5-2017

A Martingale Difference-Divergence-based test for specification

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Xin ZHENG
Singapore Management University, xin.zheng.2015@phdecons.smu.edu.sg

DOI: https://doi.org/10.1016/j.econlet.2017.05.002

Published in Economics Letters, 2017 May, Volume 156, Pages 162-167
http://doi.org/10.1016/j.econlet.2017.05.002

A martingale-difference-divergence-based test for specification☆

Liangjun Su*, Xin Zheng

School of Economics, Singapore Management University, Singapore

HIGHLIGHTS

- We propose a novel martingale-difference-divergence-based test for specification.
- The test does not require any nonparametric estimation.
- The test is applicable even if we have many covariates in the regression model.
- The test has superb finite sample performance and dominates its competitors.

ARTICLE INFO

Article history:
Received 20 January 2017
Received in revised form
30 March 2017
Accepted 5 May 2017
Available online 12 May 2017

JEL classification:
C12
C14
C21

Keywords:
Distance covariance
Integrated conditional moment test
Martingale difference divergence
Martingale difference correlation
Specification test

ABSTRACT

In this paper we propose a novel consistent model specification test based on the martingale difference divergence (MDD) of the error term given the covariates. The MDD equals zero if and only if error term is conditionally mean independent of the covariates. Our MDD test does not require any nonparametric estimation under the alternative and it is applicable even if we have many covariates in the regression model. We establish the asymptotic distributions of our test statistic under the null and a sequence of Pitman local alternatives converging to the null at the usual parametric rate. Simulations suggest that our MDD test has superb performance in terms of both size and power and it generally dominates several competitors. In particular, it is the only test that has well controlled size in the presence of many covariates and reasonable power against high frequency alternatives as well.

1. Introduction

In this paper we propose a new test for the correct specification of a parametric conditional mean model based on a variant of the *martingale difference divergence* (MDD hereafter) measure of conditional mean dependence between two random variables. In a sequence of papers, Székely et al. (2007), Székely and Rizzo (2009) and Székely and Rizzo (2014) propose to use distance covariance and distance correlation to measure the dependence between two random vectors which exhibit various nice properties. Such measures have been explored for feature screening in high dimensional regressions; see, e.g., Li et al. (2012). When one of the two random variables is a scalar, Shao and Zhang (2014, SZ hereafter) propose to use MDD to measure the conditional mean dependence of the scalar random variable given a random vector (see the definition of MDD in (2.4) in the next section). Like the relationship between covariance and correlation, the MDD can also be rescaled to ensure that it lies between 0 and 1, yielding the *martingale difference correlation* (MDC) measure of a scalar variable given a random vector. MDD measures the departure of the conditional mean independence between a scalar response variable and a vector of covariates, which is a natural extension of the distance correlation measure proposed by Székely et al. (2007). MDD and MDC have many nice properties. First, both of them are nonnegative and equal zero if and only if the scalar response variable is conditionally mean independent of the covariates. This suggests that we can propose a test for the conditional mean independence hypothesis which is widely used in econometrics and statistics. Second, both measures have a closed-form formula that is only involved with certain expectation and norm calculations so that they can be easily estimated from the data based on the sample analogue principle. Third, the measures are dimension-free in the sense that the dimension of the conditioning variable is allowed to be large but finite. Indeed, SZ use MDC as a method to conduct high-dimensional variable selection to screen out variables that do not contribute to the conditional mean of the response variable given the covariates.

One drawback of SZ's original MDD and MDC measure is that when they are used for variable screening, both the response variable and covariates need to be observed. Therefore, we propose a variant of MDD that is used to measure the conditional mean independence of a scalar random error term given the covariates. With this variant, we propose a new consistent test for the null hypothesis that a parametric conditional mean model is correctly specified. Under the null hypothesis, the error term from the correctly specified model is conditionally mean independent of the regressors in the model and has mean zero. Since the error term is not observed, we propose to estimate it from the null model and construct a test statistic based on the sample analogue of this new MDD measure. We study the asymptotic distributions of the test statistic under the null and under a sequence of Pitman local alternatives. Our test shares many nice properties that a typical nonsmoothing test might have. First, its limiting distribution under the null is a mixture of central chi-square distributions that is not asymptotically pivotal. So we propose a wild bootstrap method to obtain the bootstrap p-value or critical value. Second, our test has nontrivial asymptotic power against local alternatives converging to the null at the usual parametric rate. More importantly, our test is free of the choice of any smoothing parameter (e.g., the bandwidth in kernel-based tests or the number of sieve approximating terms in sieve-based tests) and it does not suffer from the curse of dimensionality associated with kernel- or sieve-based tests. In principle, our test works for any finite dimensional regression problem where the number of covariates, q, can be large. But for the derivation of our asymptotic distribution theory, we still need restrict q to be fixed. We conduct some Monte Carlo simulations and compare our test with some popular tests in the literature. Our simulation results indicate that our MDD-based test generally outperforms its competitors, especially for the case of high frequency alternatives and for the case of many covariates (e.g., $q = 10, 20$). To the best of our knowledge, this paper is the first to consider consistent model specification test in the presence of many covariates where existing tests tend to fail due to the notorious curse of dimensionality.

The rest of the paper is organized as follows. We introduce the hypotheses and the test statistic in Section 2. We study the asymptotic distributions of the test statistic under the null hypothesis and under a sequence of Pitman local alternatives in Section 3. We compare the MDD test with several popular tests through Monte Carlo simulations in Section 4. Section 5 concludes. The proofs of all results are relegated to the online supplementary Appendix.

Notation. For any matrix or vector A, $\|A\|$ denotes its Euclidean norm. The operators $\stackrel{p}{\to}$ and $\stackrel{d}{\to}$ denote convergence in probability and distribution, respectively.

2. The hypotheses and statistic

In this section we state the hypotheses and introduce the test statistic.

2.1. The hypotheses

We consider the following parametric regression model

$$Y_i = g(X_i; \beta) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where Y_i is a scalar dependent variable, X_i is a $q \times 1$ vector of covariates, β is a $d \times 1$ vector of unknown parameters, and ε_i is the unobserved error term. We assume that the functional form of $g(\cdot;\cdot)$ is known up to the finite dimensional parameter β. We are interested in testing the correct specification of $g(\cdot;\cdot)$. That is, we test the null hypothesis

$$\mathbb{H}_0 : P\left\{E(Y_i|X_i) = g(X_i; \beta_0)\right\} = 1 \quad \text{for some } \beta_0 \in \mathcal{B} \tag{2.2}$$

versus the alternative hypothesis

$$\mathbb{H}_1 : P\left\{E(Y_i|X_i) = g(X_i; \beta)\right\} < 1 \quad \text{for all } \beta \in \mathcal{B}, \tag{2.3}$$

where $\mathcal{B}$ is the parameter space.

2.2. Test statistic

To motivate our test statistic, we follow SZ and consider the MDD of ε given X whose square is defined by

$$\mathrm{MDD}(\varepsilon|X)^2 = \int_{\mathbb{R}^q} \left|\mathbb{E}\left[\varepsilon \exp(\mathbf{i}s'X)\right] - \mathbb{E}(\varepsilon)\,\mathbb{E}\left[\exp(\mathbf{i}s'X)\right]\right|^2 \times W(s)ds, \tag{2.4}$$

where $\mathbf{i} = \sqrt{-1}$, $W(s) = \frac{1}{c_q\|s\|^{(1+q)}}$, $c_q = \frac{\pi^{(1+q)/2}}{\Gamma((1+q)/2)}$, and $\Gamma(\cdot)$ is the complete gamma function: $\Gamma(z) = \int_0^\infty t^{z-1}\exp(-t)\,dt$. Let $\left(\varepsilon^\dagger, X^\dagger\right)$ be an independent copy of (ε, X). By Theorem 1 in SZ, we have

$$\mathrm{MDD}(\varepsilon|X)^2 = -\mathbb{E}\left\{[\varepsilon - \mathbb{E}(\varepsilon)]\left[\varepsilon^\dagger - \mathbb{E}\left(\varepsilon^\dagger\right)\right]\left\|X - X^\dagger\right\|\right\}, \tag{2.5}$$

and $\mathrm{MDD}(\varepsilon|X)^2 = 0$ if and only if $\mathbb{E}(\varepsilon|X) = \mathbb{E}(\varepsilon)$.

In our setup, ε denotes the error term in a regression such that $\mathbb{E}(\varepsilon) = 0$ is always maintained. This motivates us to consider the following variant of $\mathrm{MDD}(\varepsilon|X)^2$

$$\mathrm{MDD}^*(\varepsilon|X)^2 = -\mathbb{E}\left[\varepsilon\varepsilon^\dagger\left\|X - X^\dagger\right\|\right] + 2\mathbb{E}\left[\varepsilon\left\|X - X^\dagger\right\|\right]\mathbb{E}\left[\varepsilon^\dagger\right]. \tag{2.6}$$

The following proposition establishes the properties of $\mathrm{MDD}^*(\varepsilon|X)^2$ that serve as the basis of our test statistic.

Proposition 2.1. *Let $\left(\varepsilon^\dagger, X^\dagger\right)$ be an independent copy of (ε, X), where ε is a scalar random variable and X is a $q \times 1$ random vector. Suppose that $0 < \mathbb{E}\left[\varepsilon^2\right] < \infty$ and $0 < \mathbb{E}[\|X\|^2] < \infty$. Then*

(i) *$\mathrm{MDD}^*(\varepsilon|X)^2 \geq 0$;*
(ii) *$\mathrm{MDD}^*(\varepsilon|X)^2 = 0$ if and only if $\mathbb{E}(\varepsilon|X) = 0$ almost surely (a.s.).*

An important implication of Proposition 2.1 is that we can test (2.2) by testing whether $\mathrm{MDD}^*(\varepsilon_i|X_i)^2 = 0$, where $\varepsilon_i = Y_i - g(X_i; \beta_0)$. In practice, ε_i is not observed. We propose to estimate the model (2.1) by the nonlinear least squares (NLS) to obtain the NLS estimator $\hat{\beta}$ of β. Let $\hat{\varepsilon}_i = Y_i - g(X_i; \hat{\beta})$. We propose to estimate $n\mathrm{MDD}^*(\varepsilon|X)^2$ by the following object

$$T_n = -\frac{1}{n}\sum\sum_{1 \leq i \neq j \leq n} \hat{\varepsilon}_i\hat{\varepsilon}_j\kappa_{i,j} + \frac{2}{n}\sum\sum_{1 \leq i \neq j \leq n} \hat{\varepsilon}_i\kappa_{i,j}\frac{1}{n}\sum_{k=1}^n \hat{\varepsilon}_k, \tag{2.7}$$

where $\kappa_{i,j} \equiv \|X_i - X_j\|$. In the special case where $g(X_i; \beta)$ is linear in X_i and β, i.e., $g(X_i; \beta) = \left(1, X_i'\right)\beta$, we have $\sum_{i=1}^n \hat{\varepsilon}_i = 0$ and

$$T_n = -\frac{1}{n}\sum\sum_{1 \leq i \neq j \leq n} \hat{\varepsilon}_i\hat{\varepsilon}_j\kappa_{i,j} \equiv T_n^\ell. \tag{2.8}$$

Other than this case, $\sum_{i=1}^n \hat{\varepsilon}_i$ is generally nonzero and second term in (2.7) is necessary.

Remark 1. Interestingly, MDD $(\varepsilon|X)^2$ in (2.4) is closely related to Bierens' (1982) and Bierens and Ploberger's (1997) integrated conditional moment (ICM) test that takes the form

$$B = \int_{\mathbb{R}^q} \left| \mathbb{E}\left[\varepsilon \exp(\mathbf{i}s'\Phi(X)) \right] \right|^2 W_B(s)ds, \tag{2.9}$$

where $W_B(\cdot)$ is a nonnegative weight function and $\Phi(\cdot) : \mathbb{R}^q \to \mathbb{R}^q$ is a smooth function. But this test requires the delicate choices of both W_B and Φ and may not be tractable in practice; see Bierens (1990) and Bierens and Ploberger (1997). When $\mathbb{E}(\varepsilon) = 0$, we can also write B as

$$B^* = \int_{\mathbb{R}^q} \Big| \mathbb{E}\left[\varepsilon \exp(\mathbf{i}s'\Phi(X)) \right] - \mathbb{E}(\varepsilon)\,\mathbb{E}\left[\exp(\mathbf{i}s'\Phi(X)) \right] \Big|^2 W_B(s)ds. \tag{2.10}$$

Apparently, $B^* = \text{MDD}(\varepsilon|X)^2$ by choosing $\Phi(X) = X$ and $W_B(s) = W(s)$. In this case, we can regard MDD $(\varepsilon|X)^2$ as a special example of B. As a result, our test is tied closely to Bierens' ICM test as a nonsmoothing test.

Interestingly, as a referee kindly points out, our test can be regarded as a nonsmoothing version of Zheng's (1996) and Li and Wang's (1998) kernel-based smoothing test. There are two major differences between our test and theirs. First, we use the weigh function $\kappa_{i,j}$ in (2.7) whereas Zheng's and Li and Wang's tests use the kernel weight functions. Second, we have the second term in (2.7) which enforces $\mathbb{E}(\varepsilon) = 0$ while the latter tests do not require this term. This is simply because our test is based on the fact that $\mathbb{E}(\varepsilon|X) = \mathbb{E}(\varepsilon)$ and that $\mathbb{E}(\varepsilon) = 0$ under the null while existing tests are based on $\mathbb{E}(\varepsilon|X) = 0$ directly. In fact, Fan and Li (2000) conclude that the smoothing tests in Härdle and Mammen (1993), Zheng (1996), and Li and Wang (1998) are also closely related to the ICM test, although they are developed from completely different ideas. For more details, see Fan and Li (2000).

3. Asymptotic properties

In this section we study the asymptotic properties of T_n under the null hypothesis and under a sequence of Pitman local alternatives.

3.1. Basic assumptions

To facilitate the study of the local power property of our test, we consider the triangular array $\{(Y_{in}, X_{in}, \varepsilon_{in}),\ i = 1, \ldots, n\}$. Let $Q_n(\beta) = \frac{1}{n}\sum_{i=1}^n [Y_{in} - g(X_{in}; \beta)]^2$ and $Q(\beta) = \lim_{n\to\infty} \mathbb{E}[Y_{in} - g(X_{in}; \beta)]^2$. Let $g_{i\beta}(\beta) \equiv \partial g(X_{in}; \beta)/\partial\beta$, and $S(\beta) \equiv \lim_{n\to\infty} \mathbb{E}\left[g_{i\beta}(\beta) g_{i\beta}(\beta)'\right]$. Frequently we suppress the dependence of $(Y_{in}, X_{in}, \varepsilon_{in})$ on n.

We make the following assumptions.

Assumption A.1. (Y_{in}, X_{in}), $i = 1, 2, \ldots, n$, are independently and identically distributed (IID).

Assumption A.2. The NLS estimator $\hat{\beta}$ has the following representation

$$\hat{\beta} - \beta_0 = S^{-1}\frac{1}{n}\sum_{i=1}^n g_{i\beta}\varepsilon_i + o_P\left(n^{-1/2}\right),$$

where $g_{i\beta} = g_{i\beta}(\beta_0)$ and $S = S(\beta_0)$ is positive definite. There exists a constant $C \in (0, \infty)$ such that $E\left\| g_{i\beta} g_{i\beta}' \varepsilon_i^2 \right\| < C$.

Assumption A.3. (i) There exists a constant $C \in (0, \infty)$ such that $\mathbb{E}\left(\varepsilon_i^4\right) \le C$ and $\mathbb{E}\|X_i\|^4 \le C$.

(ii) There exists a positive definite matrix H such that

$$\sup_{\beta \in N_{\epsilon_n}(\beta_0)} \left\| \frac{1}{n}\sum_{i=1}^n \frac{\partial^2 g(X_{in}; \beta)}{\partial\beta\partial\beta'} - H \right\| = o_P(1),$$

where $N_\epsilon(\beta_0) = \{\beta \in \mathbb{B} : \|\beta - \beta_0\| \le \epsilon\}$ and $\epsilon_n = o(1)$.

(iii) $\frac{1}{n}\sum_{i=1}^n g_{i\beta} \xrightarrow{p} S_0$, $\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}\left(g_{i\beta}\kappa_{i,j}\right) \xrightarrow{p} S_1$, and $\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n g_{i\beta}g_{j\beta}'\kappa_{i,j} \xrightarrow{p} S_2$, where $S_0 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(g_{i\beta}\right)$, $S_1 = \lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}\left(g_{i\beta}\kappa_{i,j}\right)$, and $S_2 = \lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}\left(g_{i\beta}g_{j\beta}'\kappa_{i,j}\right)$.

We assume that the observations are IID in Assumption A.1 to facilitate the asymptotic analysis. We conjecture that our result below can be extended to allow for weakly dependent time series observations but restrict ourselves to IID observations for simplicity. Assumption A.2 requires $\hat{\beta}$ follow a Bahadur representation with certain well behaved influence function. One can verify A.2 under some primitive conditions given in the literature; see, e.g., Jennrich (1969), Wu (1981), and Amemiya (1985). Assumption A.3 imposes some additional conditions to study the asymptotic distribution of our test statistics. Assumption A.3(i) imposes some moment conditions for X_{in} and ε_{in}; Assumption A.3(ii) imposes uniform convergence of the Hessian function in the neighborhood of β_0; Assumption A.3(iii) imposes some convergence conditions associated with $g_{i\beta}$.

3.1.1. Asymptotic distribution under the null

The following theorem reports the asymptotic distribution of T_n.

Theorem 3.1. *Suppose that Assumptions A.1–A.3 hold. Then under* $\mathbb{H}_0$ *we have as* $n \to \infty$,

$$T_n \xrightarrow{d} \sum_{\nu=1}^\infty \lambda_\nu z_\nu^2,$$

where z_ν's are IID $N(0, 1)$, λ_ν's are the eigenvalues of the integral equation

$$\int_{-\infty}^{\infty} \varepsilon_2^2 h(X_1, X_2) f_\nu(X_2) dF(\xi_2) = \lambda_\nu f_\nu(X_1),$$

$\{\varepsilon_{in} f_\nu(X_{in})\}_{\nu=1}^\infty$ is an orthonormal sequence of eigenfunctions, $h(X_1, X_2)$ is defined in Equation A.2 in the Online Appendix, and $F(\cdot)$ denotes the limiting cumulative distribution function of $\xi_i \equiv \xi_{in} \equiv \left(\varepsilon_{in}, X_{in}'\right)'$.

The proof of Theorem 3.1 is tedious and the expression for $h(X_1, X_2)$ appears complicated. Since h depends on the underlying data generating process (DGP), T_n is not asymptotically pivotal under the null and thus we cannot tabulate its critical values. In the following we will propose a bootstrap method to obtain the bootstrap p-value to make statistical inference.

Apparently, T_n shares the same type of asymptotic null distribution as the ICM test. This is not surprising given Remark 1. As mentioned, our test does not need to specify a transformation function or weight function that an ICM test needs.

3.1.2. Local power analysis

To study the asymptotic local power of T_n, we consider the following sequence of Pitman local alternatives:

$$\mathbb{H}_1(n^{-1/2}) : \mathbb{E}(\varepsilon_{in}|X_{in}) = n^{-1/2}\delta(X_{in}) \quad \text{for all } i. \tag{3.1}$$

The next theorem describes the asymptotic distribution of MDD test under the above sequence of local alternatives.

Theorem 3.2. *Suppose Assumptions A.1–A.3 hold. Then under $\mathbb{H}_1(n^{-1/2})$, we have as $n \to \infty$,*

$$T_n \xrightarrow{d} \sum_{\nu=1}^{\infty} \lambda_\nu (z_\nu + a_\nu)^2,$$

where $a_\nu = \lim_{n\to\infty} \mathbb{E}[\delta(X_{in}) f_\nu(X_{in})]$ and $f_\nu(\cdot)$ is defined in Theorem 3.1.

Since $\{z_\nu\}_{\nu=1}^{\infty}$ are IID $N(0, 1)$, $(z_\nu + a_\nu)^2$ is stochastically larger than z_ν^2 for $a_\nu \neq 0$. This implies that our test has nontrivial asymptotic local power against local alternatives that converge to the null at rate $n^{-1/2}$. See Fan (1998) for a similar remark.

4. Monte Carlo simulation

In this section we conduct a sequence of Monte Carlo simulations to evaluate the finite sample performance of our test and compare it with some existing tests.

4.1. Data generating processes

We consider the following data generating processes:

$$\text{DGP1}(m) : Y_i = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ji} + \sigma_i^{(m)} \varepsilon_i,$$

$$\text{DGP2}(m) : Y_i = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ji} + n^{-1/2} \sum_{j=1}^{m} X_{ji}^2 + \sigma_i^{(m)} \varepsilon_i,$$

$$\text{DGP3}(m) : Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + 2\sin(mX_{1i})\sin(mX_{2i}) + \sigma_i \varepsilon_i.$$

DGP1(m) specifies m covariates and is used to evaluate the size performance of various tests. DGP2(m) specifies m covariates and is used to evaluate the local power of various tests. DGP3(m) specifies two covariates with m-dependent frequency under the alternative. We allow for conditional heteroskedasticity in all models and generate the covariates and heteroskedasticity as follows. In DGP1 and DGP2, when $m = 2$, $X_1 \sim U(0, 1)$, $X_2 \sim N(0, 1)$, and $\sigma^{(2)} = \{0.1 + X_1 + X_2^2\}^{1/2}$; when $m = 5$, $X_j \sim U(0, j)$ for $j = 1, 2, 3$, $X_j \sim N(0, (j-3)^2)$ for $j = 4, 5$, and $\sigma^{(5)} = \{0.1 + \sum_{j=1}^{3} X_j + \sum_{j=4}^{5} X_j^2\}^{1/2}$; when $m = 10$, $X_j \sim U(0, j)$ for $j = 1, \ldots, 5$, $X_j \sim N(0, (j-5)^2)$ for $j = 6, \ldots, 10$, and $\sigma^{(10)} = \{0.1 + \sum_{j=1}^{5} X_j + \sum_{j=6}^{10} X_j^2\}^{1/2}$; when $m = 20$, $X_j \sim U(0, j)$ for $j = 1 \ldots, 10$, $X_j \sim N(0, (j-10)^2)$ for $j = 11, \ldots, 20$, and $\sigma^{(20)} = \{0.1 + \sum_{j=1}^{10} X_j + \sum_{j=11}^{20} X_j^2\}^{1/2}$. In DGP3, $X_j \sim N(0, 1)$ for $j = 1, 2$ and $\sigma = \{0.1 + X_1^2 + X_2^2\}^{1/2}$. We specify $m = 1/2$, $m = 1$, and $m = 2$ in DGP3 (m), corresponding to low-, moderate-, and high-frequency alternatives, respectively. In all cases, we generate ε_i independently from the standard normal distribution and set β_j's to be 1.

We will test $\mathbb{H}_0 : E(Y_i|X_i) = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ji}$ for some $(\beta_0, \ldots, \beta_m)$ in DGP1 (m) and DGP2 (m) and $\mathbb{H}_0 : E(Y_i|X_i) = \beta_0 + \sum_{j=1}^{2} \beta_j X_{ji}$ for some $(\beta_0, \beta_1, \beta_2)$ in DGP3 (m).

4.2. Test statistics

We will implement our test statistic T_n and denote it as MDD in the following tables. For the purpose of comparison, we consider three popular tests for the correct specification of functional forms in the literature.

The first one is Zheng's (1996) and Li and Wang's (1998) residual-based test:

$$\text{Z\&LW test} : T_n^{Z\&LW} = \frac{1}{n(n-1)} \sum\sum_{1\le i\neq j\le n} \frac{1}{\Pi_{l=1}^{q} h_l} \times K\left(\frac{X_i - X_j}{h}\right) \hat{\varepsilon}_i \hat{\varepsilon}_j,$$

where $\hat{\varepsilon}_i$ is the residual from the parametric regression under the null, q denotes the dimension of X_i, $K(\cdot)$ is a product of univariate Epanechnikov kernel, $h = (h_1, \ldots, h_q)'$ is a bandwidth vector, and $a/b = (a_1/b_1, \ldots, a_q/b_q)'$ when $a = (a_1, \ldots, a_q)'$ and $b = (b_1, \ldots, b_q)'$ are both $q \times 1$ vectors.

The second one is Härdle and Mammen's (1993, HM) test that is based on the comparison of the nonparametric estimate and the smoothed parametric estimate of the conditional mean regression function under the null:

$$\text{HM test} : T_n^{HM} = n\left(\Pi_{l=1}^{q} h_l\right)^{1/2} \sum_{i=1}^{n} \left[\hat{g}_h(x_i) - \mathcal{K}_{h,n} g(x_i, \hat{\beta})\right]^2,$$

where, $\mathcal{K}_{h,n}$ denotes the smoothing operator

$$\mathcal{K}_{n,h} g(x, \hat{\beta}) = \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) g(X_i, \hat{\beta})}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)},$$

$\hat{\beta}$ denotes the least squares estimate of the regression coefficient under the null, $\hat{g}_h(x)$ is the Nadaraya–Watson kernel estimator of $E(Y_i|X_i = x)$ by using the kernel function $K(\cdot)$ and bandwidth h.

The last one is the ICM test Bierens and Ploberger's (1997) ICM test:

$$\text{ICM test} : T_n^{B} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{\varepsilon}_j \hat{\varepsilon}_i \times \prod_{k=1}^{q} \exp\left\{\left[\Phi(X_{ki}) + \Phi(X_{kj})\right]^2 / 2\right\},$$

where $\hat{\varepsilon}_i$ is the residual from the parametric regression under the null and Φ is a one-to-one mapping function from the support of X to itself: $\Phi(X_{li}) = \tan^{-1}((X_{li} - \bar{X}_l)/s_l)$, where $\bar{X}_l$ and s_l denotes the sample mean and sample standard deviation of $\{X_{li}\}_{i=1}^{n}$ with X_{li} being the lth component of X_i. Fan and Li (2000) also consider the above specification for the ICM test.

In all cases, we choose the bandwidth according to Silverman's rule of thumb: $h_l = 1.06 s_l n^{-1/(4+q)}$ for $l = 1, \ldots, q$. After suitable normalization, both $T_n^{Z\&LW}$ and T_n^{HM} are asymptotically standard normally distributed under the null and they can detect local alternatives converging to the null at the nonparametric rate. In contrast, the ICM test has asymptotic null distribution similar to our MDD test and it can detect local alternatives converging to the null at the usual parametric rate.

To implement all tests, we consider the wild bootstrap to obtain the bootstrap p-values despite the fact the two kernel-based tests are asymptotically $N(0, 1)$ under the null. The wild bootstrap procedure is the same as that in Wu (1986) and Härdle and Mammen (1993) and the justification of its asymptotic validity is standard. See, e.g., Su et al. (2015b) and Su et al. (2015a).

We will consider various sample sizes. When we have two covariates, we let n change from 50 to 400; when we have 5 or more covariates, we let n change from 200 to 800. The number of bootstrap resamples is 400 and the number of replications is 1000 in each scenario.

Table 1
Empirical size under the null hypothesis.

DGP	Level	0.1				0.05				0.01			
	n	MDD	Z&LW	HM	ICM	MDD	Z&LW	HM	ICM	MDD	Z&LW	HM	ICM
DGP1(2)	50	0.130	0.136	0.126	0.123	0.050	0.062	0.060	0.045	0.006	0.010	0.007	0.006
	100	0.119	0.117	0.115	0.116	0.060	0.063	0.054	0.054	0.012	0.013	0.011	0.006
	200	0.103	0.102	0.125	0.105	0.052	0.052	0.059	0.050	0.015	0.009	0.008	0.009
	400	0.096	0.091	0.096	0.102	0.039	0.040	0.041	0.052	0.009	0.008	0.008	0.007
DGP1(5)	200	0.131	0.125	0.079	0.021	0.067	0.054	0.040	0.006	0.021	0.008	0.005	0.000
	400	0.109	0.107	0.091	0.033	0.053	0.059	0.035	0.005	0.007	0.015	0.005	0.000
	800	0.108	0.099	0.103	0.038	0.057	0.047	0.045	0.008	0.009	0.008	0.006	0.000
DGP1(10)	200	0.113	0.129	0.003	0.000	0.047	0.071	0.000	0.000	0.013	0.017	0.000	0.000
	400	0.113	0.105	0.000	0.000	0.056	0.052	0.000	0.000	0.010	0.009	0.000	0.000
	800	0.092	0.096	0.000	0.000	0.040	0.049	0.000	0.000	0.009	0.009	0.000	0.000
DGP1(20)	200	0.202	0.183	0.006	0.000	0.083	0.098	0.000	0.000	0.007	0.023	0.000	0.000
	400	0.113	0.142	0.000	0.000	0.051	0.061	0.000	0.000	0.005	0.013	0.000	0.000
	800	0.120	0.125	0.000	0.000	0.052	0.062	0.000	0.000	0.007	0.010	0.000	0.000

Table 2
Empirical power under the local alternatives.

DGP	Level	0.1				0.05				0.01			
	n	MDD	Z&LW	HM	ICM	MDD	Z&LW	HM	ICM	MDD	Z&LW	HM	ICM
DGP2(2)	50	0.759	0.679	0.674	0.529	0.651	0.554	0.577	0.289	0.285	0.229	0.253	0.063
	100	0.812	0.726	0.749	0.570	0.724	0.636	0.640	0.395	0.395	0.308	0.354	0.090
	200	0.860	0.751	0.790	0.640	0.718	0.612	0.669	0.418	0.491	0.392	0.444	0.145
	400	0.846	0.772	0.799	0.641	0.746	0.660	0.691	0.445	0.501	0.396	0.449	0.185
DGP2(5)	200	0.801	0.634	0.448	0.109	0.692	0.501	0.295	0.025	0.454	0.259	0.115	0.001
	400	0.837	0.640	0.514	0.132	0.752	0.512	0.367	0.049	0.525	0.295	0.147	0.001
	800	0.838	0.612	0.474	0.135	0.743	0.490	0.343	0.051	0.516	0.262	0.160	0.006
DGP2(10)	200	0.979	0.880	0.038	0.000	0.953	0.786	0.006	0.000	0.820	0.522	0.000	0.000
	400	0.987	0.886	0.025	0.000	0.969	0.811	0.003	0.000	0.890	0.580	0.000	0.000
	800	0.977	0.834	0.023	0.000	0.961	0.757	0.001	0.000	0.873	0.548	0.000	0.000
DGP2(20)	200	0.709	0.420	0.006	0.000	0.545	0.283	0.000	0.000	0.248	0.092	0.000	0.000
	400	0.643	0.345	0.000	0.000	0.523	0.218	0.000	0.000	0.283	0.088	0.000	0.000
	800	0.611	0.325	0.000	0.000	0.492	0.216	0.000	0.000	0.279	0.075	0.000	0.000

Table 3
Empirical power under alternatives with different frequencies.

DGP	Level	0.1				0.05				0.01			
	n	MDD	Z&LW	HM	ICM	MDD	Z&LW	HM	ICM	MDD	Z&LW	HM	ICM
DGP3(1/2)	50	0.758	0.583	0.504	0.715	0.613	0.433	0.367	0.550	0.286	0.156	0.155	0.237
	100	0.962	0.865	0.806	0.906	0.901	0.764	0.687	0.833	0.688	0.493	0.421	0.548
	200	0.997	0.986	0.978	0.983	0.997	0.971	0.955	0.969	0.970	0.907	0.840	0.893
	400	1.000	1.000	0.999	0.999	1.000	0.999	1.000	0.998	0.999	0.998	0.994	0.993
DGP3(1)	50	0.984	0.792	0.900	0.848	0.964	0.675	0.820	0.768	0.852	0.439	0.615	0.603
	100	0.999	0.932	0.984	0.913	0.995	0.872	0.970	0.837	0.990	0.794	0.935	0.783
	200	1.000	0.988	1.000	0.973	1.000	0.978	1.000	0.949	1.000	0.955	0.997	0.893
	400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	1.000	0.998	1.000	0.988
DGP3(2)	50	0.766	0.184	0.489	0.170	0.600	0.117	0.307	0.091	0.226	0.041	0.109	0.013
	100	0.951	0.209	0.743	0.154	0.904	0.136	0.603	0.064	0.654	0.074	0.301	0.012
	200	1.000	0.318	0.974	0.142	0.997	0.213	0.933	0.069	0.982	0.120	0.732	0.010
	400	1.000	0.424	1.000	0.169	1.000	0.343	1.000	0.094	1.000	0.278	0.991	0.015

4.3. Simulation results

We report the simulation results in Tables 1–3 for DGP1 (m)-DGP3 (m), respectively, where the nominal significance levels are given by 0.01, 0.05, and 0.1. Table 1 reports the empirical levels of the four tests for DGP1 (m) with different numbers of covariates. The findings are interesting. First, when the number of covariates is small $(m = 2)$, all four tests perform quite well in terms of empirical level for the number of observations as small size as 50, and the empirical levels generally improve as n increases. Second, as m increases, the levels for both HM and ICM tests diminish rapidly to zero and the degeneracy of the levels does not improve when the sample size increases from 200 to 800. This indicates that either the HM test or the ICM test has severe size distortions due to the curse of dimensionality in nonparametrics. In particular, the HM test requires nonparametric estimation under the alternative. Third, both MDD and Z&LW tests perform very well unless m is too big (20) and n is small (200). As for the Z&LW test, even though it is a kernel-based nonparametric tests, it does not require the estimation of the regression model under the alternative. Perhaps, this explains why it is not sensitive to the number of covariates. Overall, our MDD dominates the other three tests in terms of empirical level.

Table 2 reports the empirical power for DGP2 (m) when m takes different values. We summarize some important findings from Table 2. First, the ICM test has reasonable power when $m = 2$. But as m increase, the ICM test does not have any power to detect local deviations from the null. It is even inferior to the two kernel-based tests (Z&LW and HM) which have power to detect local alternatives converging to the null at a slower rate than $n^{-1/2}$. This is due to the fact that the approximation of integrated moments in ICM would be inaccurate in the case of large m. Second, HM test has certain power when m increases from 2 to 5 but it loses power when m increases further. We conjecture that this is due to the fact that the implementation of HM test requires the nonparametric estimation of functional form which suffers curse of dimensionality. Third, as expected both MDD and Z&LW tests have power even in the presence of a large number of covariates. In general, our MDD test dominates the Z&LW test in terms of local empirical power. This is also consistent with the theory because our test can detect $n^{-1/2}$-local alternatives while Z&LW test can detect local alternatives converging to the null at a slower rate than $n^{-1/2}$. In sum, for the usual $n^{-1/2}$-local alternatives, our MDD test outperforms all of its competitors under investigation.

Table 3 reports the empirical power for DGP3 (m) when the alternatives are at different frequencies. First, when the frequency is low $(m = 1/2)$ or moderate $(m = 1)$, all four tests have reasonable power. Second, when the frequency is low and the sample size is small, the ICM test performs fairly well and it outperforms the Z&LW and HM tests. Third, the ICM test does not have power in the high-frequency case as expected. Fourth, our MDD test is almost always the best of all.

In summary, our MDD test generally has well-controlled size and it is not sensitive to the inclusion of many covariates in the regression model. It also has higher empirical power than its competitors against both local alternatives and global alternatives.

5. Conclusion

In this paper we have proposed a novel consistent model specification test based on the MDD of the error term given the covariates. The MDD equals zero if and only if error term is conditionally mean independent of the covariates. It does not require any nonparametric estimation under the null or alternative and is applicable even if we have many covariates in the regression model. We have established the asymptotic distributions of our test statistic under the null and a sequence of Pitman local alternatives converging to the null at the usual parametric rate. Simulations demonstrate that our MDD test has superb performance and generally dominates its competitors in a variety of scenarios.

Several extensions are possible. First, it is easy to extend our method to test the correct specification of a semiparametric models, e.g., partially linear, additive, or single index models. In this case, one needs to estimate the semiparametric model under the null and apply undersmoothing to ensure that the bias in the semiparametric estimation is asymptotically vanishing. Second, one can extend our test to test for the correct specification of a conditional mean model in panel data models where complications will arise due to the presence of unobserved individual heterogeneity. Third, we conjecture that it is also possible to extend the distance covariance or MDD to measure the dependence between two random vectors/variables conditional on a third one that is dimension-free. Recently there is a growing interest in testing conditional independence; see, e.g., Su and White (2007, 2008, 2014), Song (2009), Linton and Gozalo (2014), and Huang et al. (2016). But all of these tests are subject to the curse of dimensionality issue and are generally not applicable when the dimension of conditioning variable is large (e.g., larger than 6). So it is worthwhile to consider a dimension-free measure of conditional dependence based on which a sample analogue can be constructed and used to test for the null of conditional independence. We leave these topics for future research.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.econlet.2017.05.002.

References

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, Massachusetts.
Bierens, H.J., 1982. Consistent model specification tests. J. Econometrics 20, 105–134.
Bierens, H.J., 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443–1458.
Bierens, H.J., Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests. Econometrica 65, 1129–1151.
Fan, Y., 1998. Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. Econometric Theory 14, 604–621.
Fan, Y., Li, Q., 2000. Consistent model specification tests. Econometric Theory 16, 1016–1041.
Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Ann. Statist. 1926–1947.
Huang, M., Sun, Y., White, H., 2016. A flexible nonparametric test for conditional independence. Econometric Theory 32, 1434–1482.
Jennrich, R.I., 1969. Asymptotic properties of nonlinear least squares estimators. Ann. Math. Statist. 40, 633–643.
Li, Q., Wang, S., 1998. A simple consistent bootstrap test for a parametric regression function. J. Econometrics 87, 145–165.
Li, R., Zhong, W., Zhu, L., 2012. Feature screening via distance correlation learning. J. Amer. Statist. Assoc. 107, 1129–1139.
Linton, O., Gozalo, P., 2014. Testing conditional independence restrictions. Econometric Rev. 33, 523–552.
Shao, X., Zhang, J., 2014. Martingale divergence correlation and its use in high dimensional variable screening. J. Amer. Statist. Assoc. 109, 1302–1318.
Song, K., 2009. Testing conditional independence via Rosenblatt transform. Ann. Statist. 37, 4011–4045.
Su, L., Hoderlein, S., White, H., 2015a. Testing monotonicity in unobservables with panel data. Working Paper, Singapore Management University.
Su, L., Jin, S., Zhang, Y., 2015b. Specification test for panel data models with interactive fixed effects. J. Econometrics 186, 222–244.
Su, L., White, H., 2007. A consistent characteristic function-based test for conditional independence. J. Econometrics 141, 807–834.
Su, L., White, H., 2008. A nonparametric Hellinger metric test for conditional independence. Econometric Theory 24, 829–864.
Su, L., White, H., 2014. Testing conditional independence via empirical likelihood. J. Econometrics 182, 27–44.
Székely, G.J., Rizzo, M.L., 2009. Brownian distance covariance. Ann. Appl. Stat. 3, 1236–1265.
Székely, G.J., Rizzo, M.L., 2014. Partial distance correlation with methods for dissimilarities. Ann. Statist. 42, 2382–2412.
Székely, G.J., Rizzo, M.L., Bakirov, N.K., 2007. Measuring and testing dependence by correlation of distances. Ann. Statist. 35, 2769–2794.
Wu, C.-F., 1981. Asymptotic theory of nonlinear least squares estimation. Ann. Statist. 9, 501–513.
Wu, C.-F., 1986. Jackknife, bootstrap and other resampling methods in regression analysis. Ann. Statist. 14, 1261–1295.
Zheng, J.X., 1996. A consistent test of functional form via nonparametric estimation techniques. J. Econometrics 75, 263–289.