Journal of Applied Econometrics

WILEY

Journal of APPLIED ECONOMETRICS

RESEARCH ARTICLE OPEN ACCESS

Manipulation Test for Multidimensional RDD

Federico Crippa

Department of Economics, Northwestern University, Evanston, Illinois, USA

Correspondence: Federico Crippa (federicocrippa2025@u.northwestern.edu)

Received: 24 June 2024 | **Revised:** 11 February 2025 | **Accepted:** 9 April 2025

Keywords: manipulation test | multidimensional RDD | regression discontinuity design

ABSTRACT

The causal inference model for the regression discontinuity design (RDD) relies on assumptions that imply the continuity of the density of the assignment (running) variable. The test for this implication is commonly referred to as the manipulation test and is regularly reported in applied research to strengthen the design's validity. The multidimensional RDD (MRDD) extends the RDD to contexts where treatment assignment depends on several running variables. This paper introduces a manipulation test for the MRDD. First, it develops a theoretical model for causal inference with the MRDD, which is used to derive a testable implication on the conditional marginal densities of the running variables. Then, it constructs the test for the implication based on a quadratic form of a vector of statistics separately computed for each marginal density. Finally, the proposed test is compared with alternative procedures commonly employed in applied research.

JEL Classification: C12, C14

1 | Introduction

Regression discontinuity design (RDD) is widely used in policy evaluation and causal inference analysis to establish credible causal relationships under mild assumptions. RDD requires that units are assigned to a treatment based on some observable characteristic, the running variable: the probability of being treated must discontinuously change when the value of the running variable exceeds a certain threshold, called the cutoff. The fact that policies are often designed in this way (scholarship for students with GPA exceeding a threshold, welfare benefits for households with a certain income, etc.) explains RDD popularity.[1]

To identify the average treatment effect at the cutoff, the causal inference model for the RDD proposed by Lee (2008) relies on assumptions about unobservable potential outcomes. Although these assumptions are not directly testable, the model implies two testable conditions on observable quantities. The first condition requires continuity at the cutoff for the probability density function of the running variable, as discussed by McCrary (2008). The second condition imposes continuity at the cutoff for the conditional expectation (given the running variable) of additional observable characteristics measured before treatment. Tests of these conditions provide evidence supporting the validity of the RDD and are commonly reported in empirical applications, as highlighted in the survey by Canay and Kamat (2018). The test for the continuity of the running variable's density is commonly referred to as the manipulation test, as it checks whether units manipulate their scores to secure treatment assignments. Several manipulation tests have been proposed in the literature, starting with the seminal work by McCrary (2008) and followed by more recent approaches, such as those by Cattaneo et al. (2020) and Bugni and Canay (2021).

This paper introduces a manipulation test for the multidimensional RDD (MRDD), valid for both cases of perfect (sharp MRDD) and imperfect (fuzzy MRDD) compliance. MRDD is a model where the treatment assignment depends on multiple running variables. I consider the version of MRDD where the probability of receiving the treatment changes discontinu-

https://doi.org/10.1002/jae.3135

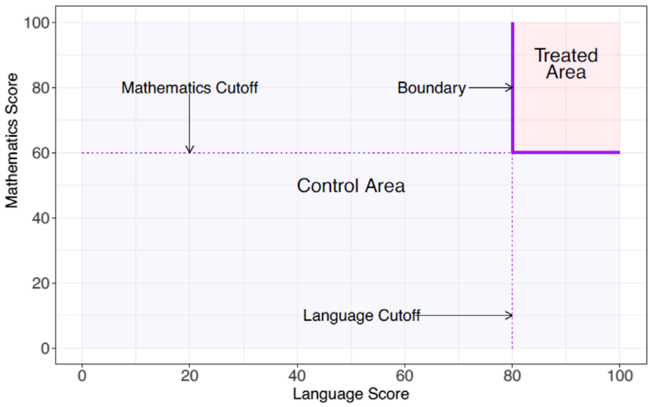

FIGURE 1 | MRDD running variables and thresholds. *Source:* Cattaneo et al. (2024).

ously when all the running variables exceed their cutoffs, and the cutoff of each running variable is fixed.[2] Compared to the single-dimensional RDD, the main novelty is that the cutoff is not a point in a single-dimensional space but a set of infinite points in the multidimensional space of the running variables. Consider, for example, a scholarship for students who score above certain thresholds in language and mathematics tests, as illustrated in Figure 1. The MRDD allows the researcher to identify and estimate the average effect of the scholarship on students with scores at the boundary. In this case, the cutoff is the solid purple boundary in the bidimensional space of language and math scores.

The main contributions of this paper are as follows: First, I extend the model of Lee (2008) to a multidimensional setting and demonstrate how one of the model's assumptions on unobservable quantities leads to a testable implication for the conditional marginal densities of the running variables, analogous to the condition proposed by McCrary (2008) for the density of the single running variable.

Second, I construct a manipulation test for this condition to help corroborate the credibility of the MRDD. Intuitively, the test procedure divides the space of running variables into subspaces where each running variable individually determines the treatment assignment. In each subspace, the model's implication requires the marginal density of the single running variable to be continuous at its respective threshold. This leads to a set of conditions on the continuity of the conditional marginal densities of all running variables, and I propose a multidimensional manipulation test based on a quadratic form of the test statistics introduced by Cattaneo et al. (2020) for the single-dimensional case, computed for each condition. Asymptotically, these statistics converge to a multivariate normal distribution, and the proposed test statistic converges to a chi-square distribution with degrees of freedom equal to the number of running variables.

Finally, I compare my procedure with alternative methods, providing a review of other approaches commonly used in the literature.

I am not the first to study MRDD from a theoretical perspective. Identification and estimation in the MRDD setting, and how they differ from the single-dimensional RDD, have been investigated by Imbens and Zajonc (2009) and Papay et al. (2011), respectively; other results are also discussed in Wong et al. (2013) and Imbens and Wager (2019). So far, to my knowledge, there is no research explicitly dealing with extending the framework proposed by Lee (2008) and discussing manipulation tests in the MRDD context. Interestingly, though, manipulation tests are run by several applied papers employing MRDD (see the survey in Table 1): They appeal to disparate approaches, with different null hypotheses, assumptions, and test statistics. None of these approaches justifies the implemented procedures,[3] while my test is supported by a model and backed by statistical theory. Local asymptotic analysis and Monte Carlo simulations confirm my test's advantages in terms of size control and power in realistic settings.

Three main strands of literature resort to MRDD as a tool for causal inference. First, it is exploited to evaluate policies that assign a treatment when more than one condition on observable continuous quantities is met. Examples can be found in several fields, mainly in education (Matsudaira 2008; Clark and Martorell 2014; Cohodes and Goodman 2014; Elacqua et al. 2016; Evans 2017; Smith et al. 2017; Londoño-Vélez et al. 2020) but also in corporate finance (Becht et al. 2016), political economy (Hinnerich and Pettersson-Lidbom 2014; Frey 2019), development (Salti et al. 2022), industrial organization (Snider and Williams 2015), and public economics (Egger and Wamser 2015). In these cases, MRDD provides reliable results on treatment effects from a clean identification strategy.

A second application is the geographic or spatial RDD to study the effect of treatments only assigned to specific areas. Running variables are latitude and longitude, and the boundary at which the ATE is computed coincides with actual (or historical) national, regional, or municipal borders. Keele and Titiunik (2015) discuss how this setting relates to MRDD in detail. Note, however, that I am considering a model where the treatment is assigned when each running variable exceeds its cutoff: As such, my results do not directly apply to the spatial RDD, and if my test works in this setting, it is a case-specific issue.

Third, recently there has been an increasing interest in MRDD in a theory literature at the intersection of market design and machine learning (Abdulkadiroglu et al. 2022; Narita and Yata 2021). When algorithms determine a treatment assignment, they may consider multiple thresholds and running variables in a setting that mimics an MRDD. This literature is primarily theoretical, but it will likely encourage new empirical research, potentially relying on my proposed MT.

The rest of the paper is organized as follows. Section 2 introduces the theoretical model for MRDD and derives the testable implication. Section 3 provides a manipulation test for the implication. Section 4 compares the manipulation test with alternative approaches used in the literature. Section 5 reports Monte Carlo simulations. Section 6 applies the manipulation test to Frey (2019). Section 7 concludes.

TABLE 1 | Published papers using MRDD. Most studies utilize either separate tests (ST), where each running variable's density continuity is tested individually, or distance as running variable tests (DT), which consider the distance of observations from the boundary as the unique running variable.

Authors (year)	Manipulation test	ST	DT
Frey (2019)	×		
Matsudaira (2008)	✓	✓	
Hinnerich and Pettersson-Lidbom (2014)	✓	✓	
Elacqua et al. (2016)	✓	✓	
Egger and Wamser (2015)	✓	✓	
Evans (2017)	✓	✓	
Smith et al. (2017)	✓	✓	
Londoño-Vélez et al. (2020)	✓	✓	
Clark and Martorell (2014)	✓		✓
Cohodes and Goodman (2014)	✓		✓
Becht et al. (2016)	✓		✓

2 | Model and Testable Implication

In this section, I outline the model for the MRDD and present the testable implication for the manipulation test in the multidimensional setting. This model extends some of the results previously proposed by Lee (2008) and McCrary (2008) to the multidimensional case. The comprehensive discussion of the model, including identification results and analysis of various multidimensional RD designs, is provided in Appendix S1.

Let $Z \in \mathbb{R}^d$ be a random vector consisting of d observable continuous running variables, with a joint cumulative distribution function $F(z)$ and a joint probability density function $f(z)$.

The treatment status D depends on Z. This paper focuses on a sharp design, where the treatment status is deterministic. However, the framework can also be extended to a fuzzy design, where the probability of receiving treatment changes discontinuously at the threshold. Consider a multiple-threshold treatment rule, where units receive the treatment ($D = 1$) if all running variables exceed their respective cutoffs:

$$D = D(Z) = \mathbf{1}\{Z \geq c\} = \mathbf{1}\{Z_1 \geq c_1\}\mathbf{1}\{Z_2 \geq c_2\} \ \dots \ \mathbf{1}\{Z_d \geq c_d\} \tag{1}$$

Without loss of generality, assume a rescaling of Z such that $c_j = 0$ for all j. In this case, units are treated if, and only if, $Z_j \geq 0$ for all j.

Let $\mathcal{T}$ denote the set of values of Z for which $D(Z) = 1$, and let $\overline{\mathcal{T}^C}$ represent the closure of the complement of $\mathcal{T}$. Define the boundary $\mathcal{B}$ as the set of points with both treated and untreated units in any neighborhood: formally, $\mathcal{B} = \mathcal{T} \cap \overline{\mathcal{T}^C}$.

The MRDD leverages the treatment assignment mechanism described in Equation (1) to identify the causal effect of the treatment at the boundary, where treated and untreated units are comparable and differ only in their treatment status. In Appendix S1, I formally discuss the assumptions under which different causal parameters can be identified. These parameters include the conditional average treatment effect (CATE), the average treatment effect at a specific point on the boundary, and the integrated CATE (ICATE), which summarizes the average effect along the entire boundary.

I then demonstrate how one of the key assumptions for identification leads to a testable implication for the observable distribution of Z, analogous to the condition derived by McCrary (2008) in the single running variable case. This implication, formalized as Proposition 4 in Appendix S1.3, is presented in the following proposition.

Testable Implication 1 (Continuity of the conditional marginal densities). To identify the CATE, the MRDD relies on assumptions with the following implication:

$$f_{Z_j|Z_{-j}}(z_j|z_{-j} \geq 0) \text{ continuous at } z_j = 0, \forall j \tag{2}$$

where

$$f_{Z_j|Z_{-j}}(z_j|z_{-j} \geq 0) = \frac{\int_{z_{-j} \geq 0} f_{Z_j, Z_{-j}}(z_j, z_{-j}) d z_{-j}}{\int_{z_{-j} \geq 0} f_{Z_{-j}}(z_{-j}) d z_{-j}}$$

denotes the conditional probability density function of the running variable Z_j, conditional on all the other running variables $Z_{-j} = (Z_1, \dots, Z_{j-1}, Z_{j+1}, \dots, Z_d)$ being larger than 0.

The implication in Equation (2) requires the continuity at the boundary of the conditional marginal density of each running variable. This condition is equivalent to the continuity of the marginal density of each random variable j at its threshold ($z_j = 0$), given that all other running variables exceed their respective cutoffs ($z_{-j} \geq 0$). In terms of restrictions on agents' behavior, this implication rules out any influence over individual running variables that would change the treatment status at the boundary.

The implication is derived by combining the multiple-threshold assignment rule with the assumption that units on either side of the boundary are comparable. As a result, it does not depend on the specific parameter of interest (e.g., CATE or ICATE) or the estimation strategy chosen by the researcher. Consequently, the derivation remains valid even in scenarios that deviate from the baseline model, as long as the treatment is assigned using rules analogous to Equation (1). For example, the implication also holds in the case studied by Papay et al. (2011), where two running variables define four distinct treatment regions, each characterized by a combination of threshold rules.

In the next section, I propose a manipulation test for the implication in Equation (2), which stems from the assumption required for identification in the MRDD. This manipulation test should be regarded as a robustness check and, in any specific application, should complement rather than replace a thorough discussion of the validity of the MRDD assumptions.

3 | Manipulation Test

In this section, I first introduce the manipulation test for the implication in Equation (2), presenting the test statistic and the critical values. Practitioners interested in applying the manipulation test will find a practical description of how to implement the procedure. Then, in Section 3.1, I discuss the assumptions necessary to establish the asymptotic validity and consistency of the test. Finally, in Section 3.2, I derive the formal results.

The MRDD manipulation test I propose for the implication in Equation (2) is defined as follows:

$$\phi(\hat{t}, \alpha) = \begin{cases} 1, & \text{if } \hat{t} > c(\alpha), \\ 0, & \text{if } \hat{t} \leq c(\alpha), \end{cases} \tag{3}$$

where $\hat{t}$ is the test statistic, α the significance level, and $c(\alpha)$ the critical value. Whenever $\phi(\hat{t}, \alpha)$ equals 1, the null hypothesis is rejected.

The construction of the test statistic $\hat{t}$ involves two steps. First, for each running variable j, compute the statistic $\hat{\theta}_j$ along with its variance estimator $\hat{\sigma}_j^2$. The expression for $\hat{\theta}_j$ is given by

$$\hat{\theta}_j = \hat{f}^{+}_{Z_j|Z_{-j}}(0|z_{-j} \geq 0) - \hat{f}^{-}_{Z_j|Z_{-j}}(0|z_{-j} \geq 0),$$

where $\hat{f}^{+}_{Z_j|Z_{-j}}$ and $\hat{f}^{-}_{Z_j|Z_{-j}}$ are estimators of the conditional marginal density of Z_j, as defined in Section 3.1. It is worth noting that $\hat{\theta}_j$ resembles the test statistic proposed by Cattaneo et al. (2020) for testing the continuity of the density in a single-dimensional RDD. However, in this context, the test focuses on the continuity of a *conditional* marginal density, which requires adaptations for the statistic and the formal proofs.

In practice, $\hat{\theta}_j$ and $\hat{\sigma}_j$ can be computed as follows: for any $j = 1, \ldots, d$, consider the subsample of data that includes only observations where $Z_{-j} \geq 0$. Within this subsample, use available packages in R and Stata to run the manipulation test proposed by Cattaneo et al. (2020) and obtain the test statistic $\hat{\theta}_j$ along with its standard error estimator $\hat{\sigma}_j$.

Next, construct the test statistic $\hat{t}$ based on the vector $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_d)$ and the diagonal matrix $\hat{\Sigma} = \text{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_d^2)$, using the quadratic form:

$$\hat{t} = \hat{\theta}' \hat{\Sigma}^{-1} \hat{\theta} \tag{4}$$

In Section 3.2, I derive the asymptotic distribution of $\hat{\theta}$ and prove that the test statistic $\hat{t}$ converges to a χ^2 distribution with d degrees of freedom. Consequently, the critical value $c(\alpha)$ can be chosen as the $1-\alpha$ quantile of a χ^2 distribution with d degrees of freedom.

With these test statistics and critical value, the manipulation test $\phi(\hat{t}, \alpha)$ is asymptotically valid and consistent for the null hypothesis in Equation (2).

3.1 | Assumptions

To prove the properties of the manipulation test, the following assumption is required.

Assumption 1. (Smoothness). $\{z_i\}_{i \in \{1,\ldots,n\}}$ is an iid random sample of Z with cumulative distribution function F. In neighborhoods of points on the boundary $\mathcal{B}$, F is at least four times continuously differentiable.

The assumption that F has at least four continuous derivatives allows for consistent estimation of the conditional densities $f^{+}_{Z_j|Z_{-j}}(0)$ and $f^{-}_{Z_j|Z_{-j}}(0)$. This is analogous to the assumption required by the manipulation test of McCrary (2008) for the single-dimensional case.

To simplify the notation, define $f_j(z_j) = f_{Z_j|Z_{-j}}(z_j|z_{-j} \geq 0)$, and consider the local polynomial estimator $\hat{f}_{j,p}(z_j)$:

$$\hat{f}_{j,p}(z_j) = e_1' \hat{\beta}(z_j),$$

$$\hat{\beta}(z_j) = \text{argmin}_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \left[\tilde{F}_j(z_j|z_{-j} \geq 0) - r_p(z_{ji} - z_j)' b\right]^2 K\left(\frac{z_{ji} - z_j}{h_j}\right) \mathbf{1}\{z_{-j} \geq 0\},$$

where $e_1' \in \mathbb{R}^{p+1}$ such that $e_1' = (0, 1, 0, \ldots, 0)$; $n_j = \sum_{i=1}^{n} \mathbf{1}\{z_{-j} \geq 0\}$ is the number of observations actually considered for the test; $\tilde{F}_j(z_j|z_{-j} \geq 0) = \frac{1}{n_j} \sum_i \mathbf{1}\{z_{ji} \leq z_j\} \mathbf{1}\{z_{-ji} \geq 0\}$ is the empirical distribution function for the marginal conditional distribution $F(z_j|z_{-j} \geq 0)$; $r_p(u) = (1, u, u^2, \ldots, u^p)$ is a one-dimensional polynomial expansion; h_j is a bandwidth, which will be better specify later; and $K(\cdot)$ is a kernel function satisfying the following assumption.

Assumption 2. (Kernel). The kernel function $K(.)$ is nonnegative, symmetric, continuous, and integrates to one: $\int K(u)du = 1$. It has support $[-1, 1]$.

The local polynomial approach for estimating derivatives of the cumulative distribution function is extensively discussed in Cattaneo et al. (2020), although it was mentioned already by Jones (1993). Chapter 3 of Fan and Gijbels (1996) examines several properties of the local polynomial estimator that make it particularly well suited for the manipulation test. Notably, it achieves the optimal rate of convergence both at interior points and boundaries, while being boundary-adaptive. No adjustments are needed when estimating points near the boundary if the object of interest is the νth derivative and $p - \nu$ is odd. For my MT, where $\nu = 1$, I employ an estimator with $p = 2$ to take full advantage of its boundary adaptiveness.

3.2 | Manipulation Test

To establish asymptotic validity and consistency of the test $\phi(\hat{\iota}, \alpha)$, it is necessary to derive some intermediate results. The first result regards the asymptotic properties of the density estimator $\hat{f}_{j,p}(z_j)$. Formulas for bias $B(x)$, variance $V(x)$, and consistent variance estimator $\hat{V}(x)$ are reported in Appendix S2.

Proposition 1. **(Asymptotic distribution of $\hat{f}_{j,p}(z_j)$).** *Under Assumptions 1 and 2, with $p = 2$, $nh_j^2 \to \infty$ and $nh_j^{2p+1} = O(1)$, $\hat{f}_{j,p}(z_j)$ is a consistent estimator for $f_j(z_j)$. Furthermore,*

$$\sqrt{n_j h_j}(\hat{f}_{j,p}(z_j) - f_j(z_j) - h_j^p B(z_j)) \to^d \mathcal{N}(0, V(z_j))$$

where $B(z_j)$ is the asymptotic bias and $V(z_j)$ the asymptotic variance.

When $nh_j^{2p+1} = O(1)$, the bandwidth h_j has the MSE-optimal rate and can be computed by cross-validation.

The presence of asymptotic bias $B(z_j)$ is standard in nonparametric settings and must be considered to ensure valid hypothesis testing. In this paper, I adopt the robust bias correction method proposed by Calonico et al. (2018). Alternative approaches include the critical values correction method suggested by Armstrong and Kolesár (2020).

Bias-corrected inference for the density estimator $\hat{f}_{j,p}(z_j)$ can be obtained by considering the estimator $\hat{f}_{j,q}(z_j)$ with $q = p + 1$, computed with the bandwidth $h_{j,p}$, the MSE-optimal bandwidth for $\hat{f}_{j,p}(z_j)$ (see Calonico et al. (2022) and Cattaneo et al. (2022) for an extensive discussion on the procedure). Moving forward, I will consider the estimator $\hat{f}_{j,p}(z_j)$ for point estimates and the estimator $\hat{f}_{j,q}(z_j)$ with bandwidth $h_{j,p}$ to construct bias-corrected confidence intervals for $\hat{f}_{j,p}(z_j)$.

Let $n_{j+} = \sum_{i=1}^n \mathbf{1}\{z_j \geq 0\}\mathbf{1}\{z_{-j} \geq 0\}$ and $n_{j-} = \sum_{i=1}^n \mathbf{1}\{z_j < 0\}\mathbf{1}\{z_{-j} \geq 0\}$, and denote by $\frac{n_{j+}}{n_j}\hat{f}_{j+,p}(z_j)$ and $\frac{n_{j-}}{n_j}\hat{f}_{j-,p}(z_j)$ the estimators of conditional density $f_j(z_j) = f_{Z_j|Z_{-j}}(z_j|z_{-j} \geq 0)$ computed considering only observations in $\{z : z \geq 0\}$ and $\{z : z_j < 0, z_{-j} \geq 0\}$, respectively.

Consider $\theta_j = \lim_{z_j \to 0^+} f_j(z_j) - \lim_{z_j \to 0^-} f_j(z_j)$, and note that, when the implication in Equation (2) is true, $\theta_j = 0$ for all j. Define the statistic $\hat{\theta}_{j,p}$:

$$\hat{\theta}_{j,p} = \frac{n_{j+}}{n_j}\hat{f}_{j+,p}(0) - \frac{n_{j-}}{n_j}\hat{f}_{j-,p}(0) \tag{5}$$

The following result derives the asymptotic distribution of the statistic.

Proposition 2. **(Asymptotic distribution of $\hat{\theta}_{j,q}$).** *Under Assumptions 1 and 2 holding separately for $\{Z : Z \geq 0\}$ and $\{Z : Z_{-j} \geq 0, Z_j < 0\}$, with $p = 2$, $q = p + 1$, $n\min\{h_{j-}, h_{j+}\} \to \infty$, and $n\max\{h_{j-}^{1+2q}, h_{j+}^{1+2q}\} \to 0$, when the implication $\theta_j = 0$ is true:*

$$\frac{1}{\sigma_j}\hat{\theta}_{j,q} \to^d \mathcal{N}(0, 1)$$

where

$$\sigma_j^2 = \frac{\pi_{j+}}{h_{j+}\pi_j n}V_{j+}(0) + \frac{\pi_{j-}}{h_{j-}\pi_j n}V_{j-}(0).$$

A consistent estimator $\hat{\sigma}_j^2$ for σ_j^2 can be obtained by

$$\hat{\sigma}_j^2 = \frac{n_{j+}}{h_{j+}n_j^2}\hat{V}_{j+,q}(0) + \frac{n_{j-}}{h_{j-}n_j^2}\hat{V}_{j-,q}(0).$$

Proposition 2 is valid for any j. I am interested in the asymptotic distribution of the vector $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_d)$, whose distribution under the null hypothesis of continuity of $f_{Z_j|Z_{-j}}(z_j|z_{-j} \geq 0)$ is derived in the next theorem.

Theorem 1. **(Asymptotic distribution of $\hat{\theta}$).** *Under Assumptions 1 and 2 holding separately for $\{Z : Z \geq 0\}$ and $\{Z : Z_{-j} \geq 0, Z_j < 0\}$ for all j, with $p = 2$, $q = p + 1$, $n\min\{h_{j-}, h_{j+}\} \to \infty$ and $n\max\{h_{j-}^{1+2q}, h_{j+}^{1+2q}\} \to 0$ for all j, when $\theta = 0$,*

$$\hat{\Sigma}^{-\frac{1}{2}}\hat{\theta} \to^d \mathcal{N}(0, I),$$

where $\hat{\Sigma}_{jj} = \hat{\sigma}_j^2$ as defined in Proposition 2, and $\hat{\Sigma}_{ji} = 0$ for all $i \neq j$.

Theorem 1 shows that, even if the number of observations simultaneously considered by any pair of estimators θ_j and θ_i goes to infinite, they are asymptotically independent.

The asymptotic distribution of the quadratic form test statistic $\hat{\iota}$ defined in Equation (4) is immediately derived from Theorem 1. The quadratic form is, in fact, a continuous function of $\hat{\Sigma}^{-\frac{1}{2}}\hat{\theta}$ and is hence distributed as the sum of d squared independent normals: a χ^2 distribution with d degrees of freedom.

Theorem 1 allows to derive the asymptotic distribution of any continuous function of the vector $\hat{\Sigma}^{-\frac{1}{2}}\hat{\theta}$. To test the null hypothesis in Equation (2), it is useful to consider the ℓ^p-norm statistic $\|\hat{\Sigma}^{-\frac{1}{2}}\hat{\theta}\|_p$: Large values of this statistic give evidence against the null hypothesis, suggesting to choose the critical value $c(\alpha)$ as the $1 - \alpha$ quantile of its asymptotic distribution. The continuity of the ℓ^p-norm ensures that $\|\hat{\Sigma}^{-\frac{1}{2}}\hat{\theta}\|_p \to^d \|X\|_p$, where $X \in \mathbb{R}^d$ is a random vector with distribution $\mathcal{N}(0, I)$. In general, the quantiles of $\|X\|_p$ and hence the critical value for the test can be obtained through simulations, drawing random samples from X and computing $\|X\|_p$. The quadratic form can be seen as a special case, where the Euclidean distance (ℓ^2-norm) is squared to obtain a χ^2 distribution whose quantiles can be analytically computed.

With these test statistics and critical values, the MRDD manipulation test $\phi(\hat{\iota}, \alpha)$ defined in Equation (3) is asymptotically valid and consistent: when the null hypothesis in Equation (2) is true, the test has an asymptotic rejection probability of α. When the null hypothesis is false, the test asymptotically rejects with probability one. The following corollary formalizes this result.

Corollary 1. (**Manipulation Test**). *Let H_0 be the null hypothesis in Equation (2), and consider the MT $\phi(\hat{t}, \alpha)$ defined in Equation (3) with the test statistic $\hat{t}$ defined in Equation (4) and as critical value $c(\alpha)$ the $1-\alpha$ quantile of the χ^2 distribution. Under the assumptions of Theorem 1, when H_0 is true:*

$$\lim_{n\to\infty} P(\phi(\hat{t}, \alpha) = 1) = \alpha.$$

When H_0 is false:

$$\lim_{n\to\infty} P(\phi(\hat{t}, \alpha) = 1) = 1.$$

In the next section, the MT is compared to some alternatives used in the literature, studying its finite sample properties in terms of power and size control.

Remark 1. (Choice of the test statistic). The manipulation test is asymptotically valid and consistent for any ℓ^p-norm test statistic, with $p > 0$. Different ℓ^p-norms imply differences in power against different alternatives. The quadratic form test statistic with $p = 2$ performs well in detecting discontinuities diffused across all running variables. These behaviors are of primary interest: When the treatment's benefits lead agents to manipulate their running variables for eligibility, manipulation is likely widespread. In contrast, if only one running variable is manipulated, other test statistics may be more appropriate. For this case, I discuss the ℓ^∞-norm statistic, equivalent to the max-statistic $\hat{t}_m = \max\left(\left|\frac{\hat{\theta}_1}{\hat{\sigma}_1}\right|, \ldots, \left|\frac{\hat{\theta}_d}{\hat{\sigma}_d}\right|\right)$, in Appendix S3.2.

4 | Alternative Approaches

It is common for applied papers utilizing the RDD to include manipulation tests for their running variables, as highlighted in the survey by Canay and Kamat (2018). Although a theoretical foundation for manipulation tests in the multidimensional case is lacking, Table 1 demonstrates the prevalence of such tests in papers employing the MRDD. These papers typically use one of two approaches: conducting multiple tests, one for each running variable separately (separate tests [ST]), or aggregating the running variables by considering the distance of each observation from the boundary and then running the MT using the distance as a single running variable (distance as running variable test [DT]).

The ST approach does not control the size for the null hypothesis in Equation (2), while the DT approach is not consistent against certain alternatives and is not robust to changes in the units of measurement. In the following sections, I compare these approaches with the proposed manipulation test [MT], highlighting their limitations and demonstrating how they can be adapted to properly test the null hypothesis in Equation (2).

It is important to emphasize that the primary goal of this comparison is not to evaluate different test statistics but to demonstrate how each test statistic can be adapted to the multidimensional setting and to analyze the advantages and drawbacks of each adaptation. In other words, the goal is not to determine whether the test statistic based on the local polynomial estimator of the derivatives of the empirical distribution function, proposed by Cattaneo et al. (2020) and discussed in the previous section, is superior to those based on histogram smoothing (McCrary 2008) or order statistics (Bugni and Canay 2021). Such a comparison would not specifically address the multidimensional setting and is therefore outside the scope of this extension.

Instead, the goal is to compare different procedures (the one I proposed, multiple hypothesis tests, and tests based on the distance from the boundary) for the MT in the MRDD that could be implemented using various test statistics. To ensure that the observed differences are attributable to the procedures rather than the test statistic itself, all tests in this analysis are specified using the test statistic proposed by Cattaneo et al. (2020). However, they could also be implemented with alternative test statistics, such as those proposed by McCrary (2008) or Bugni and Canay (2021).

4.1 | Separate Tests (ST)

The ST procedure in the context of MRDD treats each running variable separately and applies existing MTs designed for single-dimensional RDD (McCrary 2008; Cattaneo et al. 2020; Bugni and Canay 2021). In some cases (Egger and Wamser 2015; Evans 2017; Londoño-Vélez et al. 2020), the test is conducted on the conditional marginal densities by considering only the units that meet the threshold for the other running variables. Either way, without accounting for multiple hypotheses testing, the ST is invalid, as it does not control the size for the null in Equation (2).

4.2 | Multiple Hypotheses Test With Bonferroni Correction (BCT)

A straightforward fix for the ST is to account for multiple hypotheses testing using Bonferroni correction (BCT). In case the test by Cattaneo et al. (2020) is employed, the resulting procedure partly overlaps with the test I proposed. To test the implication in Equation (2) at level α, statistics $\hat{\theta}_j$ defined in Equation (5) are used to conduct ST for each running variable, with the critical values adjusted for multiple testing (for a review on multiple hypotheses testing, see Chapter 9 in Lehmann and Romano (2022)). The null hypothesis of running variable j continuity is tested at significance level $\frac{\alpha}{d}$, where d is the number of running variables. The implication in Equation (2) is rejected if the continuity of any of the running variables is rejected. The correction for the number of hypotheses ensures correct coverage, meaning that the asymptotic family-wise error rate, which is the probability of rejecting one or more true null hypotheses (and hence the probability of rejecting the implication in Equation (2) when it is true), is not greater than α.[4] In this context, alternative multiple hypotheses corrections (e.g., stepwise methods or Holm correction) would coincide with the BCT, as rejecting continuity for the density of just one running variable is equivalent to rejecting the implication in Equation (2).

MT and BCT rely on the same vector of statistics $\hat{\theta}$. Local power analysis can be used to compare the power of the two tests against local alternative hypotheses, letting the discontinuity of the density at the threshold get smaller as the sample size increases. I consider a framework where all the running variables are discontinuous, and the discontinuity is equal to $k/\sqrt{nh}$, such that, asymptotically, $\hat{\theta}_j \to^d \mathcal{N}(k, 1)$ for all j. This framework mimics a setting where all the running variables are manipulated to get the treatment: If treatment is desirable (or undesirable), I expect all the agents close to the treatment region to manipulate their running variables to get in (or out) the region.

Figure 2 reports power curves for MT (in red) and BCT (in blue), considering different numbers of running variables $d \in \{2, 5, 8\}$. When all running variables exhibit discontinuity, MT outperforms BCT in terms of power. This is because MT combines information from all the running variables and effectively detects manipulation when it is widespread across them. On the other hand, BCT considers each running variable separately, which results in lower power in case of widespread manipulation.

The local power analysis confirms that aggregating information and testing a single hypothesis is better than testing multiple hypotheses for the continuity of each running variable separately. The MT is less conservative, at least against alternative hypotheses where manipulation is spread across all running variables.

4.3 | Distance as Running Variable Test (DT)

The second approach for manipulation tests in the multidimensional setting employed in applied research involves dimension reduction: the MRDD is reduced to a single-dimensional design, with the scalar distance between the vector of running variables and the boundary $\mathcal{B}$ as the only running variable.[5] The distance is used to estimate the CATE, similar to the classical RDD, and to conduct a manipulation test using one of the available tests (McCrary 2008; Cattaneo et al. 2020; Bugni and Canay 2021). This approach appears simple because it directly relates to the single-dimensional RDD case. Nonetheless, it comes with some caveats that need to be considered.

First, the choice of distance metric and measurement units can significantly impact the test results. Different distance metrics used to measure the distance between the running variables and the boundary lead to different test statistics and test outcomes, as well as different units of measurement for the running variables. To address this second issue, one possible solution is to standardize the running variables to have unit variance before conducting the manipulation test, but this practice may worsen the properties of the test (as shown in Monte Carlo simulations).

A second flaw of the DT is that it is inconsistent against certain fixed alternatives of the null hypothesis in Equation (2). For instance, if there are opposite discontinuities in the marginal distributions of different running variables, and these discontinuities balance each other out, the asymptotic probability that the DT will reject the false null hypothesis in Equation (2) is equal to α, rather than one. A design where the DT test is inconsistent is studied in Section 5.2.

Despite the lack of a clear theoretical background and the lack of power against specific alternative hypotheses, it may still be the case that the DT performs better than MT and BCT in contexts plausible in applications. However, if any, the evidence suggests poorer finite sample performance for DT, as shown by Wong et al. (2013) or in the Monte Carlo simulation in the following section.

5 | Monte Carlo Simulation

In this section, I conduct Monte Carlo simulations to evaluate the finite sample performance of the manipulation test (MT) proposed in this paper. The MT is compared to the multiple hypotheses test with the BCT[6] and the single running variable test, both with standardization to unit variance (SDT) and without it (DT), in each case using the signed Euclidean distance from the boundary as the single running variable.

Without loss of generality, the cutoff is set at 0 for all running variables: Units are treated when all the running variables are nonnegative. The boundary is the set of points with all nonnegative coordinates and at least one coordinate equal to zero. Figure 3

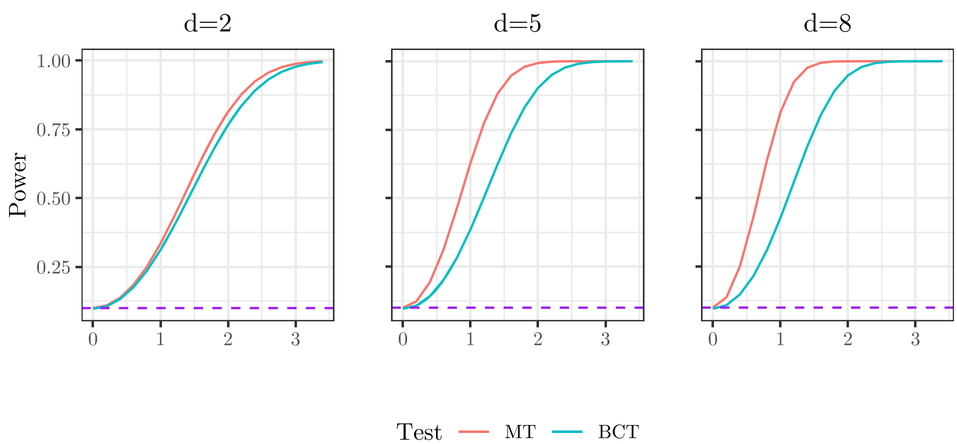

FIGURE 2 | Local asymptotic power curves for the manipulation test proposed in this paper (MT, in red) and the multiple hypotheses test with Bonferroni correction (BCT, in blue). Plots consider different numbers d of running variables. The significance level $\alpha = 0.1$ is represented by the dotted horizontal purple line. All the running variables are discontinuous, and the discontinuity is equal to $k/\sqrt{nh}$, such that $\hat{\theta}_j \to^d \mathcal{N}(k, 1)$ for all j.

reports a realization of the simulated samples for the four models, illustrating the joint distribution of the running variables.

Models 1 and 2 show how the tests are comparable in controlling the size, while Models 3 and 4 attest the better power properties for MT discussed in Section 4.

5.1 | Models 1 and 2

Model 1. Consider d running variables uniformly distributed:

$$Z_j \sim U(-1,1) \quad \text{for } j \text{ in } \{1, \dots, d\}.$$

In Model 1, densities are symmetrical to the threshold, and the density function is flat. Because the setting can be particularly convenient for the tests, Model 2 considers densities with different behaviors at the two sides of the boundary.

Model 2. Consider d running variables normally distributed and centered at 1:

$$Z_j \sim \mathcal{N}(1,1) \quad \text{for } j \text{ in } \{1, \dots, d\}.$$

Both Models 1 and 2 are simulated considering sample sizes $n \in \{500, 2000, 5000\}$ and total numbers of running variables $d \in \{2, 3, 4\}$. The simulation results are presented in Table 2. Overall, for all the sample sizes considered, the tests tend to underreject, with empirical rejection rates closer to the theoretical ones for DT and SDT. MT and BCT exhibit similar performances across different models and parameters specification: rejection rates get closer to asymptotic ones as the effective sample size grows, when d decreases for a fixed n or n increases for a fixed d. For the same values of parameters d and n, underrejection is larger in Model 1 than Model 2: As expected, the steeper is the probability density function at the cutoff, the higher is the probability for the test to reject the true null.

Unsurprisingly, all the tests have a comparable performance: No theoretical reason suggests discrepancies for the three tests in controlling size. Differences arise when finite sample power is studied, as shown by Models 3 and 4.

5.2 | Model 3

Model 3. Define random vector $Z^* = (Z_1^*, Z_2^*)$, where $Z_1^* \sim U(-1,1)$, $Z_2^* \sim U(-1,1)$, and Z_1^* and Z_2^* independent. Define sets $A_1 = \{(z_1, z_2) : z_1 < 0, -z_1 < z_2\}$ and $A_2 = \{(z_1, z_2) : z_1 > z_2, z_2 > 0\}$.

Consider two running variables Z_1 and Z_2 distributed as follows:

$$Z_1 \sim \begin{cases} Z_1^*, & \text{if } Z^* \notin A_1 \\ Z_1^*, & \text{with probability } 1-\gamma_1 \text{ if } Z^* \in A_1 \\ -Z_1^*, & \text{with probability } \gamma_1 \text{ if } Z^* \in A_1 \end{cases}$$

$$Z_2 \sim \begin{cases} Z_2^*, & \text{if } Z^* \notin A_2 \\ Z_2^*, & \text{with probability } 1-\gamma_1 \text{ if } Z^* \in A_2 \\ -Z_2^*, & \text{with probability } \gamma_1 \text{ if } Z^* \in A_2 \end{cases}$$

Model 3 mimics a setting where the two running variables are manipulated, but in opposite directions: When $Z^* \in A_1$, Z_1 is manipulated to get the treatment; when $Z^* \in A_2$, Z_2 is manipulated to avoid the treatment. Parameter γ_1 governs the extent of manipulation: When $\gamma_1 = 0$, the joint density of Z_1 and Z_2 is continuous; when $\gamma_1 = 1$, the joint density becomes zero in regions A_1 and A_2, resulting in the maximum discontinuity.

The curves depicted in Figure 4 illustrate the finite sample performance of the tests. For both MT and BCT, the power of the tests increases with the degree of manipulation γ_1, as expected. For DT and SDT, the power always remains equal to the test size. This design corresponds to the situation described in Section 4.3: The condition in Equation (2) is not satisfied, because neither the marginal densities of Z_1 nor Z_2 are continuous at the threshold (as shown in Figure 3). Nonetheless, the probability density function of the distance from the boundary is continuous. Consequently, the null hypothesis tested by DT and SDT is true, resulting in trivial power for these tests.

5.3 | Model 4

Model 4. Consider two running variables distributed as follows:

$$Z_1 \sim \mathcal{N}(0,1)$$

$$Z_2 \sim \begin{cases} U(0,1), & \text{with probability } \gamma_2, \\ U(-1,0), & \text{with probability } 1-\gamma_2. \end{cases}$$

Model 4 is a design where only Z_2 is manipulated. The manipulation is determined by the parameter γ_2. When $\gamma_2 = 0.5$, Z_2 has a continuous density, following a uniform distribution between -1 and 1. When $\gamma_2 = 1$, the density of Z_2 becomes zero at the left of the boundary. The degree of discontinuity increases as the value of Z_2 deviates further from 0.5.

The curves in Figure 4 show how the finite sample power depends on γ_2. The power increases with higher values of γ_2 for all tests, but it is lower for DT and SDT. Additionally, despite two similar versions of the same test, DT and SDT's performances are different. As discussed in Section 4.3, the choice of the unit of measure affects the result of DT. In this case, the standardization applied in SDT reduces its power compared to DT. Standardization is not the solution to the issue.

MT and BCT exhibit similar behavior in the context of Model 4. It mimics a framework different from the one studied in the local asymptotic analysis, and there is no theoretical reason to expect the MT to perform better in this setting.

Overall, the Monte Carlo simulations confirm that the MT proposed in this paper has better finite sample properties than alternative tests. The simulations demonstrate advantages in terms of

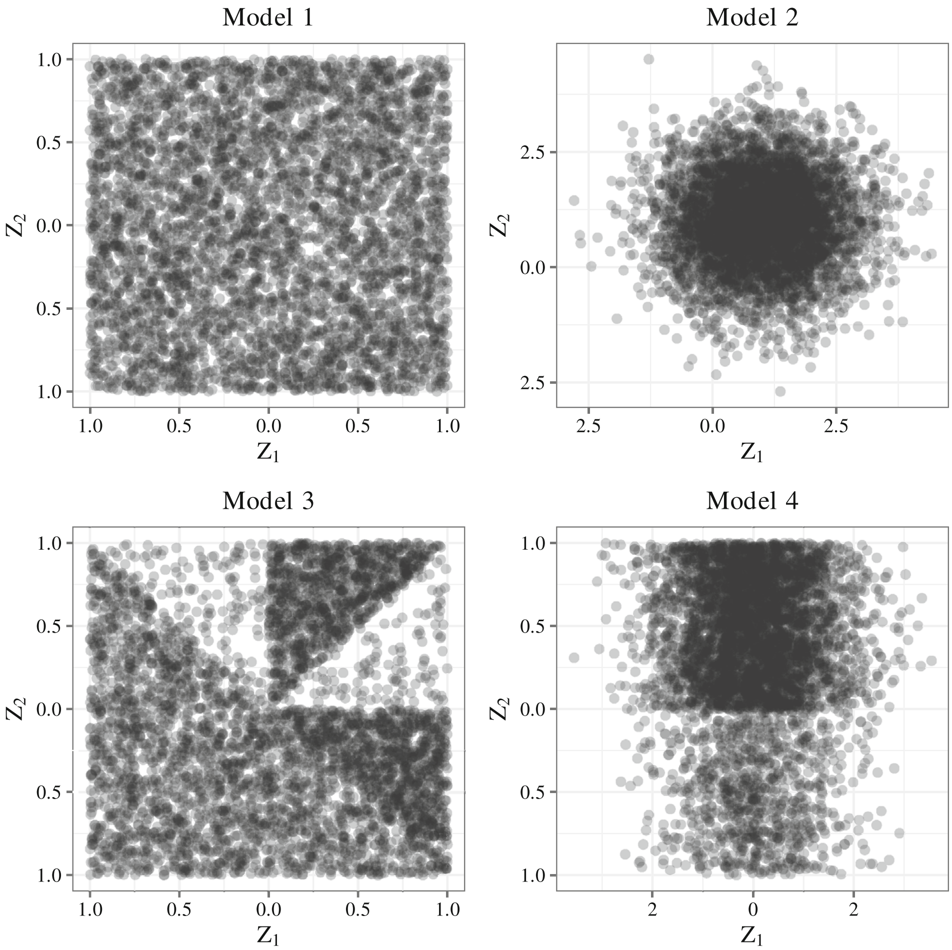

FIGURE 3 | Scatter plots of a sample of size $n = 5000$ from the four models illustrate the joint distribution of Z_1 and Z_2. For Models 1 and 2, parameter d is set to 2 (two running variables). Joint density is continuous, and the condition in Equation (2) is satisfied. For Models 3 and 4, parameters γ_1 and γ_2 are 0.8: Joint density is not continuous, and condition in Equation (2) is not satisfied.

power and robustness, reinforcing the findings derived from the local asymptotic analysis discussed earlier.

As outlined in Section 3, the proposed MT can be readily implemented using existing packages in popular statistical software, such as R and Stata, with just a few lines of code. This ease of implementation enhances the practical applicability of the test. The next section implements the MT in a real-world application, illustrating its simplicity.

6 | Application: Frey (2019)

I apply my manipulation test to the MRDD considered by Frey (2019) investigating the political economy of redistributive policies. In the original analysis, no manipulation test is reported. The paper studies the impact of cash transfers implemented by the Brazilian federal government on the dynamics of clientelism at the municipal level. The main hypothesis suggests that these cash transfers, by reducing the vulnerability of the poor, diminish the attractiveness of clientelism as a strategy for incumbent mayors.

The Bolsa Família (BF) program is the largest conditional cash transfer program globally and has been implemented in households across Brazil since 2003. The coverage of BF across different municipalities exhibits a positive correlation with the funding allocated to the Family Health Program (FHP), a household-based healthcare program run by municipalities since 1995. The positive correlation between BF coverage and FHP funding can be attributed to the fact that FHP teams have a significant penetration among poor households, potentially beneficiaries of BF. This enables them to effectively disseminate information about the BF program and encourage enrollment among eligible households.

To estimate the causal effect of the cash transfers on local clientelism, Frey (2019) exploits the link between BF and FHP, along with a specific discontinuity in the design of the FHP. The FHP provides municipalities with an additional 50% funding if they meet two criteria: a population of fewer than 30,000 inhabitants and a human development index (HDI) below 0.70. This discontinuity, determined by the joint thresholds of population and HDI, is directly reflected in the diffusion of the BF program: Consequences of cash transfers can be analyzed using an MRDD.

Figure 5 provides a visualization of the MRDD. In the space of the two running variables (population and HDI), treated municipalities are depicted in light blue, and untreated municipalities

TABLE 2 | Rejection rates under the true null hypothesis of continuity of marginal densities of the running variables, computed through 5000 Monte Carlo simulations, at 5% significance level. MT is the manipulation test proposed in this paper, BCT is the multiple hypotheses test with Bonferroni correction, and DT and SDT consider as single running variables the Euclidean distance from the boundary: For SDT, running variables are standardized to have unitary variance before computing the distance, while for DT they are not. d is the number of running variables, and n is the sample size.

		Model 1				**Model 2**			
d	***n***	**MT**	**BCT**	**DT**	**SDT**	**MT**	**BCT**	**DT**	**SDT**
	500	0.025	0.030	0.043	0.044	0.037	0.036	0.043	0.043
2	2000	0.036	0.039	0.047	0.046	0.042	0.041	0.050	0.049
	5000	0.032	0.031	0.045	0.045	0.040	0.039	0.052	0.052
	500	0.024	0.025	0.036	0.037	0.029	0.033	0.041	0.040
3	2000	0.027	0.029	0.041	0.040	0.038	0.039	0.039	0.039
	5000	0.033	0.033	0.037	0.037	0.037	0.040	0.045	0.044
	500	0.025	0.019	0.047	0.047	0.037	0.032	0.043	0.039
4	2000	0.020	0.019	0.041	0.040	0.036	0.036	0.042	0.042
	5000	0.030	0.035	0.044	0.044	0.039	0.041	0.041	0.041

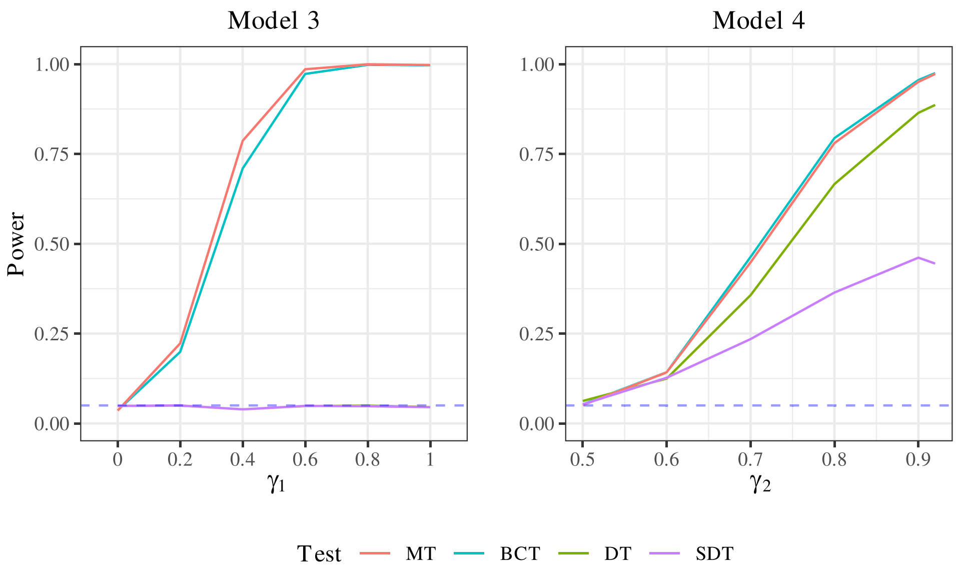

FIGURE 4 | Power of different manipulation tests with $n = 2000$, computed though 5000 Monte Carlo simulations. The dotted line indicates the nominal size of the tests (5%). MT is the manipulation test proposed in this paper, BCT is the multiple hypotheses test with Bonferroni correction, and DT and SDT consider as single running variables the Euclidean distance from the boundary: For SDT, running variables are standardized to have unitary variance before computing the distance; for DT, they are not. Parameters γ_1 and γ_2 determine the degree of manipulation. In Model 3, lines for DT and SDT overlap.

in dark blue. The red line represents the boundary. In this specific context, the treatment corresponds to the additional FHP funding, which leads to variations in the adoption rate of the BF program.

In this context, the MRDD requires the assumption that the joint density of population and HDI is continuous at the boundary, thereby ensuring the continuity of their respective marginal densities. The manipulation test is employed to validate the design and enhance the credibility of the study's findings. The test is conducted considering two running variables, resulting in a p value of 0.490, as reported in Table 3. With a significance level of $\alpha = 0.05$, the null hypothesis is not rejected, indicating no evidence of manipulation. In this context, the same conclusion of absence of manipulation is also reached by the other tests discussed in Section 4 (BCT, DT, and SDT). This is not surprising, as all the tests are expected to control the size similarly and mostly differ in power. It is nonetheless interesting to observe how, even in real applications, the DT and SDT yield different p values, despite using the same procedure just with a rescale of the running variables.

It is important to emphasize that the MT alone does not establish the model's validity. It serves as a robustness check and provides supporting evidence for the continuity of the densities of the running variables, but cannot substitute a discussion on why the assumptions of the MRDD are likely to hold in this setting,

TABLE 3 | p values are reported for various manipulation tests for the MRDD considered by Frey (2019). For the testing procedure with Bonferroni corrections (BCT), two p values are presented (for population and HDI, respectively), as the procedure involves testing the continuity of the conditional marginal densities separately, each at a $\frac{\alpha}{2}$ level. For DT and SDT, the tests are performed using the signed Euclidean distance from the boundary as the single running variable.

Test	MT	BCT	DT	SDT
p value	0.49	0.35, 0.45	0.19	0.38

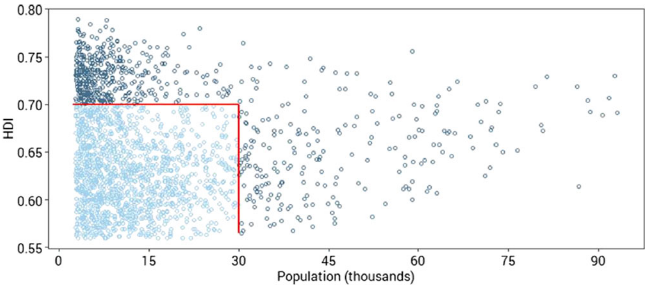

FIGURE 5 | Running variables for the MRDD considered by Frey (2019). Municipalities are assigned to the policy (light blue dots) when the population is below 30,000 inhabitants (x axis), and human development index is below 0.7 (y axis).

a discussion that remains essential for drawing valid conclusions from the analysis.

7 | Conclusion

This paper introduced a manipulation test for the MRDD. I extended the model proposed by Lee (2008) for the single-dimensional RDD to the multidimensional setting, and demonstrated, similar to McCrary (2008), how an assumption on unobservable quantities in the model leads to a testable implication on observable quantities—the continuity of the conditional marginal densities of the multiple running variables. I proposed a manipulation test for this implication and compared it with alternative approaches commonly used in applied research. While these approaches vary, they generally lack clear theoretical justification, and some are inconsistent for the considered implication. Through Monte Carlo simulations and local power analysis, I explored the finite sample properties of the proposed tests.

The manipulation test should be seen as a robustness check to strengthen the credibility of the assumptions required by the MRDD, and it is not intended as a pretest. It can be readily implemented with already existing packages without the need for additional tuning parameters. Considering the application in Frey (2019), I showed how to use the test in practice.

I focused on the case where treatment is assigned when each running variable exceeds its threshold, using this rule to derive a testable implication for the conditional marginal densities. Notably, my approach does not directly extend to the geographic RDD, where the boundary can follow more complex shapes. Developing a test for the continuity of the joint density distribution in this more general setting remains an interesting open problem.

Acknowledgments

I am grateful to Federico Bugni and Ivan Canay for their guidance in this project. I am also thankful to the co-editor, the referees, and all the participants of the Econometrics Reading Group at Northwestern University for their comments and suggestions.

Data Availability Statement

The data that support the findings of this study are openly available in "Manipulation Test for Multidimensional RDD (replication data)" at https://doi.org/10.15456/jae.2025116.2114256586.

ENDNOTES

[1] See Abadie and Cattaneo (2018), Cattaneo et al. (2019), and Cattaneo et al. (2024) for recent comprehensive reviews on RDD applications, identification, estimation, and inference.

[2] This shape of the assignment region is the most popular in practice (see references below), but may exclude the spatial RDD.

[3] Snider and Williams (2015) recognize that formal results are missing, asserting that "Extending formal tests to check for the strategic manipulation […] with a two-dimensional predictor vector is not immediately clear."

[4] In Appendix S3.1, I show that the asymptotic family-wise error rate is strictly less than α, and hence, the BCT is conservative. However, I also show that the difference compared to the MT analyzed in this section is primarily attributable to the testing procedure itself and persists even when the test is adjusted to achieve exact size control.

[5] The distance of each point z from the boundary is defined as $\min_{b \in B} d(z, b)$, where $d(z, b)$ denotes a metric in $\mathbb{R}^d$ (e.g., Euclidean, Manhattan, and cosine). In Sections 5 and 6, I implement the test using the Euclidean distance, assigning each point a scalar running variable defined as $\min_{b \in B} \sqrt{\sum_{i=1}^{d} (z_i - b_i)^2}$, with the sign determined by the treatment status.

[6] In Section 4.2 and Appendix S3.1, I show that the BCT is conservative: a test with nominal size α asymptotically rejects the null with a probability strictly lower than α. However, this difference is minor—for instance, a test with $\alpha = 0.05$ rejects the true null with probability 0.049375—and hence, all comments in this section remain valid even if the BCT were adjusted to achieve the exact rejection probability of α.

References

Abadie, A., and M. D. Cattaneo. 2018. "Econometric Methods for Program Evaluation." *Annual Review of Economics* 10, no. 1: 465–503.

Abdulkadiroglu, A., J. D. Angrist, Y. Narita, and P. Pathak. 2022. "Breaking Ties: Regression Discontinuity Design Meets Market Design." *Econometrica* 90, no. 1: 117–151.

Armstrong, T. B., and M. Kolesár. 2020. "Simple and Honest Confidence Intervals in Nonparametric Regression." *Quantitative Economics* 11, no. 1: 1–39.

Becht, M., A. Polo, and S. Rossi. 2016. "Does Mandatory Shareholder Voting Prevent Bad Acquisitions?" *Review of Financial Studies* 29, no. 11: 3035–3067.

Bugni, F. A., and I. A. Canay. 2021. "Testing Continuity of a Density via *g*-Order Statistics in the Regression Discontinuity Design." *Journal of Econometrics* 221, no. 1: 138–159.

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2018. "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference." *Journal of the American Statistical Association* 113, no. 522: 767–779.

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2022. "Coverage Error Optimal Confidence Intervals for Local Polynomial Regression." *Bernoulli* 28, no. 4: 2998–3022.

Canay, I. A., and V. Kamat. 2018. "Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design." *Review of Economic Studies* 85, no. 3: 1577–1608.

Cattaneo, M. D., N. Idrobo, and R. Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press.

Cattaneo, M. D., N. Idrobo, and R. Titiunik. 2024. *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Cambridge University Press.

Cattaneo, M. D., M. Jansson, and X. Ma. 2020. "Simple Local Polynomial Density Estimators." *Journal of the American Statistical Association* 115, no. 531: 1449–1455.

Cattaneo, M. D., M. Jansson, and X. Ma. 2022. "lpdensity: Local Polynomial Density Estimation and Inference." *Journal of Statistical Software* 101, no. 2: 1–25.

Clark, D., and P. Martorell. 2014. "The Signaling Value of a High School Diploma." *Journal of Political Economy* 122, no. 2: 282–318.

Cohodes, S. R., and J. S. Goodman. 2014. "Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy." *American Economic Journal: Applied Economics* 6, no. 4: 251–285.

Egger, P. H., and G. Wamser. 2015. "The Impact of Controlled Foreign Company Legislation on Real Investments Abroad. A Multi-Dimensional Regression Discontinuity Design." *Journal of Public Economics* 129: 77–91.

Elacqua, G., M. Martinez, H. Santos, and D. Urbina. 2016. "Short-Run Effects of Accountability Pressures on Teacher Policies and Practices in the Voucher System in Santiago, Chile." *School Effectiveness and School Improvement* 27, no. 3: 385–405.

Evans, B. J. 2017. "Smart Money: Do Financial Incentives Encourage College Students to Study Science?" *Education Finance and Policy* 12, no. 3: 342–368.

Fan, J., and I. Gijbels. 1996. *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

Frey, A. 2019. "Cash Transfers, Clientelism, and Political Enfranchisement: Evidence From Brazil." *Journal of Public Economics* 176: 1–17.

Hinnerich, B. T., and P. Pettersson-Lidbom. 2014. "Democracy, Redistribution, and Political Participation: Evidence From Sweden 1919–1938." *Econometrica* 82, no. 3: 961–993.

Imbens, G., and S. Wager. 2019. "Optimized Regression Discontinuity Designs." *Review of Economics and Statistics* 101, no. 2: 264–278.

Imbens, G., and T. Zajonc. 2009. "Regression Discontinuity Design With Vector-Argument Assignment Rules." Unpublished Paper.

Jones, M. C. 1993. "Simple Boundary Correction for Kernel Density Estimation." *Statistics and Computing* 3, no. 3: 135–146.

Keele, L. J., and R. Titiunik. 2015. "Geographic Boundaries as Regression Discontinuities." *Political Analysis* 23, no. 1: 127–155.

Lee, D. S. 2008. "Randomized Experiments From Non-Random Selection in US House Elections." *Journal of Econometrics* 142, no. 2: 675–697.

Lehmann, E. L., and J. P. Romano. 2022. *Testing Statistical Hypotheses*. 4th ed. Springer.

Londoño-Vélez, J., C. Rodríguez, and F. Sánchez. 2020. "Upstream and Downstream Impacts of College Merit-Based Financial Aid for Low-Income Students: Ser Pilo Paga in Colombia." *American Economic Journal: Economic Policy* 12, no. 2: 193–227.

Matsudaira, J. D. 2008. "Mandatory Summer School and Student Achievement." *Journal of Econometrics* 142, no. 2: 829–850.

McCrary, J. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142, no. 2: 698–714.

Narita, Y., and K. Yata. 2021. "Algorithm as Experiment: Machine Learning, Market Design, and Policy Eligibility Rules." Preprint, arXiv, December 6, https://doi.org/10.48550/arXiv.2104.12909.

Papay, J. P., J. B. Willett, and R. J. Murnane. 2011. "Extending the Regression-Discontinuity Approach to Multiple Assignment Variables." *Journal of Econometrics* 161, no. 2: 203–207.

Salti, N., J. Chaaban, W. Moussa, et al. 2022. "The Impact of Cash Transfers on Syrian Refugees in Lebanon: Evidence From a Multidimensional Regression Discontinuity Design." *Journal of Development Economics* 155: 102803.

Smith, J., M. Hurwitz, and C. Avery. 2017. "Giving College Credit Where It Is Due: Advanced Placement Exam Scores and College Outcomes." *Journal of Labor Economics* 35, no. 1: 67–147.

Snider, C., and J. W. Williams. 2015. "Barriers to Entry in the Airline Industry: A Multidimensional Regression-Discontinuity Analysis of AIR-21." *Review of Economics and Statistics* 97, no. 5: 1002–1022.

Wong, V. C., P. M. Steiner, and T. D. Cook. 2013. "Analyzing Regression-Discontinuity Designs With Multiple Assignment Variables: A Comparative Study of Four Estimation Methods." *Journal of Educational and Behavioral Statistics* 38, no. 2: 107–141.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Appendix S1:** Supplementary.pdf.