.061
	50	1.000	0.040	1.000	0.048
	100	1.000	0.027	1.000	0.038
(v)	5	0.680	0.173	0.934	0.208
	15	0.742	0.097	0.982	0.133
	25	0.718	0.075	0.992	0.094
	50	0.641	0.058	0.997	0.078
	100	0.591	0.045	0.999	0.053

Table 2. Empirical sizes and powers in Example 2 with $n = 10, 30$ at significance level 0.05.

Scenario	n = 10		n = 30	
	pMDAD	pMDD	pMDAD	pMDD
(i)	0.046	0.042	0.061	0.035
(ii)	0.648	0.009	0.979	0.003
(iii)	0.659	0.986	0.971	1.000
(iv)	0.645	0.033	0.982	0.030
(v)	0.642	0.187	0.987	0.353

Table 3. Empirical sizes and powers in Example 3 with $n = 30, 50, 100$ at significance level 0.05.

Scenario	n = 30		n = 50		n = 100	
	pMDAD	pMDD	pMDAD	pMDD	pMDAD	pMDD
(i)	0.049	0.047	0.051	0.062	0.040	0.047
(ii)	0.420	0.744	0.525	0.924	0.666	0.983
(iii)	0.414	0.080	0.514	0.100	0.672	0.140
(iv)	0.425	0.058	0.557	0.092	0.665	0.132
(v)	0.434	0.062	0.511	0.092	0.686	0.125

(v) *The setting is the same as in* (ii)*, except that* $\gamma_s = \eta_s + \xi_s^3 \times e_{2s}$*,* $s = 1, \cdots, 50$.

Under the setting of Example 2, the null hypothesis H_0 holds exclusively in scenario (i). Based on [17], we choose the Fourier basis as the eigenfunctions and set the number of basis as 101 for pMDAD, i.e., $i = 1, 2, \cdots, 101$. Table 2 shows the empirical sizes and powers when $n = 10, 30$. It can be seen that the empirical sizes in scenario (i) of both tests are close to the nominal significance level. In the other scenarios almost all the powers are strengthened with the increase of sample size. The test we propose is generally better than pMDD, which loses power in most scenarios.

Example 3. *Let* $Z(t)$ *be generated by Gaussian process with exponential variogram* (*a Gaussian process with mean 0 and covariance* $E\{X(s)X(t)\} = \exp(-5|s-t|)$, $s, t \in [0, 1]$). *Then, we consider the following scenarios:*

(i) *Let* $X(t), Y(t)$ *be generated independently by Gaussian process with exponential variogram.*
(ii) *Let* $X(t) = 2Z(t) + \epsilon_1(t)$*,* $Y(t) = \exp(Z(t)) + \exp(X(t)) + \epsilon_2(t)$*, where* $\epsilon_1(t), \epsilon_2(t)$ *are generated by Wiener processes.*
(iii) *The setting is the same as scenario* (ii)*, except that* $Y(t) = \exp(Z(t)) + \cos(X(t)) + \epsilon_2(t)$.
(iv) *The setting is the same as scenario* (ii)*, except that* $Y(t) = \exp(Z(t)) + X(t) \times \epsilon_2(t)$.
(v) *The setting is the same as scenario* (ii)*, except that* $X(t) = \cos(Z(t)) + \epsilon_1(t)$ *and* $Y(t) = \exp(Z(t)) + \cos(X(t)) \times \epsilon_2(t)$.

In this example, we select the spline basis as the eigenfunctions and set the number of basis functions as 100 for pMDAD, as suggested by [17]. The results for $n = 30, 50, 100$ are presented in Table 3. It is shown that the empirical sizes in scenario (i) of both tests are close to the nominal significance level, and almost all the power is strengthened as the increase of sample size in the other scenarios. In scenarios (iii)–(v), the pMDAD test performs well; however, pMDD almost always fails when the sample size is small.

6. Real data analysis

We introduce the application of the proposed statistic through two real datasets in this section.

6.1. The CONTENT child growth data

The CONTENT child growth study was funded by the Sixth Framework Programme of the European Union, Project CONTENT (INCO-DEV-3-032136), which can be found in the *R* package *refund*. This dataset contains variables bmi $Y(t)$, height $X_1(t)$, and weight $X_2(t)$ recorded at irregular time points between 0 and 701 days after birth for 127 children. The number of data points recorded for each child ranges from 16 to 38, indicating that the observations for each individual are sparse. BMI $Y(t)$ is an indicator derived from height $X_1(t)$ and weight $X_2(t)$ ($Y(t) = X_2(t)/X_1^2(t)$) that can be used to assess whether body weight is healthy. By letting BMI $Y(t)$ be the response variable, we can consequently validate the applicability of the proposed statistical test by investigating whether height $X_1(t)$ or weight $X_2(t)$ can be omitted when constructing the regression model (i.e., conducting the omitted variable test).

Because the data are observed sparsely, we first use the method introduced by [22] to recover the curves. The relevant tools can be found in the *R* package *fdapace*. Under the setting of the adaptive permutation method, when controlling for height $X_1(t)$, the resulting p-values are 0.004 (pMDAD) and 0.004 (pMDD), which means the variable weight cannot be omitted. This aligns with the standard BMI formula. When controlling for weight $X_2(t)$, the resulting p-values are 0.004 (pMDAD) and 1.000 (pMDD), respectively. Given the BMI formula, we can conclude that the proposed test exhibits superior nonlinear detection capability.

6.2. The dietary calcium absorption data

In this subsection, we apply the proposed test to a real dataset, the dietary calcium absorption data from [23], to perform a significance test.

The dataset includes four variables: calcium absorption (calabs) $Y(t)$, dietary calcium intake (caldiet) $X_1(t)$, body surface area (bsa) $X_2(t)$, and body mass index (bmi) $X_3(t)$ for 188 patients. Measurements were taken at irregular time points between 35 and 64 years of age. Each patient had a varying number of repeated measurements, ranging from 1 to 4, which indicates sparse observations for each individual.

Let calabs $Y(t)$ be the response variable. [23] used their proposed method to screen out caldiet $X_1(t)$ as an important variable. We aim to test whether bsa $X_2(t)$ or bmi $X_3(t)$ has a significant influence on calabs $Y(t)$ after controlling for caldiet $X_1(t)$.

We first recover the curves and use 100 spline bases as the eigenfunctions for pMDAD. Under 225 permutations, we obtain the following p-values: 0.714 for pMDAD$(Y \mid X_2, X_1)$, 0.004 for pMDD$(Y \mid X_2, X_1)$, 0.758 for pMDAD$(Y \mid X_3, X_1)$, and 0.753 for pMDD$(Y \mid X_3, X_1)$. This indicates that, after selecting the variable caldiet $X_1(t)$, the influence of bsa $X_2(t)$ on calabs $Y(t)$ requires further clarification, while bmi $X_3(t)$ shows no significant effect on calcium absorption calabs $Y(t)$.

7. Conclusions

This paper proposes a new test, partial martingale difference-angle divergence, for testing and measuring the (in)dependence of partial conditional mean for Hilbert space random elements. We develop an estimator and present its large-sample properties. The proposed test exhibits superior power in detecting nonlinear relationships compared to pMDD. A notable feature of the test is that the regression estimator is given by the adaptive kernel weighting approach, thereby eliminating inaccuracies stemming from misspecified models. Furthermore, how to test and measure partial conditional mean (in)dependence in more general spaces, such as Banach spaces, is well worthy of further research.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work is partly supported by National Natural Science Foundation of China with No.12271014.

Conflict of interest

There is no interest conflict with others in the research work.

References

1. M. A. Delgado, W. G. Manteiga, Significance testing in nonparametric regression based on the bootstrap, *Ann. Stat.*, **29** (2001), 1469–1507. https://doi.org/10.1214/aos/1013203462
2. Y. Fan, Q. Li, Consistent model specification tests: omitted variables and semiparametric functional forms, *Econometrica*, **64** (1996), 865–890. https://doi.org/10.2307/2171848
3. X. Shao, J. Zhang, Martingale difference correlation and its use in high-dimensional variable screening, *J. Am. Stat. Assoc.*, **109** (2014), 1302–1318. https://doi.org/10.1080/01621459.2014.887012
4. C. E. Lee, X. Zhang, X. Shao, Testing conditional mean independence for functional data, *Biometrika*, **107** (2020), 331–346. https://doi.org/10.1093/biomet/asz070
5. T. Lai, Z. Zhang, Y. Wang, A kernel-based measure for conditional mean dependence, *Comput. Stat. Data Anal.*, **160** (2021), 107246. https://doi.org/10.1016/j.csda.2021.107246
6. T. Lai, Z. Zhang, Local influence detection of conditional mean dependence, *Commun. Math. Stat.*, **11** (2023), 1–27. https://doi.org/10.1007/s40304-023-00365-3
7. T. Zhou, L. Zhu, C. Xu, R. Li, Model-free forward screening via cumulative divergence, *J. Am. Stat. Assoc.*, **115** (2020), 1393–1405. https://doi.org/10.1080/01621459.2019.1632078
8. L. Su, X. Zheng, A martingale-difference-divergence-based test for specification, *Econ. Lett.*, **156** (2017), 162–167. https://doi.org/10.1016/j.econlet.2017.05.002

9. Z. Tian, T. Lai, Z. Zhang, Variation of conditional mean and its application in ultra-high dimensional feature screening, *Commun. Stat. Theory Methods*, **54** (2024), 352–382. https://doi.org/10.1080/03610926.2024.2310690

10. T. Park, X. Shao, S. Yao, Partial martingale difference correlation, *Electron. J. Stat.*, **9** (2015), 1492–1517. https://doi.org/10.1214/15-EJS1047

11. G. J. Székely, M. L. Rizzo, Partial distance correlation with methods for dissimilarities, *Ann. Stat.*, **42** (2014), 2382–2412. https://doi.org/10.1214/14-AOS1255

12. L. Cai, X. Guo, W. Zhong, Test and measure for partial mean dependence based on machine learning methods, *J. Am. Stat. Assoc.*, **120** (2025), 833–845. https://doi.org/10.1080/01621459.2024.2366030

13. A. R. Lundborg, I. Kim, R. D. Shah, R. J. Samworth, The projected covariance measure for assumption-lean variable significance testing, *Ann. Stat.*, **52** (2024), 2851–2878. https://doi.org/10.1214/24-AOS2447

14. B. D. Williamson, P. B. Gilbert, M. Carone, N. Simon, Nonparametric variable importance assessment using machine learning techniques, *Biometrics*, **77** (2021), 9–22. https://doi.org/10.1111/biom.13392

15. B. Dai, X. Shen, W. Pan, Significance tests of feature relevance for a black-box learner, *IEEE Trans. Neural Netw. Learn. Syst.*, **35** (2022), 1898–1911. https://doi.org/10.1109/TNNLS.2022.3185742

16. B. D. Williamson, P. B. Gilbert, N. R. Simon, M. Carone, A general framework for inference on algorithm-agnostic variable importance, *J. Am. Stat. Assoc.*, **118** (2023), 1645–1658. https://doi.org/10.1080/01621459.2021.2003200

17. T. Lai, Z. Zhang, Y. Wang, L. Kong, Testing independence of functional variables by angle covariance, *J. Multivar. Anal.*, **182** (2021), 104711. https://doi.org/10.1016/j.jmva.2020.104711

18. Y. Zhou, Y. Zhang, L. Zhu, A projective approach to conditional independence test for dependent processes, *J. Bus. Econ. Stat.*, **40** (2022), 398–407. https://doi.org/10.1080/07350015.2020.1826952

19. F. Ferraty, A. Mas, P. Vieu, Nonparametric regression on functional data: inference and practical aspects, *Aust. N. Z. J. Stat.*, **49** (2007), 267–286. https://doi.org/10.1111/j.1467-842X.2007.00480.x

20. F. Ferraty, A. Laksaci, A. Tadj, P. Vieu, Kernel regression with functional response, *Electron. J. Stat.*, **5** (2011), 159–171. https://doi.org/10.1214/11-EJS600

21. Z. Jin, D. S. Matteson, Generalizing distance covariance to measure and test multivariate mutual dependence via complete and incomplete V-statistics, *J. Multivar. Anal.*, **168** (2018), 304–322. https://doi.org/10.1016/j.jmva.2018.08.006

22. F. Yao, H. G. Müller, J. L. Wang, Functional data analysis for sparse longitudinal data, *J. Am. Stat. Assoc.*, **100** (2005), 577–590. https://doi.org/10.1198/016214504000001745

23. R. Ghosal, A. Maity, T. Clark, S. B. Longo, Variable selection in functional linear concurrent regression, *J. R. Stat. Soc. Ser. C Appl. Stat.*, **69** (2020), 565–587. https://doi.org/10.1111/rssc.12408

24. N. Pfister, P. Bühlmann, B. Schölkopf, J. Peters, Kernel-based tests for joint independence, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **80** (2017), 5–31. https://doi.org/10.1111/rssb.12235

Appendix

Proof of Theorem 2. Denote $\mathbb{E}_i = \mathbb{E}(Y \mid Z_i), \hat{\mathbb{E}}_i = \hat{\mathbb{E}}(Y \mid Z_i)$, and by a direct calculation, we can get

$$\begin{aligned} p_{ij} &= \langle \hat{U}_i, \hat{U}_j \rangle \\ &= \langle \hat{U}_i - U_i + U_i, \hat{U}_j - U_j + U_j \rangle \\ &= \langle U_i, U_j \rangle + \langle \hat{U}_i - U_i, U_j \rangle + \langle U_i, \hat{U}_j - U_j \rangle + \langle \hat{U}_i - U_i, \hat{U}_j - U_j \rangle \\ &= \langle U_i, U_j \rangle + \langle \mathbb{E}_i - \hat{\mathbb{E}}_i, U_j \rangle + \langle U_i, \mathbb{E}_j - \hat{\mathbb{E}}_j \rangle + \langle \mathbb{E}_i - \hat{\mathbb{E}}_i, \mathbb{E}_j - \hat{\mathbb{E}}_j \rangle \\ &=: p_{ij}^{(0)} + p_{ij}^{(1)} + p_{ij}^{(2)} + p_{ij}^{(3)}, \end{aligned}$$

where $p_{ij}^{(0)} = \langle U_i, U_j \rangle$, $p_{ij}^{(1)} = \langle \mathbb{E}_i - \hat{\mathbb{E}}_i, U_j \rangle$, $p_{ij}^{(2)} = \langle U_i, \mathbb{E}_j - \hat{\mathbb{E}}_j \rangle$ and $p_{ij}^{(3)} = \langle \mathbb{E}_i - \hat{\mathbb{E}}_i, \mathbb{E}_j - \hat{\mathbb{E}}_j \rangle$. Then, pMDAD_n can be rewritten in the following form,

$$\begin{aligned} &\text{pMDAD}_n(Y \mid X, Z) \\ =&\frac{1}{n^2}\sum_{i,j=1}^{n} p_{ij}a_{ij} \\ =&\frac{1}{n^2}\sum_{i,j=1}^{n} (p_{ij}^{(0)} + p_{ij}^{(1)} + p_{ij}^{(2)} + p_{ij}^{(3)})a_{ij} \\ =&\frac{1}{n^2}\sum_{i,j=1}^{n} p_{ij}^{(0)}a_{ij} + \frac{1}{n^2}\sum_{i,j=1}^{n} p_{ij}^{(1)}a_{ij} + \frac{1}{n^2}\sum_{i,j=1}^{n} p_{ij}^{(2)}a_{ij} + \frac{1}{n^2}\sum_{i,j=1}^{n} p_{ij}^{(3)}a_{ij} \\ =&:\text{pMDAD}_n^{(0)} + \text{pMDAD}_n^{(1)} + \text{pMDAD}_n^{(2)} + \text{pMDAD}^{(3)}. \end{aligned} \tag{A.1}$$

Step 1: Negligibility of $\text{pMDAD}_n^{(1)}$, $\text{pMDAD}_n^{(2)}$.
Under assumptions (A1)–(A7), we have

$$\|\mathbb{E}_j - \hat{\mathbb{E}}_j\| = O(h^b) + O_{a.co.}\Big(\sqrt{\frac{\Psi_{S_Z}(\frac{\log n}{n})}{n\psi(h)}}\Big),$$

uniformly with respect to Z_i from Remark 2 in [20], where *a.co.* means almost complete convergence. This implies that $\|\mathbb{E}_j - \hat{\mathbb{E}}_j\| = o(1)$ holds almost surely, due to the condition $h \to 0$ and (A.2) in [20]. By the Cauchy-Schwarz inequality,

$$|p_{ij}^{(1)}| \le \|\mathbb{E}_i - \hat{\mathbb{E}}_i\|\|U_j\|.$$

Under the assumption (A2), for $\text{pMDAD}_n^{(1)}$, it holds that

$$\frac{1}{n^2}\sum_{i,j=1}^{n} p_{ij}^{(1)}a_{ij} \le C\Big(\frac{\sum\limits_{i=1}^{n}\|\mathbb{E}_i - \hat{\mathbb{E}}_i\|}{n}\Big)\Big(\frac{\sum_{j=1}^{n}\|U_j\|}{n}\Big) = o(1),$$

almost surely, where $C = \max\{a_{ij}, 1 \le i, j \le n\}$. Then, we obtain

$$\text{pMDAD}_n^{(1)} \to 0, \ \text{a.e.}$$

Similarly, we have $\text{pMDAD}_n^{(2)} \to 0$, a.e.
Step 2: Negligibility of $\text{pMDAD}_n^{(3)}$.
By a direct calculation and using the Cauchy-Schwarz inequality twice, we have

$$\begin{aligned}|p_{ij}^{(3)}| &\le (\|\hat{\mathbb{E}}_i - E_i\|^2 + \|\hat{\mathbb{E}}_j - \mathbb{E}_j\|^2)/2\\ &= O(h^{2b}) + O_{a.co.}(\frac{\Psi_{S_Z}(\frac{\log n}{n})}{n\psi(h)}),\end{aligned}$$

uniformly with respect to (Z_i, Z_j). Thus $|p_{ij}^{(3)}| = o(1)$, almost surely uniformly in i, j.
For $\text{pMDAD}_n^{(3)}$, we have

$$\frac{1}{n^2}\sum_{i,j=1}^{n} p_{ij}^{(3)} a_{ij} \le C\Big(\frac{\sum_{i=1}^{n} \|\mathbb{E}_i - \hat{\mathbb{E}}_i\|^2}{n}\Big) = o(1).$$

Then,

$$\text{pMDAD}_n^{(3)} \to 0, \text{ a.s.}$$

Step 3: Asymptotic property of $\text{pMDAD}_n^{(0)}$.
Let $M_i = (X_i, Y_i, Z_i)$. Recall that

$$\text{pMDAD}_n^{(0)} = \frac{1}{n^2}\sum_{i,j=1}^{n} p_{ij}^{(0)} a_{ij} := \frac{1}{n^2}\sum_{i,j=1}^{n} \eta(M_i, M_j). \tag{A.2}$$

Then, $\text{pMDAD}_n^{(0)}$ is a V-statistic with core function $\eta(M_i, M_j)$ of degree 2. Finally, by the strong law of large numbers for V-statistics (Theorem C.4 in [24]), and combining Steps 1 and 2, we have $\text{pMDAD}_n(Y \mid X, Z)$ converges to pMDAD almost surely.

Proof of Theorem 3. By Proposition 1, $\text{pMDAD}(Y \mid X, Z) > 0$ under H_1. Then, in combination with Theorem 2, this conclusion is obvious.