Global Testing in Multivariate Regression Discontinuity Designs*

Artem Samiahulin†

January 27, 2026

Abstract

Regression discontinuity (RD) designs with multiple running variables arise in a growing number of empirical applications, including geographic boundaries and multi-score assignment rules. Although recent methodological work has extended estimation and inference tools to multivariate settings, far less attention has been devoted to developing global testing methods that formally assess whether a discontinuity exists anywhere along a multivariate treatment boundary. Existing approaches perform well in large samples, but can exhibit severe size distortions in moderate or small samples due to the sparsity of observations near any particular boundary point. This paper introduces a complementary global testing procedure that mitigates the small-sample weaknesses of existing multivariate RD methods by integrating multivariate machine learning estimators with a distance-based aggregation strategy, yielding a test statistic that remains reliable with limited data. Simulations demonstrate that the proposed method maintains near-nominal size and strong power, including in settings where standard multivariate estimators break down. The procedure is applied to an empirical setting to demonstrate its implementation and to illustrate how it can complement existing multivariate RD estimators.

*I thank EunYi Chung, Marcelo Medeiros, Joshua Shea, and all of the participants of the UIUC Econometrics Lab Lunch Seminar and Applied Micro Research Lunch for thoughtful comments and suggestions.

†Department of Economics, University of Illinois at Urbana-Champaign

1 Introduction

Regression discontinuity (RD) designs are widely regarded as one of the "...leading quasi-experimental empirical strategies in economics, political science, education, and many other social and behavioral sciences" (Calonico, Cattaneo, and Titiunik (2014)). This paper develops a global testing framework for multivariate regression discontinuity designs, allowing researchers to determine whether discontinuities in conditional expectations or densities exist anywhere along a multivariate treatment boundary, even in settings with limited sample sizes. From a theoretical perspective, the paper establishes a global testing procedure whose validity does not rely heavily on precise estimation of local multivariate discontinuities. Practically, the proposed procedure exhibits good finite-sample size control even when the sample size is modest.

While traditional RD designs use a single running variable, recent research has increasingly explored settings with multiple running variables (Papay, Willett, and Murnane (2011), Reardon and Robinson (2012), and Wong, Steiner, and Cook (2013)). For example, Keele and Titiunik (2015) examine geographic RD designs, where treatment and control groups are separated by geographic boundaries in two-dimensional space. Matsudaira (2008) studies the effects of mandatory summer school, where treatment assignment depends on both math and reading scores. Londoño-Vélez, Rodríguez, and Sánchez (2020) explore a government tuition subsidy in Colombia where eligibility was determined by both merit (via a test score) and economic need (via an index of wealth). For more applications, see papers such as Frey (2019), Narita and Yata (2021), Elacqua et al. (2016), Egger and Wamser (2015), Evans (2017), Cohodes and Goodman (2014), and Becht, Polo, and Rossi (2016).

In response to these empirical applications, several methodological papers have been written to estimate treatment effects in multivariate RD settings. Notable contributions include Imbens and Wager (2019), who use convex optimization, Cheng (2023), who applies thin plate splines, and Gunsilius and Van Dijcke

(2023), who develop methods for cases where the treatment boundary is unknown. More recently, Cattaneo, Titiunik, and Yu (2025a) and Cattaneo, Titiunik, and Yu (2025b) explore using local polynomial regressions to estimate treatment effects along two-dimensional boundaries, as well as aggregated effects.

Despite these methodological developments, relatively little attention has been devoted to global testing: the task of determining whether a discontinuity exists at any point along a multivariate treatment boundary. A straightforward strategy is to estimate treatment effects at many boundary locations and construct uniform confidence intervals at each point, assessing whether zero lies outside any of the resulting bands. When the sample size is large, this approach is appealing: it not only detects whether a discontinuity is present but also reveals where along the boundary it occurs. As a result, uniform confidence bands provide a rich and granular view of underlying treatment effect patterns. However, their performance deteriorates when the number of running variables increases or when the sample size is modest. In such settings, the number of observations effectively informing each boundary point becomes small, leading to unreliable asymptotic approximations. Through a simple simulation in the next subsection, this standard approach is shown to exhibit poor finite-sample performance.

To combat this issue, this paper proposes a new approach to global testing where a multivariate estimator is used to pool individual treatment effects together, which allows for more robust global inference even with smaller sample sizes. In simulations, a test size closer to a target 5% is achieved even with a modest sample size. Even with two or more running variables, this approach can be used to test for discontinuities in multivariate conditional expectation functions and multivariate densities. This method is intended to complement existing multivariate RD estimators by providing a robust tool for assessing the reliability of estimated effects and confidence intervals.

The proposed global testing procedure works in two stages. First, a multivariate estimator is used to estimate whether the treatment effects along the boundary are positive or negative. The local linear forest estimator of Friedberg et al. (2020) is used to estimate treatment effect heterogeneity, while a variation of the random forest density estimator of Wen and Hang (2022) is used for the density discontinuity test. Afterwards, these estimates are incorporated into a univariate test statistic that is used in the final hypothesis test. This approach has a number of advantages. First, by pooling together estimates along the boundary, the asymptotics for the final test statistic are made more reliable than they would be for each estimate individually. Second, the random forest estimators help maintain reasonable performance even when the number of running variables becomes relatively large. Third, the final test statistic only uses the signs of the multivariate estimates rather than the complete point estimates. This makes the final test statistic robust even when the initial treatment effects are estimated poorly, which translates to powerful theoretical properties.

This paper makes two contributions. First, it contributes to the literature on multivariate regression discontinuity designs by proposing a global test that remains robust even in relatively small samples. Second, it contributes to the literature on bunching and manipulation testing in regression discontinuity designs. Several papers have been written on univariate manipulation tests (Cattaneo, Jansson, and Ma (2020), McCrary (2008), and Cattaneo, Jansson, and Ma (2024)). Crippa (2025) extends these types of tests to accommodate many running variables. However, this test only applies to boundaries like the one in Figure 1 in the next subsection. The test outlined in this paper can be applied to a wide range of boundary shapes, making it applicable even in settings with complex treatment rules, as in geographic regression discontinuities.

The remainder of this paper is organized as follows. Subsection 1.1 discusses a simulation example that

demonstrates where the uniform confidence bands approach fails, while subsection 1.2 discusses notation and assumptions. Section 2 discusses identification for both the conditional expectation and density cases. Section 3 discusses estimation and inference for the final test statistics used for global testing. Section 4 shows simulation results that demonstrate the properties of the outlined global testing procedure. Section 5 discusses an application, and section 6 concludes.

1.1 Simple Simulation Example

To illustrate the limitations of the uniform confidence band approach in finite samples, consider the following process:

$$Y_i = \frac{X_{1,i} + X_{2,i}}{3} + \epsilon_i$$

where $\epsilon_i \sim \mathcal{N}(0, 0.05)$, $X_{1,i}, X_{2,i}$ come from independent uniform distributions on the interval $[-1, 1]$, and all observations are i.i.d across i. Further, define the treatment indicator as follows:

$$T_i = \begin{cases} 1, & X_{1,i}, X_{2,i} \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

For this treatment rule, the treatment region looks as follows:

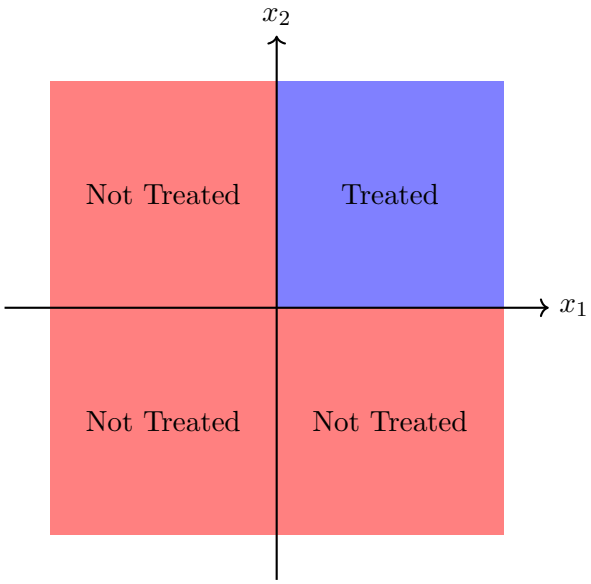


Figure 1: Treatment and control regions for simulated treatment rule.

The goal is to test whether there is a discontinuity in $\mathbb{E}[Y_i|X_i = x]$ for some x along the boundary between the treated and non-treated regions. To do this, the local polynomial estimator of Cattaneo, Titiunik, and Yu (2025b) is used to estimate treatment effect at 79 evenly-spaced points along the boundary and apply 95% uniform confidence bands for inference. Afterwards, rejection rates are computed over 5,000 simulations. Since there are no discontinuities in this example, the rejection rate should be approximately 5%. To assess robustness, the rejection rates from the sum of treatment effects and sum of squared treatment effects are also reported. Additionally, these three approaches are performed only on "safe" points $x = (0.5, 0)$ and $x = (0, 0.5)$ to avoid estimation at edges and kink points. Using number of observations $n = 20,000$ and $n = 1,000$, the rejection rates are as follows:

Table 1: Simulated Rejection Rates for Bivariate Local Polynomial Regression

	Uniform Bands		Sum of $\hat{\tau}(x)$		Sum of $\hat{\tau}(x)^2$	
n	Full	Safe	Full	Safe	Full	Safe
1,000	0.3652	0.1540	0.1006	0.1080	0.1714	0.1276
20,000	0.0748	0.0560	0.0564	0.0566	0.0616	0.0598

In the case where $n = 20,000$, the rejection rates are fairly close to 0.05. However, with a smaller number of observations, the rejection rate becomes much larger than the expected rejection rate of 0.05. Summation mitigates this problem only partially, because the aggregated statistic remains sensitive to inaccuracies in the individual local estimates used to construct it.

1.2 Setup and Notation

Let Y_i denote the outcome variable for observation i and let X_i be a vector of running variables with support $\mathbb{X}$. Let Ω denote the set of treated units and let $\partial\Omega$ be the boundary of Ω. Let $T_i = 1$ for $i \in \Omega$ and $T_i = 0$ for $i \notin \Omega$. Also, let $Y_i(t)$ denote the potential outcome for unit i with treatment status t. Let $g_\Omega(X_i)$ be the signed (Euclidean) distance function relative to $\partial\Omega$, with $f_{g_\Omega}(g)$ being the density of $g_\Omega(X_i)$. For simplicity, define $g_\Omega(X_i) \equiv g_i$. With that, here are some necessary assumptions:

Assumption 1. *The set Ω is closed (without loss of generality), bounded, and has a piecewise smooth boundary.*

Assumption 2. *All observations are independently and identically distributed (i.i.d) across i.*

Assumption 3. *All Y_i, X_i take values in bounded supports.*

Assumption 4. $\mathbb{V}(Y_i(1)|X_i = x), \mathbb{V}(Y_i(0)|X_i = x) > 0$ *are continuous and bounded for all $x \in \mathbb{X}$ within some neighborhood of $\partial\Omega$.*

Assumption 5. *The density $f_{g_\Omega}(g)$ is bounded away from zero in some neighborhood around $g = 0$.*

Assumption 6. *The kernel function $K(\cdot)$ is a bounded and symmetric function that takes values on the interval $[-1, 1]$.*

Assumption 7. *$\mathbb{E}[Y_i(t)|X_i = x]$ is twice continuously differentiable and bounded for all $x \in \mathbb{X}$ and for all $t \in \{0, 1\}$.*

Assumption 8. *The joint density of the running vector, $f_X(\cdot)$, is twice continuously differentiable and bounded within some neighborhood of $x \in \partial\Omega$.*

Assumption 9. *The joint density of the running vector, $f_X(\cdot)$, is twice continuously differentiable and bounded above and away from zero within some neighborhood of $x \in \partial\Omega$, except where $x \in \partial\Omega$.*

2 Identification

This section develops the identification framework for two proposed global test statistics, covering both conditional expectation discontinuities (treatment effect heterogeneity) and density discontinuities (manipulation).

2.1 Treatment Effect Heterogeneity

Define $\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]$. Here, the hypothesis of interest is as follows:

$$H_0 : \tau(x) = 0 \text{ for all } x \in \partial\Omega$$

$$H_a : \tau(x) \neq 0 \text{ for some } x \in \partial\Omega$$

To test the hypothesis above without uniform confidence bands, the running vector X_i is first aggregated into a scalar g_i. Using g_i as a running variable leads to the following:

Theorem 1. *Under Assumptions 1-3, 5, and 7, we have:*

$$\lim_{\epsilon \to 0^+} \mathbb{E}[Y_i \mid g_i = \epsilon] - \lim_{\epsilon \to 0^-} \mathbb{E}[Y_i \mid g_i = \epsilon] = \frac{1}{f_{g_\Omega}(0)} \int_{x: g_\Omega(x)=0} \tau(x) f_X(x)\, dS_\Omega,$$

where dS_Ω denotes a Hausdorff measure over $\partial\Omega$.

Proofs of Theorem 1 and all subsequent theorems are deferred to the appendix. Intuitively, Theorem 1 implies that when the signed distance function is used as a running variable, the standard RD difference identifies a weighted average across $\tau(x)$ for all $x \in \partial\Omega$. Although this quantity may be of interest to researchers, it is not sufficient to test the main hypothesis because $\mathbb{E}[Y_i(1) - Y_i(0) | g_i = 0]$ may equal zero even when there exists points $x \in \partial\Omega$ where $\tau(x) \neq 0$. This can happen when the positive discontinuities mix with negative discontinuities to produce a null result. To address this issue, let $\gamma(X_i) \equiv \gamma_i$ denote the closest boundary point to X_i and let $\Gamma(X_i) \equiv \Gamma_i = 1$ if $\tau(\gamma(X_i)) \geq 0$ and $\Gamma_i = 0$ otherwise. This leads to the following result:

Theorem 2. *Under Assumptions 1-3, 5, and 7, we have:*

$$\begin{aligned}
\tau_\Gamma \equiv &\left(\lim_{\epsilon \to 0^+} \mathbb{E}[Y_i \mid g_i = \epsilon, \Gamma_i = 1] - \lim_{\epsilon \to 0^-} \mathbb{E}[Y_i \mid g_i = \epsilon, \Gamma_i = 1] \right) \mathbb{E}[\Gamma_i | g_i = 0] - \\
&\left(\lim_{\epsilon \to 0^+} \mathbb{E}[Y_i \mid g_i = \epsilon, \Gamma_i = 0] - \lim_{\epsilon \to 0^-} \mathbb{E}[Y_i \mid g_i = \epsilon, \Gamma_i = 0] \right) (1 - \mathbb{E}[\Gamma_i | g_i = 0]) \\
= &\frac{1}{f_{g_\Omega}(0)} \int_{x: g_\Omega(x)=0} |\tau(x)| f_X(x)\, dS_\Omega = \frac{1}{f_{g_\Omega}(0)} \int_{x: g_\Omega(x)=0} (2\Gamma(x) - 1)\tau(x) f_X(x)\, dS_\Omega
\end{aligned}$$

Here, τ_Γ represents the weighed average of $|\tau(x)|$ over $x \in \partial\Omega$. Intuitively, Γ_i serves as a sign-normalization factor: it reverses the sign of Y_i whenever the corresponding treatment effect is negative, which aligns all effects in the same direction. In other words, the original hypothesis can be rewritten as:

$$H_0 : \tau_\Gamma = 0$$

$$H_a : \tau_\Gamma > 0$$

To make estimation easier and more efficient, notice that τ_Γ can be rewritten as:

$$\tau_\Gamma = \lim_{\epsilon \to 0^+} \mathbb{E}[(2\Gamma_i - 1) Y_i | g_i = \epsilon] - \lim_{\epsilon \to 0^-} \mathbb{E}[(2\Gamma_i - 1) Y_i | g_i = \epsilon]$$

A key advantage of this approach is that it only uses the sign of $\tau(\gamma(X_i))$. As a result, Γ_i can be estimated reliably even if the underlying $\tau(\gamma(X_i))$ is noisy, which is an important property in small samples or in settings with many running variables. In contrast, procedures that aggregate raw treatment effects along the boundary are fragile: a single poorly estimated effect can materially distort the resulting test statistic. This phenomenon is evident in Table 1.

2.2 Density Manipulation

Consider the problem of testing whether the joint density of the running variables is continuous at the boundary. Such a test serves as a diagnostic for endogenous manipulation of the assignment variables. In this sense, the density-discontinuity test of McCrary (2008) is being extended to a multivariate setting. Define the following:

$$f_X^+(x) = \lim_{x' \to x, g_\Omega(x') \geq 0} f_X(x'), \quad f_X^-(x) = \lim_{x' \to x, g_\Omega(x') < 0} f_X(x')$$

$$\tau_f(x) = f_X^+(x) - f_X^-(x)$$

In this case, the hypothesis of interest is as follows:

$$H_0 : \tau_f(x) = 0 \text{ for all } x \in \partial\Omega$$

$$H_a : \tau_f(x) \neq 0 \text{ for some } x \in \partial\Omega$$

Using a change-of-variables for densities:

$$f_{g_\Omega}(g) = \int_{x: g_\Omega(X) = g} f_X(x) dS_\Omega \implies \lim_{\epsilon \to 0^+} f_{g_\Omega}(\epsilon) - \lim_{\epsilon \to 0^-} f_{g_\Omega}(\epsilon) = \int_{x: g_\Omega(x) = 0} \tau_f(x) dS_\Omega$$

Let $\Lambda(X_i) \equiv \Lambda_i = 1$ if $\tau_f(\gamma(X_i)) \geq 0$ and $\Lambda_i = 0$ otherwise. With that, the result below follows:

Theorem 3. *Under Assumption 1-3:*

$$\tau_\Lambda^f \equiv \left[\lim_{\epsilon\to 0^+} f_{g_\Omega|\Lambda}(g=\epsilon|\Lambda_i=1) - \lim_{\epsilon\to 0^-} f_{g_\Omega|\Lambda}(g=\epsilon|\Lambda_i=1)\right]\mathbb{P}(\Lambda_i=1) -$$
$$\left[\lim_{\epsilon\to 0^+} f_{g_\Omega|\Lambda}(g=\epsilon|\Lambda_i=0) - \lim_{\epsilon\to 0^-} f_{g_\Omega|\Lambda}(g=\epsilon|\Lambda_i=0)\right]\mathbb{P}(\Lambda_i=0)$$
$$= \int_{X:g_\Omega(X)=0} |\tau_f(x)| dS_\Omega = \int_{X:g_\Omega(X)=0} (2\Lambda(X)-1)\tau_f(x) dS_\Omega$$

Here, τ_Λ^f is equivalent to the sum of $|\tau_f(x)|$ over all $x \in \partial\Omega$. Like in the boundary heterogeneity case, a multivariate density test can be written through the following hypothesis:

$$H_0 : \tau_\Lambda^f = 0$$

$$H_a : \tau_\Lambda^f > 0$$

To make this expression easier to estimate, τ_Λ^f can be rewritten as follows:

$$\tau_\Lambda^f = \lim_{\epsilon\to 0^+} f_{(2\Lambda-1)g}(\epsilon) - \lim_{\epsilon\to 0^-} f_{(2\Lambda-1)g}(\epsilon)$$

In other words, τ_Λ^f can be computed as the discontinuity in the density of $g_i^* \equiv (2\Lambda_i - 1)g_i$ at $g_i^* = 0$.

3 Estimation and Inference

The following subsections describe estimation and inference procedures for both the treatment effect heterogeneity test and the density manipulation test. Although the two procedures are closely related, there are several important distinctions between them.

3.1 Treatment Effect Heterogeneity

Based on the identification results of the previous section, the quantity of interest is the following:

$$\tau_\Gamma = \lim_{\epsilon\to 0^+} \mathbb{E}[Y_i^*|g_i=\epsilon] - \lim_{\epsilon\to 0^-} \mathbb{E}[Y_i^*|g_i=\epsilon]$$

where $Y_i^* = (2\Gamma(X_i) - 1)Y_i$. However, $\Gamma(X_i)$ is not observed, so it will need to be estimated as $\Gamma_n(X_i)$. Now define $\hat{Y}_i = (2\Gamma_n(X_i) - 1)Y_i$. With that, notice the following:

$$\begin{aligned}\hat{Y}_i - Y_i^* &= Y_i(2\Gamma_n(X_i) - 1 - (2\Gamma(X_i) - 1)) \\ &= 2Y_i(\Gamma_n(X_i) - \Gamma(X_i)) \\ &\equiv \hat{\eta}_i\end{aligned}$$

In other words, $\hat{\eta}_i$ can be thought of as a measurement error term, where Γ_i is a higher dimensional nuisance parameter. As in Chernozhukov et al. (2018), estimating Γ_i and τ_Γ using the same data will cause problems for estimation and inference. To address these issues, $\hat{Y}_i$ will be generated as follows:

1. Split the n observations into K folds.
2. Denote all observations in fold $k \in \{1, 2, ..., K\}$ as I_k. For each fold k, use data from all other folds to estimate $\Gamma(\cdot)$. Call this estimate $\Gamma_{n,-k}(\cdot)$.
3. Define $\hat{Y}_{i,k} = (2\Gamma_{n,-k}(X_{i,k}) - 1)Y_{i,k}$

where the index k denotes the fold of observation i. To estimate $\Gamma(\cdot)$, the Local Linear Forest of Friedberg et al. (2020) will be used. The advantage of this method is that it easily allows for efficient boundary point estimation with any number of running variables.

Lemma 1. *Let $\mu_n^{(llf)}(x_0)$ represent a local linear forest estimator of $\mu(x_0) \equiv \mathbb{E}[Y_i(t)|X_i = x_0]$ at boundary point x_0. Under assumptions 1-4 and 7-8, the local linear forest estimator is consistent for $\mu(x_0)$.*

Once the outcome variable is defined, the fold-level version of τ_Γ (defined as $\tau_{\Gamma_{k,p}}$) will be estimated as follows:[1]

$$\begin{aligned}
\hat{\tau}_{\Gamma_{k,p}} &= e_1'(\hat{\beta}^+_{k,p} - \hat{\beta}^-_{k,p}) \\
\hat{\beta}^\pm_{k,p} &= \underset{\beta \in \mathbb{R}^{p+1}}{\arg\min} \sum_{i \in I_k} \delta^\pm_{i,k} \left(\hat{Y}_{i,k} - r_p(g_{i,k})\beta\right)^2 K\left(\frac{g_{i,k}}{h_k}\right) \\
\delta^+_{i,k} &= \mathbb{1}(g_{i,k} \geq 0), \delta^-_{i,k} = \mathbb{1}(g_{i,k} < 0)
\end{aligned}$$

where $K(\cdot)$ is some bounded, symmetric kernel function, $r_p(\cdot)$ is a vector of polynomial terms up to order p, and h_k is the bandwidth. In this case, $K(\cdot)$ will be an Epanechnikov kernel and $p = 1$. Additionally, the multiplier bootstrap approach of Chiang, Hsu, and Sasaki (2019) will be used for inference and bandwidth selection, with polynomial order $q = 2$ used for bias correction. More specifically, bootstrap estimates will be created as follows:

$$\phi^\pm_{b,q} = \sum_{i=1}^n u_{i,b} \frac{e_1'(\Gamma^\pm_q)^{-1}(\hat{Y}_{i,k} - r_q(g_{i,k})\hat{\beta}^\pm_{k,q}) r_q\left(\frac{g_{i,k}}{h_k}\right) K\left(\frac{g_{i,k}}{h_k}\right) \delta^\pm_{i,k}}{n h_k \hat{f}_g(0)}$$

where $u_{i,b}$ are independent and take values in $\{-1, 1\}$ (with equal probability), and $\Gamma^\pm_p$ is defined as follows:

$$\Gamma^+_p = \int_0^1 r_p(u) r_p(u)' K(u) du, \quad \Gamma^-_p = \int_{-1}^0 r_p(u) r_p(u)' K(u) du$$

The variance of $\hat{\tau}_{\Gamma_{k,q}}$ can be computed using the variance of $\phi^+_{b,q} - \phi^-_{b,q}$. With these components, the result below follows:[2]

Theorem 4. *Under assumptions 1-8, the following holds:*

$$\frac{1}{K}\sum_{k=1}^K \hat{\tau}_{\Gamma_{k,p}} \xrightarrow{p} \tau_\Gamma, \quad \frac{\frac{1}{K}\sum_{k=1}^K \hat{\tau}_{\Gamma_{k,q}} - \tau_\Gamma}{\sqrt{\frac{1}{B-1}\sum_{b=1}^B (\phi^+_{b,q} - \phi^-_{b,q} - (\bar{\phi}^+_q - \bar{\phi}^-_q))^2}} \xrightarrow{d} \mathcal{N}(0,1)$$

where

$$\bar{\phi}^\pm_q = \frac{1}{B}\sum_{b=1}^b \phi^\pm_{b,q}$$

[1]The superscript $\pm$ is used as shorthand, with $+$ indexing quantities computed for observations with $g_{i,k} \geq 0$ and $-$ indexing quantities computed for observations with $g_{i,k} < 0$.

[2]The theorem below can be trivially modified to accommodate two different bandwidths on either side of the cutoff.

An important observation is that asymptotic normality holds even though the local linear forest estimator converges more slowly than the second-stage local polynomial regression. The key reason is that, when a true discontinuity exists, inference depends only on correctly determining the signs of the treatment effects. Because the data lie on a bounded support, the probability of sign errors can be made arbitrarily small, regardless of the slower convergence rate of the first-stage estimator. As a result, the asymptotic properties of the standard local polynomial regression remain valid.

3.2 Density Test

Estimation and inference for the aggregated univariate density are carried out using a bias-corrected kernel density estimator, with sampling variability assessed via a multiplier bootstrap. A standard kernel density estimator is given as follows:

$$\hat{f}_{g^*}(0)^+ - \hat{f}_{g^*}(0)^- = \sum_{i=1}^{n} \left(\frac{1}{nh_k}\right) K\left(\frac{\hat{g}_{i,k}}{h_k}\right)(2\delta_i - 1)$$

where $K(\cdot)$ is a symmetric and bounded kernel function, and $\delta_i = \mathbb{1}(\hat{g}_{i,k} \geq 0)$. Since the estimator is known to exhibit boundary bias, an approach similar to Hazelton and Marshall (2009) is adopted, whereby the base kernel is adjusted to achieve unbiasedness at the boundary. The resulting estimator is given by:

$$\hat{f}_{g^*}(0)^+ - \hat{f}_{g^*}(0)^- = \sum_{i=1}^{n} \left(\frac{1}{nh_k}\right)\left(\alpha_1 + \alpha_2 \left|\frac{\hat{g}_{i,k}}{h_k}\right|\right) K\left(\frac{\hat{g}_{i,k}}{h_k}\right)(2\delta_i - 1)$$

where α_1, α_2 solves the following system of equations:

$$\begin{pmatrix} \int_0^1 K(u)du & \int_0^1 uK(u)du \\ \int_0^1 uK(u)du & \int_0^1 u^2K(u)du \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

With this construction of α_1 and α_2, the boundary bias inherent in the standard kernel density estimator is eliminated, as the modified kernel integrates to one on the relevant side of the cutoff. These coefficients also remove the first-order bias, yielding behavior analogous to that of a kernel density estimator evaluated

at an interior point. The optimal bandwidth is selected using the following MSE-optimal expression:

$$h_k = \left(\frac{V_k}{B_k^2}\right)^{\frac{1}{5}}$$

where V_k is the variance term and B_k is the bias term. Their exact expressions are given as follows:

$$V_k = \frac{1}{n}\left(\hat{f}_g(0)^+ \int_0^1 (\alpha_1 + \alpha_2 u)^2 K(u)^2 du + \hat{f}_g(0)^- \int_{-1}^0 (\alpha_1 - \alpha_2 u)^2 K(u)^2 du\right)$$
$$B_k = \left(\int_0^1 u^2(\alpha_1 + \alpha_2 u) K(u) du\right) \hat{f}_g''(0)^+ - \left(\int_{-1}^0 u^2(\alpha_1 - \alpha_2 u) K(u) du\right) \hat{f}_g''(0)^-$$

Here, the density and second derivative of the density will be estimated via preliminary local polynomial regressions (Cattaneo, Jansson, and Ma (2020) and Cattaneo, Jansson, and Ma (2024)). For consistency of this aggregated estimator, a variation of the random forest density estimator used to construct $\hat{g}_{i,k}$ must be consistent. When joint density estimation is conducted at boundary points, the volume of each cell generated by the partitioning algorithm of Wen and Hang (2022) must be adjusted, particularly when the support of the running variables is non-rectangular. The effective volume of a cell is obtained by multiplying its original volume by the proportion of the cell that lies within the support of the running variables. This proportion is approximated by uniformly sampling points within the cell and computing the fraction that fall inside the support. With that, the following lemma holds:

Lemma 2. *Without loss of generality, let $f_X(x|x \in \Omega)$ be the density of the running vector X within the region Ω. Also, let $\hat{f}_X(x|x \in \Omega)$ be the random forest density estimator of $f_X(x|x \in \Omega)$. Under assumptions 1-3 and 9, the random forest density estimator is consistent for $f_X(x|x \in \Omega)$.*

As with the treatment effect heterogeneity case, a K-fold cross-fitting procedure is used to separate the estimation of $\hat{g}_{i,k}$ and the final test statistic. Also, as with other kernel-based nonparametric estimators, the MSE-optimal bandwidth cannot be used directly for inference because it does not adequately control for bias. To address this issue, the kernel bias-correction procedure of Calonico, Cattaneo, and Farrell

(2018) is employed. Specifically, inference is conducted using the following re-centered estimator:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_k}\left[\left(\alpha_1+\alpha_2\left|\frac{\hat{g}_{i,k}}{h_k}\right|\right)K\left(\frac{\hat{g}_{i,k}}{h_k}\right)-L^{(2)}\left(\frac{\hat{g}_{i,k}}{h_k}\right)\int_0^1\frac{u^2(\alpha_1+\alpha_2 u)K(u)}{2!}du\right](2\delta_i-1)$$

where

$$L(u)=30(1-|u|)^2u^2\mathbb{1}(|u|\leq 1)$$

Here, $L(u)$ is designed to estimate the curvature of the given density at the boundary point $g=0$. With that, the following will be used to generate bootstrap replications to compute standard errors:

$$\begin{aligned}\phi_b^{\pm}=\sum_{i=1}^{n}\frac{u_i^{(b)}}{nh_k}\Bigg[&\left(\alpha_1+\alpha_2\left|\frac{\hat{g}_{i,k}}{h_k}\right|\right)K\left(\frac{\hat{g}_{i,k}}{h_k}\right)-h_k\hat{f}_g(0)^{\pm}\\&-\left(L^{(2)}\left(\frac{\hat{g}_{i,k}}{h_k}\right)-h_k\hat{f}_g''(0)^{\pm}\right)\int_0^1\frac{u^2(\alpha_1+\alpha_2 u)K(u)}{2!}du\Bigg](2\delta_i-1)\end{aligned}$$

where

$$\hat{f}_g(0)^{\pm}=\begin{cases}\hat{f}_g(0)^{+}, & \delta_i=1\\ \hat{f}_g(0)^{-}, & \delta_i=0\end{cases},\quad \hat{f}_g''(0)^{\pm}=\begin{cases}\hat{f}_g''(0)^{+}, & \delta_i=1\\ \hat{f}_g''(0)^{-}, & \delta_i=0\end{cases}$$

Now, the result below follows:[3]

Theorem 5. *Under assumptions 1-3, 5-7 and 9, the following result holds:*

$$\frac{1}{K}\sum_{k=1}^{K}\hat{\tau}_{\Lambda_k}\xrightarrow{p}\tau_{\Lambda}^{f},\quad \frac{\frac{1}{K}\sum_{k=1}^{K}(\hat{\tau}_{\Lambda_k}-\mathcal{B}_k)-\tau_{\Lambda}^{f}}{\sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\phi_b^{+}-\phi_b^{-}-(\bar{\phi}^{+}-\bar{\phi}^{-}))^2}}\xrightarrow{d}\mathcal{N}(0,1)$$

where $\mathcal{B}_k$ is the estimated bias from fold k.

As in the treatment effect heterogeneity setting, the asymptotic results for the bias-corrected kernel density estimator continue to hold. This is because the probability of assigning an incorrect sign to the estimated discontinuity can be driven to zero as the sample size grows, ensuring that such errors do not affect the limiting distribution.

[3]The theorem below can be trivially modified to accommodate two different bandwidths on either side of the cutoff.

3.3 Accounting for Random Split Selection

Asymptotically, the specific split used in the K-fold cross-fitting procedure is inconsequential. In finite samples, however, different splits can lead to non-negligible variation in both point estimates and standard errors. To account for this variability, the K-fold cross-fitting procedure is repeated S times, and the resulting main estimators are averaged across the S splits. For inference, influence functions are computed for each split; multiplier weights are then applied to the average influence function, with the weights for each observation i held fixed across splits. This construction parallels the clustered multiplier bootstrap extension of Chiang, Hsu, and Sasaki (2019) and incorporates the randomness of the sample split into the standard error calculation.

4 Simulations

This section illustrates the implementation of the treatment effect heterogeneity test and the density manipulation tests using simulated data. For each case, two data-generating processes (DGPs) are considered: one that is discontinuous almost everywhere, and one that is continuous everywhere.

4.1 Treatment Effect Heterogeneity Simulations

The first DGP will be constructed as follows:

$$Y = \frac{1}{3}\begin{cases} X_1 + X_2, & X_1, X_2 \in [0,1] \\ 1 - X_1, & X_1 \in [0,1], X_2 \in [-1,0] \\ 1 - X_2, & X_1 \in [-1,0], X_2 \in [0,1] \\ 1, & X_1, X_2 \in [-1,0] \end{cases} + \epsilon$$

where $\epsilon \sim N(0, \sigma_\epsilon^2)$, X_1, X_2 take values in $[-1,1]$, and $f_X(x) = 0.25$ with $X_1 \perp\!\!\!\perp X_2$. Here, $\tau_\Gamma = \frac{1}{6}$. Below is a plot of this function:

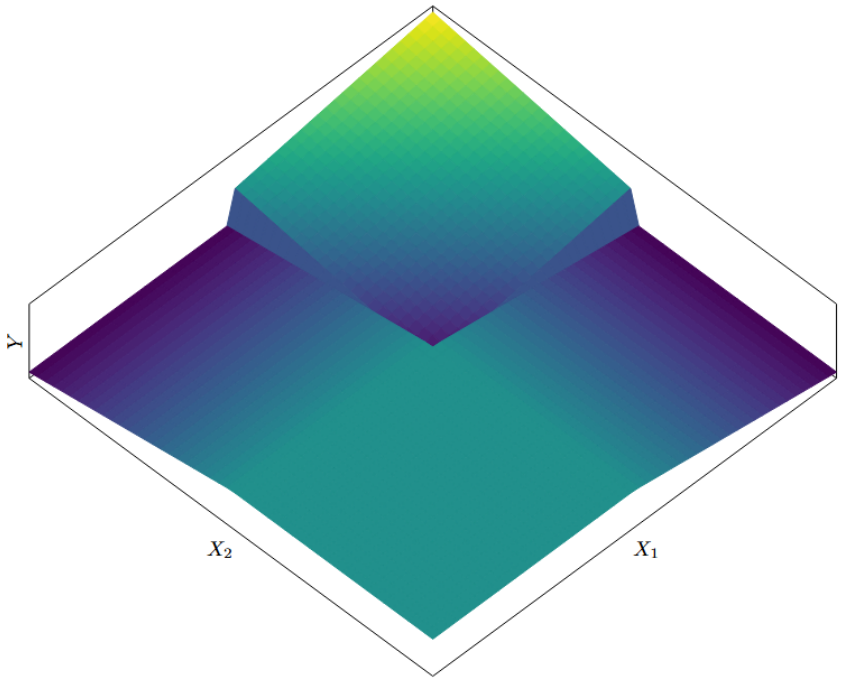


Figure 2: Graph of the DGP 1 function. Darker colors refer to lower values and lighter colors refer to higher values.

For the second DGP, the data will be generated as follows:

$$Y = \frac{X_1 + X_2}{3} + \epsilon$$

where X_1, X_2, and ϵ are constructed in the same way as in the first DGP. For each of these DGPs, the following parameters are used: sample size $n = 1000$, noise variance $\sigma_\epsilon^2 = 0.05$, number of folds $K = 2$, number of splits $S \in \{1, 10, 20\}$, and 5,000 simulation replications. Also, two different coverage-error optimal bandwidths are used on either side of the one-dimensional cutoff. The resulting simulation outcomes are presented below.

Table 2: Simulation results for treatment effect heterogeneity test using DGP 1 and DGP 2.

	DGP 1	**DGP 2**
Bias		
Splits = 1	-0.0295	-0.0003
Splits = 10	-0.0295	-0.0016
Splits = 20	-0.0294	-0.0017
Standard Errors		
Splits = 1	0.0895	0.0902
Splits = 10	0.0743	0.0539
Splits = 20	0.0734	0.0510
Rejection Rate		
Splits = 1	0.5252	0.0778
Splits = 10	0.6412	0.0774
Splits = 20	0.6518	0.0762
Rejection Rate (Local Polynomial Regression from Table 1)		
$n = 1,000$		0.3652
$n = 20,000$		0.0748

These results reveal several patterns. First, the bias under the first DGP exceeds that under the second. This occurs because, in the second DGP, the sign of the outcome is immaterial in the absence of discontinuities. In contrast, under the first DGP, misclassification of the treatment effect sign has a non-trivial influence on the test statistic, leading to greater bias. Consequently, DGPs featuring discontinuities are expected to exhibit slightly larger bias. This bias does not seem to change on average as the number of

splits increases.

Second, the standard errors exhibit substantial sensitivity to the number of splits. For both DGPs, the standard error is approximately 0.09 with a single split, but decreases markedly when the number of splits is increased to 10. Although the initial reduction is pronounced, further increases in the number of splits yield progressively smaller declines. For the discontinuous DGP, this reduction improves the rejection rate, whereas the rejection rate remains relatively stable for the continuous DGP.

4.2 Density Manipulation Simulations

For the first density test example, the following bivariate density will be used:

$$f_X(X_1, X_2) = \frac{1}{3} \begin{cases} X_1 + X_2, & X_1, X_2 \in [0, 1] \\ 1 - X_1, & X_1 \in [0, 1], X_2 \in [-1, 0] \\ 1 - X_2, & X_1 \in [-1, 0], X_2 \in [0, 1] \\ 1, & X_1, X_2 \in [-1, 0] \end{cases}$$

Note that this is the same function that was used in DGP 1 for the boundary heterogeneity simulations. Here, $\tau_\Lambda^f = 1/3$. For the second density, the following joint density will be used:

$$f_X(X_1, X_2) = 1/4$$

where X_1, X_2 are independent and take values on $[-1, 1]$. Here, the first density is discontinuous almost everywhere along the treatment boundary, whereas the second density is continuous everywhere. For bandwidth selection, a single MSE optimal bandwidth is used on either side of the one-dimensional cutoff. For $n = 1000$, $K = 2$ and $S = 1, 10, 20$, notice the following simulation results:

Table 3: Simulation results for density test using DGP 1 and DGP 2.

	DGP 1	DGP 2
Bias		
Splits = 1	-0.035	-0.002
Splits = 10	-0.036	-0.001
Splits = 20	-0.035	-0.002
Standard Errors		
Splits = 1	0.135	0.167
Splits = 10	0.111	0.095
Splits = 20	0.109	0.090
Rejection Rate		
Splits = 1	0.721	0.059
Splits = 10	0.862	0.063
Splits = 20	0.874	0.061

In this scenario, the simulation results exhibit a pattern similar to that observed in the treatment effect heterogeneity case. Specifically, the bias is slightly greater under the discontinuous DGP, and the standard errors decrease up to a certain point as the number of splits increases.

5 Application

The empirical application in this paper follows the setting of Frey (2019). In 2003, the conditional cash transfer (CCT) program *Bolsa Família* (BF) was launched across Brazil. Beyond serving as a major source of income stabilization for extremely poor households, there is substantive speculation that BF also re-

duced the scope for political influence by diminishing incentives to engage in clientelism.[4] To evaluate this claim, Frey (2019) exploit cross-municipal variation in CCT coverage induced by the Family Health Program (FHP). Beginning in August 2004, municipalities with fewer than 30,000 residents and a Human Development Index (HDI) below 0.7 became eligible for a 50% increase in FHP funding. This expansion influenced BF coverage by improving households' access to information about the program. According to a survey of 10,000 poor households, more than 10% of BF beneficiaries learned about the program through their family doctor. Consequently, the political effects of BF can be identified using variation in BF coverage generated by the FHP expansion through an instrumental variables or regression discontinuity analysis.

In the original Frey (2019) study, the author implements a multivariate regression discontinuity design along the FHP funding threshold to assess the strength of the first stage and to estimate the effects of BF on various political outcomes. However, with only a few thousand municipalities, a full boundary-based treatment effect analysis may lack reliability due to limited effective sample size near each cutoff point. To evaluate the robustness of the original findings, the methods of this paper will be applied to test for the presence of any treatment effects along the FHP boundary for the same outcomes examined in Frey (2019). For simplicity, only municipalities from 2008 are included in this analysis.

For the first set of results, the outcomes will be health funds received and CCT coverage. These outcomes measure the extent to which the intervention affects the type of treatment that was received. The first column reports the weighted average absolute treatment effect, with p-values in brackets. The second column presents the weighted average treatment effect based on simple signed distances (as in Theorem 1). The third column indicates whether a discontinuity is detected using the uniform bands of Cattaneo,

[4]Here, clientelism refers to the act of "...replacing public good distribution with private transfers targeting groups or individual voters" (Frey (2019)).

Titiunik, and Yu (2025b). The fourth column reports whether both positive and negative discontinuities are detected using the local polynomial regression approach of Cattaneo, Titiunik, and Yu (2025b). All running variables are standardized using the standard deviation.

Table 4: Treatment effect heterogeneity test for first-stage outcomes.

Variable	**Est.**	**Est. (Dist.)**	**LPR Disc.**	**Double Disc.**
Health Funds, R$mn	0.352	0.196	No	No
	[0.714]	[0.436]		
CCT Coverage, p.p.	3.217	4.036	No	No
	[0.426]	[0.373]		

Note: P-values are reported in square brackets.

Unlike Frey (2019), the analysis presented here does not uncover evidence of a strong first stage. This conclusion holds both when applying the uniform confidence bands of Cattaneo, Titiunik, and Yu (2025b) and when using the aggregation methods developed in this paper. Together, these results indicate that the available data do not provide sufficient support for a strong first-stage effect. At the same time, there may be scope to improve the power of these tests without compromising size. As noted by Crippa (2025) and Wong, Steiner, and Cook (2013), the scaling of the running variables can meaningfully affect the performance of distance-based aggregators. Intuitively, rescaling alters the relative emphasis placed on different segments of the boundary. A common remedy is to standardize each running variable so that all coordinates receive equal weight. However, there is no reason to expect equal weighting to be optimal for power. For instance, if discontinuities occur primarily along one portion of the boundary, assigning greater weight to that region may deliver more powerful tests than treating all boundary segments symmetrically. One possible way to address this issue is to treat the scaling parameters as tuning choices and select them to maximize power. Although this extension lies beyond the scope of this paper, it may improve the power

of the proposed test and potentially reveal first-stage discontinuities that remain undetected under equal weighting.

For the second set of results, the analysis will use political outcomes that will help determine whether BF had any impact on clientelism in Brazil.

Table 5: Treatment effect heterogeneity test for main political outcomes.

Variable	**Est.**	**Est. (Dist.)**	**LPR Disc.**	**Double Disc.**
Incumbent's Vote Share (%)	4.408	-2.163	No	No
	[0.135]	[0.528]		
Margin of Victory (p.p.)	0.039	-4.650	No	No
	[0.288]	[0.248]		
Candidates (Number)	-0.059	-0.056	No	No
	[0.663]	[0.531]		
Pro-Poor Spending (% share)	11.945**	2.643	Yes	Yes
	[0.015]	[0.173]		
Challenger's Entry (share with HS)	-0.017	-0.075	No	No
	[0.653]	[0.305]		
Challenger's Entry (share clientelistic)	-0.065	0.092	No	No
	[0.990]	[0.252]		
Challenger is Top 2 (share with HS)	-0.025	-0.155	No	No
	[0.530]	[0.124]		
Challenger is Top 2 (share clientelistic)	-0.090	0.124	No	No
	[0.996]	[0.197]		

Note: P-values are reported in square brackets.

In this case, all but one of the political outcomes retain their significance from Frey (2019). In particular, pro-poor spending exhibits a clear discontinuity at the FHP boundary. Notably, both the local polynomial estimator and the ML-based distance aggregation approach detect this discontinuity, whereas the standard distance-based aggregation estimator does not. A closer examination of the local polynomial estimates

suggests that pro-poor spending increases at some boundary locations and decreases at others. Below is a figure showing heterogeneity in the treatment effects estimated by the local polynomial regression.

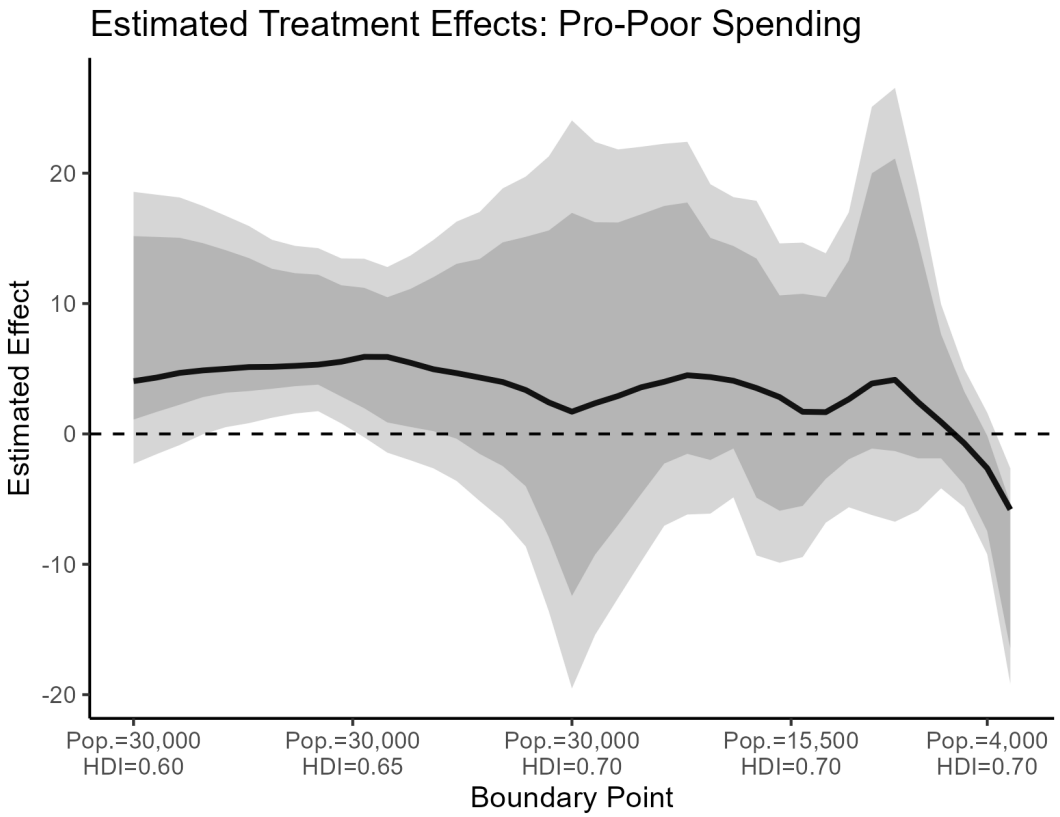


Figure 3: Treatment effects for pro-poor spending outcome using a local polynomial regression. The darker region represents pointwise bands and the lighter region represents uniform bands.

The patterns observed in pro-poor spending are intuitive: larger and less-developed municipalities tend to exhibit higher pro-poor expenditures than smaller and more-developed municipalities. Under the standard distance-based aggregation method, the positive and negative discontinuities offset one another, attenuating the aggregated statistic toward zero. In contrast, the aggregation method proposed in this paper avoids this cancellation issue and is therefore able to detect a discontinuity despite the heterogeneous pattern of effects along the boundary.

To check for robustness, a density test will be conducted to ensure that municipalities did not manipulate

their running variables to receive more funding. More specifically, a density test is conducted for four different subsets of the data, each of which were used in the main analysis. For the first subset, only municipalities in which the candidate was the incumbent mayor were retained. The second subset includes municipalities where the incumbent mayor was eligible to run for re-election. The third subset restricts the sample to municipalities in which the candidate was not the incumbent but the incumbent mayor was running for re-election. Finally, the fourth subset applies the same criteria as the third, further restricting the sample to candidates whose rank was less than three. These tests show no signs of manipulation.

Table 6: Density test results for different running variable subsets.

Variable	**Est.**	**Est. (Dist.)**
Candidate is Incumbent Mayor	-0.046	-0.045
	[0.369]	[0.591]
Incumbent Mayor Allowed for Re-election	-0.258	-0.303
	[0.428]	[0.177]
Not Incumbent Mayor, Running for Re-election	0.045	-0.062
	[0.330]	[0.327]
Not Incumbent Mayor, Running for Re-election, Rank less than 3	-0.034	-0.066
	[0.339]	[0.639]

Note: P-values are reported in square brackets.

6 Conclusion

This paper develops a testing procedure for detecting discontinuities along multidimensional regression discontinuity (RD) boundaries. A central advantage of the proposed approach is its strong performance in relatively small samples, which is a setting where many existing multivariate RD methods struggle due

to the curse of dimensionality. The simulation evidence demonstrates that the procedure yields reliable inference even with limited data, and the empirical application illustrates how it can complement and strengthen insights obtained from multivariate RD estimators.

Several promising directions for future research remain. One important extension concerns the choice of distance metric used to aggregate local estimates along the boundary. Because the power of the global test depends on how distances are scaled, developing principled data-driven methods for tuning or learning the metric may lead to substantial power gains. Another avenue involves identifying where along the boundary discontinuities occur. Combining the proposed global test with localization techniques could provide a more refined understanding of heterogeneity in treatment effects or sorting, while limiting the disadvantages of standard multivariate RD estimators.

References

[1] Marco Becht, Andrea Polo, and Stefano Rossi. "Does mandatory shareholder voting prevent bad acquisitions?" In: *The Review of financial studies* 29.11 (2016), pp. 3035–3067.

[2] Sebastian Calonico, Matias D Cattaneo, and Max H Farrell. "On the effect of bias estimation on coverage accuracy in nonparametric inference". In: *Journal of the American Statistical Association* 113.522 (2018), pp. 767–779.

[3] Sebastian Calonico, Matias D Cattaneo, and Rocio Titiunik. "Robust nonparametric confidence intervals for regression-discontinuity designs". In: *Econometrica* 82.6 (2014), pp. 2295–2326.

[4] Matias D Cattaneo, Michael Jansson, and Xinwei Ma. "Local regression distribution estimators". In: *Journal of econometrics* 240.2 (2024), p. 105074.

[5] Matias D Cattaneo, Michael Jansson, and Xinwei Ma. "Simple local polynomial density estimators". In: *Journal of the American Statistical Association* 115.531 (2020), pp. 1449–1455.

[6] Matias D Cattaneo, Rocio Titiunik, and Ruiqi Rae Yu. "Estimation and Inference in Boundary Discontinuity Designs: Distance-Based Methods". In: *arXiv preprint arXiv:2510.26051* (2025).

[7] Matias D Cattaneo, Rocio Titiunik, and Ruiqi Rae Yu. "Estimation and Inference in Boundary Discontinuity Designs: Location-Based Methods". In: *arXiv preprint arXiv:2505.05670* (2025).

[8] Alden Cheng. "Estimation of Regression Discontinuity and Kink Designs with Multiple Running Variables". In: *Available at SSRN 4516922* (2023).

[9] Victor Chernozhukov et al. *Double/debiased machine learning for treatment and structural parameters*. 2018.

[10] Harold D Chiang, Yu-Chin Hsu, and Yuya Sasaki. "Robust uniform inference for quantile treatment effects in regression discontinuity designs". In: *Journal of Econometrics* 211.2 (2019), pp. 589–618.

[11] Sarah R Cohodes and Joshua S Goodman. "Merit aid, college quality, and college completion: Massachusetts' Adams scholarship as an in-kind subsidy". In: *American Economic Journal: Applied Economics* 6.4 (2014), pp. 251–285.

[12] Federico Crippa. "Manipulation Test for Multidimensional RDD". In: *Journal of Applied Econometrics* (2025).

[13] Peter H Egger and Georg Wamser. "The impact of controlled foreign company legislation on real investments abroad. A multi-dimensional regression discontinuity design". In: *Journal of Public Economics* 129 (2015), pp. 77–91.

[14] Gregory Elacqua et al. "Short-run effects of accountability pressures on teacher policies and practices in the voucher system in Santiago, Chile". In: *School Effectiveness and School Improvement* 27.3 (2016), pp. 385–405.

[15] Brent J Evans. "SMART money: Do financial incentives encourage college students to study science?" In: *Education Finance and Policy* 12.3 (2017), pp. 342–368.

[16] Anderson Frey. "Cash transfers, clientelism, and political enfranchisement: Evidence from Brazil". In: *Journal of Public Economics* 176 (2019), pp. 1–17.

[17] Rina Friedberg et al. "Local linear forests". In: *Journal of Computational and Graphical Statistics* 30.2 (2020), pp. 503–517.

[18] Florian Gunsilius and David Van Dijcke. "Free Discontinuity Design: With an Application to the Economic Effects of Internet Shutdowns". In: *arXiv preprint arXiv:2309.14630* (2023).

[19] Martin L Hazelton and Jonathan C Marshall. "Linear boundary kernels for bivariate density estimation". In: *Statistics & probability letters* 79.8 (2009), pp. 999–1003.

[20] Guido Imbens and Stefan Wager. "Optimized regression discontinuity designs". In: *Review of Economics and Statistics* 101.2 (2019), pp. 264–278.

[21] Luke J Keele and Rocio Titiunik. "Geographic boundaries as regression discontinuities". In: *Political Analysis* 23.1 (2015), pp. 127–155.

[22] Juliana Londoño-Vélez, Catherine Rodríguez, and Fabio Sánchez. "Upstream and downstream impacts of college merit-based financial aid for low-income students: Ser Pilo Paga in Colombia". In: *American Economic Journal: Economic Policy* 12.2 (2020), pp. 193–227.

[23] Ricardo Masini and Marcelo Medeiros. "Balancing Flexibility and Interpretability: A Conditional Linear Model Estimation via Random Forest". In: *arXiv preprint arXiv:2502.13438* (2025).

[24] Jordan D Matsudaira. "Mandatory summer school and student achievement". In: *Journal of Econometrics* 142.2 (2008), pp. 829–850.

[25] Justin McCrary. "Manipulation of the running variable in the regression discontinuity design: A density test". In: *Journal of econometrics* 142.2 (2008), pp. 698–714.

[26] Yusuke Narita and Kohei Yata. "Algorithm as experiment: machine learning, market design, and policy eligibility rules". In: *arXiv preprint arXiv:2104.12909* (2021).

[27] John P Papay, John B Willett, and Richard J Murnane. "Extending the regression-discontinuity approach to multiple assignment variables". In: *Journal of Econometrics* 161.2 (2011), pp. 203–207.

[28] Sean F Reardon and Joseph P Robinson. "Regression discontinuity designs with multiple rating-score variables". In: *Journal of research on Educational Effectiveness* 5.1 (2012), pp. 83–104.

[29] Hongwei Wen and Hanyuan Hang. "Random forest density estimation". In: *International Conference on Machine Learning*. PMLR. 2022, pp. 23701–23722.

[30] Vivian C Wong, Peter M Steiner, and Thomas D Cook. "Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods". In: *Journal of Educational and Behavioral Statistics* 38.2 (2013), pp. 107–141.

7 Appendix

Here, proofs and materials for the main paper are included.

7.1 Consistency of Random Forest Estimators

7.1.1 Proof of Lemma 1

Proof. Let x_0 be an evaluation point and, without loss of generality, consider estimation from the region Ω. The standard random forest estimator can be written as follows:

$$\mu_n^{(rf)}(x_0)^+ = \sum_{i=1}^{n} \alpha_i^+(x_0) Y_i$$

where $\alpha_i^+(x_0) = 0$ if $X_i \notin \Omega$. Let $\boldsymbol{X}$ be an $n \times d+1$ matrix with $\begin{pmatrix} 1 & (X_i - x_0)' \end{pmatrix}$ in each row, let $A^+ \in \mathbb{R}^{n \times n}$ be a diagonal matrix with $\alpha_i^+(x_0)$ on the diagonal, and let Y be a vector of size n consisting of each Y_i. Also, let $J = \text{diag}(0, 1, 1, 1, ...)$ and $\lambda_n \to 0$ (as $n \to \infty$) be the ridge regularization penalty. With that, the local linear forest estimator can be written as follows:

$$\begin{aligned}
\mu_n^{(llf)}(x_0) &= e_1'(\boldsymbol{X}' A^+ \boldsymbol{X} + \lambda_n J)^{-1} \boldsymbol{X}' A^+ Y \\
&= e_1'(\boldsymbol{X}' A^+ \boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0) Y_i
\end{aligned}$$

With that, this estimator can be simplified as follows:

$$\begin{aligned}
\mu_n^{(llf)}(x_0) &= e_1'(\boldsymbol{X}' A^+ \boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0) Y_i \\
&= e_1'(\boldsymbol{X}' A^+ \boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0) (\mu(X_i) + \epsilon_i)
\end{aligned}$$

If $\mathbb{1}$ is a vector of all ones, then a Taylor expansion of $\mu(\cdot)$ gives the following:

$$\begin{aligned}
\mu_n^{(llf)}(x_0)^+ &= e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0)\epsilon_i \\
&+ e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0)\mu(x_0)^+ \\
&+ e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0)\nabla_x\mu(x_0)'^+(X_i - x_0) \\
&+ O(\bar{R}^2)
\end{aligned}$$

where $\bar{R}^2$ is the average squared radius of the leaves in each tree grown by the random forest. Now, since $\boldsymbol{X}e_1 = \mathbb{1}$ and $\lambda_n J e_l = 0$, notice the following:

$$\begin{aligned}
(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)e_1 &= \boldsymbol{X}'A^+\mathbb{1} \\
e_1 &= (\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1}\boldsymbol{X}'A^+\mathbb{1} \\
e_1'e_1 &= e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1}\boldsymbol{X}'A^+\mathbb{1} \\
&= e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0) \\
&= 1
\end{aligned}$$

With that, the estimator can be simplified as follows:

$$\begin{aligned}
\mu_n^{(llf)}(x_0)^+ &= \mu(x_0)^+ \\
&\quad + e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^n \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0)\epsilon_i \\
&\quad + e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^n \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0)\nabla_x\mu(x_0)'(X_i - x_0) \\
&\quad + O(\bar{R}^2)
\end{aligned}$$

From here, since $\lambda_n \to 0$, notice the following about the third term.

$$\begin{aligned}
&e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^n \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0)\nabla_x\mu(x_0)'(X_i - x_0) \\
&= e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^n \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0) \begin{pmatrix} 0 \\ \\ \nabla_x\mu(x_0) \end{pmatrix}' \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \\
&= e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^n \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0) \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix}' \begin{pmatrix} 0 \\ \\ \nabla_x\mu(x_0) \end{pmatrix} \\
&= e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1}\boldsymbol{X}'A^+\boldsymbol{X} \begin{pmatrix} 0 \\ \\ \nabla_x\mu(x_0) \end{pmatrix} \\
&\to 0
\end{aligned}$$

With that, the estimator can be further simplified as follows.

$$\begin{aligned}
\mu_n^{(llf)}(x_0)^+ &= \mu(x_0)^+ \\
&\quad + e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^n \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0)\epsilon_i \\
&\quad + O(\bar{R}^2)
\end{aligned}$$

Since ϵ_i is independent of $\alpha_i(x_0)$ and X_i, the expectation of the second term above is equal to zero. Therefore, we have:

$$e_1'(\boldsymbol{X}'A^+\boldsymbol{X} + \lambda_n J)^{-1} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \\ X_i - x_0 \end{pmatrix} \alpha_i^+(x_0)\epsilon_i \xrightarrow{p} 0$$

Now, note that the radius of each leaf vanishes according to Lemma 1 of Masini and Medeiros (2025). Therefore, the local linear forest estimator is consistent at any boundary point x_0. □

7.1.2 Proof of Lemma 2

Proof. The random forest density estimator is an approach that approximates the density through an averaging of randomly partitioned trees. This approach is similar to a histogram estimator, except where the partitions in the random forest density estimator are made using a more efficient algorithm.

Let T be the number of trees and let d be the number of running variables. Also, without loss of generality, let n denote the number of points in Ω and let N denote the total number of points. With that, let p_n be the depths of the splits.[5] The random forest density estimator comes from averaging over many random tree density estimators. To determine the splits used for a given tree, the first step is enclosing the support of X_i into a hyper rectangle $B_r = [-r, r]^d$. Then, a coordinate $l \in \{1, 2, ..., d\}$ is randomly chosen along which the data is split in half. This process repeats for each individual partition of the data until depth p_n is met.

Let $A_p^{j,t}$ denote leaf j of tree t with depth p and let I_p^t denote the index of of each partition for tree t and depth p. Also, let $\mu(A_p^{j,t})$ be a non-zero Lebesgue measure of each leaf and let $D(A_p^{j,t})$ be defined as follows:

$$D(A_p^{j,t}) = \frac{1}{n} \sum_{i \in \Omega} \mathbb{1}(X_i \in A_p^{j,t})$$

[5] In this case, set $p_n = p_N = d(1 + d\ln(2))^{-1} \ln(N)$, which appears in Theorem 2 of Wen and Hang (2022).

With that, the random tree density estimator is defined as follows:

$$f^p_{D,t}(x_0|x \in \Omega) = \sum_{j \in I^t_p} \frac{D(A^{j,t}_p)\mathbb{1}(x_0 \in A^{j,t}_p(x_0))}{\mu(A^{j,t}_p)}$$

The random forest density estimator is then defined as follows:

$$f_{D,E}(x_0|x \in \Omega) = \frac{1}{T}\sum_{t=1}^{T} f^p_{D,t}(x_0)$$

If $D(\cdot)$ is the empirical probability measure, let $P(\cdot)$ denote the corresponding population probability measure. With that, notice the following:

$$\begin{aligned}
\mathbb{E}[(f_{D,E}(x_0|x \in \Omega) - f(x_0|x \in \Omega))^2] &= \mathbb{E}[(f_{D,E}(x_0|x \in \Omega) - f_{P,E}(x_0|x \in \Omega))^2] \\
&\quad + \mathbb{E}[(f_{P,E}(x_0|x \in \Omega) - f(x_0|x \in \Omega))^2] \\
&\quad + 2\mathbb{E}[(f_{D,E}(x_0|x \in \Omega) - f(x_0|x \in \Omega)) \\
&\quad \cdot (f_{P,E}(x_0|x \in \Omega) - f(x_0|x \in \Omega))] \\
&\quad \leq \mathbb{E}[(f_{D,E}(x_0|x \in \Omega) - f_{P,E}(x_0|x \in \Omega))^2] \\
&\quad + \mathbb{E}[(f_{P,E}(x_0|x \in \Omega) - f(x_0|x \in \Omega))^2] \\
&\quad + 2\sqrt{\mathbb{E}[(f_{D,E}(x_0|x \in \Omega) - f_{P,E}(x_0|x \in \Omega))^2]} \\
&\quad \cdot \sqrt{\mathbb{E}[(f_{P,E}(x_0|x \in \Omega) - f(x_0|x \in \Omega))^2]}
\end{aligned}$$

Therefore, if $\mathbb{E}[(f_{D,E}(x_0|x \in \Omega) - f_{P,E}(x_0|x \in \Omega))^2]$ and $\mathbb{E}[(f_{P,E}(x_0|x \in \Omega) - f(x_0|x \in \Omega))^2]$ converge to zero with the sample size, then the random forest density estimator is consistent for the true density. From the results of Wen and Hang (2022) for $f \in C^{1,\alpha}$:

$$\mathbb{E}[(f_{P,E}(x_0|x \in \Omega) - f(x_0|x \in \Omega))^2] \leq \frac{c^2(2r)^4 d^2}{T}\exp\left(\frac{-0.75p_n}{d}\right) + 4c^2(2r)^{2d+2}d^2\exp\left(\frac{-p_n}{d}\right)$$

for some constant c.[6] Since $p_n \to \infty$ as $n \to \infty$, $\mathbb{E}[(f_{P,E}(x_0|x \in \Omega) - f(x_0|x \in \Omega))^2] \to 0$. For the other

[6]Even with an irregular support, these bounds still work because the diameter of each cell is bounded above by the diameter given in the original Wen and Hang (2022) paper.

term:

$$\begin{aligned}\mathbb{E}[(f_{D,E}(x_0|x \in \Omega) - f_{P,E}(x_0|x \in \Omega))^2] &\leq \mathbb{E}\left[\frac{P(A_p^j(x_0))}{n\mu^2(A_p^j(x_0))}\right] \\ &\leq \mathbb{E}\left[\frac{c_f}{n\mu(A_p^j(x_0))}\right]\end{aligned}$$

for some constant c_f. Note that when $\mu(A_p^j(x_0)) = 0$, $\mathbb{E}[(f_{D,E}(x_0|x \in \Omega) - f_{P,E}(x_0|x \in \Omega))^2] = 0$. Let $V_j \in [v_{\min}, 1]$ be the fraction of each cube that is within the support of X. By construction, notice the following:

$$\begin{aligned}n\mu(A_p^j(x_0)) &= \frac{n(2r)^d}{2^p}V_j \\ \ln(n\mu(A_p^j(x_0)) &= d\ln(2r) + \ln(n) - p\ln(2) + \ln(V_j) \\ &= d\ln(2r) + \ln(n) - \frac{d\ln(n)\ln(2)}{1 + d\ln(2)} + \ln(V_j) \\ &= d\ln(2r) + \ln(n)\left(1 - \frac{d\ln(2)}{1 + d\ln(2)}\right) + \ln(V_j) \\ &\to \infty\end{aligned}$$

In other words, $n\mu(A_p^j(x_0)) \to \infty$, which means that $\mathbb{E}[(f_{D,E}(x_0|x \in \Omega) - f_{P,E}(x_0|x \in \Omega))^2] \to 0$, thus showing that $f_{D,E}(x_0|x \in \Omega)$ is consistent for $f(x_0|x \in \Omega)$. Finally, to get the full density from the region Ω, $f(x_0|x \in \Omega)$ must be multiplied by $\frac{n}{N}$. □

7.2 Proofs of Theorems

7.2.1 Proof of Theorem 1

Proof. First, define $Y^{(d)} = Y(1) - Y(0)$. By Assumption 7, since g_i is a deterministic function of X_i, $\mathbb{E}[Y_i(t)|g_i = g]$ is continuous for all g and for all $t \in \{0, 1\}$. With that, notice the following:

$$\begin{aligned}
\lim_{\epsilon \to 0^+} \mathbb{E}[Y_i \mid g_i = \epsilon] - \lim_{\epsilon \to 0^-} \mathbb{E}[Y_i \mid g_i = \epsilon] &= \lim_{\epsilon \to 0^+} \mathbb{E}[Y_i(1) \mid g_i = \epsilon] - \lim_{\epsilon \to 0^-} \mathbb{E}[Y_i(0) \mid g_i = \epsilon] \\
&= \mathbb{E}[Y_i(1) - Y_i(0)|g_i = 0] \\
&= \int_y Y^{(d)} f_{Y^{(d)}|g}(Y^{(d)}|g = 0) dY^{(d)} \\
&= \frac{1}{f_{g_\Omega}(0)} \int_y Y^{(d)} f_{Y^{(d)},g}(Y^{(d)}, g = 0) dY^{(d)} \\
&= \frac{1}{f_{g_\Omega}(0)} \int_y \int_{X: g_\Omega(X)=0} Y^{(d)} f_{Y^{(d)},X}(Y, X) dS_\Omega dY^{(d)} \\
&= \frac{1}{f_{g_\Omega}(0)} \int_{X: g_\Omega(X)=0} \left(\int_y Y^{(d)} f_{Y^{(d)}|X}(Y^{(d)}|X) dY^{(d)} \right) f_X(x) dS_\Omega \\
&= \frac{1}{f_{g_\Omega}(0)} \int_{x: g_\Omega(x)=0} \tau(x) f_X(x)\, dS_\Omega
\end{aligned}$$

□

7.2.2 Proof of Theorem 2

Proof. First, note that the following holds by the definition of conditional density:

$$\begin{aligned}
\frac{\mathbb{P}(\Gamma = 1|g = 0)}{f_{g,\Gamma}(g = 0, \Gamma = 1)} &= \frac{1}{f_{g_\Omega}(0)} \\
&= \frac{\mathbb{P}(\Gamma = 0|g = 0)}{f_{g,\Gamma}(g = 0, \Gamma = 0)}
\end{aligned}$$

With that, notice the following:

$$\left(\lim_{\epsilon\to 0^+}\mathbb{E}[Y_i \mid g_i=\epsilon,\Gamma_i=1]-\lim_{\epsilon\to 0^-}\mathbb{E}[Y_i \mid g_i=\epsilon,\Gamma_i=1]\right)\mathbb{E}[\Gamma_i|g_i=0]-$$

$$\left(\lim_{\epsilon\to 0^+}\mathbb{E}[Y_i \mid g_i=\epsilon,\Gamma_i=0]-\lim_{\epsilon\to 0^-}\mathbb{E}[Y_i \mid g_i=\epsilon,\Gamma_i=0]\right)(1-\mathbb{E}[\Gamma_i|g_i=0])$$

$$=\mathbb{E}[Y_i(1)-Y_i(0) \mid g_i=0,\Gamma_i=1]\mathbb{E}[\Gamma_i|g_i=0]-\mathbb{E}[Y_i(1)-Y_i(0) \mid g_i=0,\Gamma_i=0](1-\mathbb{E}[\Gamma_i|g_i=0])$$

$$=\mathbb{P}(\Gamma_i=1|g_i=0)\int_y Y^{(d)} f_{Y^{(d)}|g,\Gamma}(Y^{(d)}|g=0,\Gamma=1)dY^{(d)}$$

$$-\mathbb{P}(\Gamma_i=0|g_i=0)\int_y Y^{(d)} f_{Y^{(d)}|g,\Gamma}(Y^{(d)}|g=0,\Gamma=0)dY^{(d)}$$

$$=\frac{\mathbb{P}(\Gamma_i=1|g_i=0)}{f_{g,\Gamma}(g_i=0,\Gamma=1)}\int_y Y^{(d)} f_{Y^{(d)},g,\Gamma}(Y^{(d)},g=0,\Gamma=1)dY^{(d)}$$

$$-\frac{\mathbb{P}(\Gamma_i=0|g_i=0)}{f_{g,\Gamma}(g_i=0,\Gamma=0)}\int_y Y^{(d)} f_{Y^{(d)},g,\Gamma}(Y^{(d)},g=0,\Gamma=0)dY^{(d)}$$

$$=\frac{1}{f_{g_\Omega}(0)}\left(\int_y Y^{(d)} f_{Y^{(d)},g,\Gamma}(Y^{(d)},g=0,\Gamma=1)dY^{(d)}-\int_y Y^{(d)} f_{Y^{(d)},g,\Gamma}(Y^{(d)},g=0,\Gamma=0)dY^{(d)}\right)$$

$$=\frac{1}{f_{g_\Omega}(0)}\left(\int_y Y^{(d)}\int_{X:g_\Omega(X)=0,\Gamma(X)=1} f_{Y^{(d)},X}(Y^{(d)},X)dS_\Omega dY^{(d)}\right.$$

$$\left.-\int_y Y^{(d)}\int_{X:g_\Omega(X)=0,\Gamma(X)=0} f_{Y^{(d)},X}(Y^{(d)},X)dS_\Omega dY^{(d)}\right)$$

$$=\frac{1}{f_{g_\Omega}(0)}\left(\int_y\int_{X:g_\Omega(X)=0,\Gamma(X)=1} Y^{(d)} f_{Y^{(d)}|X}(Y^{(d)}|X)f_X(x)dS_\Omega dY^{(d)}\right.$$

$$\left.-\int_y\int_{X:g_\Omega(X)=0,\Gamma(X)=0} Y^{(d)} f_{Y^{(d)}|X}(Y^{(d)}|X)f_X(x)dS_\Omega dY^{(d)}\right)$$

$$=\frac{1}{f_{g_\Omega}(0)}\left(\int_{X:g_\Omega(X)=0,\Gamma(X)=1}\tau(x)f_X(x)dS_\Omega-\int_{X:g_\Omega(X)=0,\Gamma(X)=0}\tau(x)f_X(x)dS_\Omega\right)$$

$$=\frac{1}{f_{g_\Omega}(0)}\int_{x:g_\Omega(x)=0}|\tau(x)|f_X(x)\,dS_\Omega=\frac{1}{f_{g_\Omega}(0)}\int_{x:g_\Omega(x)=0}(2\Gamma(X)-1)\tau(x)f_X(x)\,dS_\Omega$$

□

7.2.3 Proof of Theorem 3

Proof. Notice the following:

$$\begin{aligned}
&\left[\lim_{\epsilon\to 0^+} f_{g_\Omega|\Lambda}(g=\epsilon|\Lambda_i=1) - \lim_{\epsilon\to 0^-} f_{g_\Omega|\Lambda}(g=\epsilon|\Lambda_i=1)\right]\mathbb{P}(\Lambda_i=1) - \\
&\left[\lim_{\epsilon\to 0^+} f_{g_\Omega|\Lambda}(g=\epsilon|\Lambda_i=0) - \lim_{\epsilon\to 0^-} f_{g_\Omega|\Lambda}(g=\epsilon|\Lambda_i=0)\right]\mathbb{P}(\Lambda_i=0) \\
=&\left[\lim_{\epsilon\to 0^+} f_{g_\Omega,\Lambda}(g=\epsilon,\Lambda_i=1) - \lim_{\epsilon\to 0^-} f_{g_\Omega,\Lambda}(g=\epsilon,\Lambda_i=1)\right] - \\
&\left[\lim_{\epsilon\to 0^+} f_{g_\Omega,\Lambda}(g=\epsilon,\Lambda_i=0) - \lim_{\epsilon\to 0^-} f_{g_\Omega,\Lambda}(g=\epsilon,\Lambda_i=0)\right] \\
=&\int_{X:g_\Omega(X)=0,\Lambda(X)=1} \tau_f(x)dS_\Omega - \int_{X:g_\Omega(X)=0,\Lambda(X)=0} \tau_f(x)dS_\Omega \\
=&\int_{X:g_\Omega(X)=0} |\tau_f(x)|dS_\Omega = \int_{X:g_\Omega(X)=0} (2\Lambda(X)-1)\tau_f(x)dS_\Omega
\end{aligned}$$

□

7.2.4 Proof of Theorem 4

Proof. The estimator for each fold is defined as follows:

$$\hat{\tau}_{\Gamma_{k,p}} = e_1'(\hat{\beta}^+_{k,p} - \hat{\beta}^-_{k,p})$$

To show that the final estimator is consistent and asymptotically normal, it is sufficient to show that each fold-level estimator is consistent and asymptotically normal. Without loss of generality, notice the following about $\hat{\beta}^+_{k,p}$.

$$\begin{aligned}
\hat{\beta}^+_{k,p} &= (G_p' W_{+,k} G_p)^{-1} G_p' W_{+,k} \hat{Y} \\
&= (G_p' W_{+,k} G_p)^{-1} G_p' W_{+,k} Y^* + (G_p' W_{+,k} G_p)^{-1} G_p' W_{+,k} \eta
\end{aligned}$$

where

$$G_p = \begin{pmatrix} r_p(g_1)' \\ r_p(g_2)' \\ \vdots \end{pmatrix}, \quad W_{+,k} = \mathbb{1}(g_i \geq 0)\mathbb{1}(i \in I_k)\text{diag}\left(K\left(\frac{g_1}{h_k}\right), K\left(\frac{g_2}{h_k}\right), ...\right), \quad \eta = \begin{pmatrix} \hat{\eta}_1 \\ \hat{\eta}_2 \\ \vdots \end{pmatrix}, \quad Y^* = \begin{pmatrix} Y_1^* \\ Y_2^* \\ \vdots \end{pmatrix}$$

Notice the following for the first term (according to the results of Calonico, Cattaneo, and Titiunik (2014)):

$$(G_p' W_{+,k} G_p)^{-1} G_p' W_{+,k} Y^* \xrightarrow{p} \beta^+$$

$$\sqrt{n_k h_k}((G_q' W_{+,k} G_q)^{-1} G_q' W_{+,k} Y^* - \beta^+) \xrightarrow{d} \mathcal{N}(0, V_+)$$

where V_+ is the asymptotic variance of the local polynomial regression coefficient estimator and

$$\epsilon = \begin{pmatrix} \epsilon_1 & \epsilon_2, & \dots \end{pmatrix}', \quad n_k = \sum_{i=1}^{n} \mathbb{1}(i \in I_k)$$

From here, it is sufficient to show that the second term in $\hat{\beta}_{k,p}^+$ is asymptotically negligible for any polynomial order $p \in \{1, 2, ...\}$. With that, by the law of large numbers, $(G_p' W_{+,k} G_p)^{-1} G_p' W_{+,k} \eta = (\Gamma_p^+)^{-1} \frac{1}{n_k h_k} G_p' W_{+,k} \eta + o_p(n_k^{-1} h_k^{-1})$. Now, notice the following:[7]

$$\begin{aligned}
\mathbb{E}[G_p' W_{+,k} \eta] &= \mathbb{E}[G_p' W_{+,k} \mathbb{E}[\eta | X]] \\
&= \frac{2}{h_k} \int_{X: g(X) \geq 0} K(g(X)/h) r_p(g(X)/h) \mathbb{E}[\Gamma_n(X) - \Gamma(X) | X] \mathbb{E}[Y(1) | X] f_X(x) dX \\
&= \int_0^1 K(u) r_p(u) \left(\int_{X: g(X)/h = u} m_1(X) dX \right) du \\
&= \int_0^1 K(u) r_p(u) S_1(hu) du
\end{aligned}$$

where

$$m_1(X) = 2\mathbb{E}[\Gamma_n(X) - \Gamma(X) | X] \mathbb{E}[Y(1) | X] f_X(x), \quad S_1(z) = \int_{X: g(X) = z} m_1(X) dX$$

Using a Taylor expansion of $S_1(hu)$:

$$\mathbb{E}[G_p' W_{+,k} \eta] = S_1(0) \int_0^1 K(u) r_p(u) du + h_k S_1'(0) \int_0^1 u K(u) r_p(u) du + \frac{h_k^2}{2} S_1''(0) \int_0^1 u^2 K(u) r_p(u) du + O(h_k^2)$$

This means:

$$\mathbb{E}[G_p' W_{-,k} \eta] = S_0(0) \int_{-1}^0 K(u) r_p(u) du + h_k S_0'(0) \int_{-1}^0 u K(u) r_p(u) du + \frac{h_k^2 S_0''(0)}{2} \int_{-1}^0 u^2 K(u) r_p(u) du + O(h_k^2)$$

[7] $\mathbb{E}[\eta | X]$ can be separated in the second line since sample splitting is used.

From here, consider the following:

$$e_1'(\Gamma_p^+)^{-1}\int_0^1 u^j r_p(u)K(u)du$$
$$e_1'(\Gamma_p^-)^{-1}\int_{-1}^0 u^j r_p(u)K(u)du$$

Notice that:

$$\int_0^1 u^j r_p(u)K(u)du = \Gamma_p^+ e_{j+1}$$
$$\int_{-1}^0 u^j r_p(u)K(u)du = \Gamma_p^- e_{j+1}$$

Therefore, for $j \leq p$:

$$e_1'(\Gamma_p^+)^{-1}\int_0^1 u^j r_p(u)K(u)du = e_1'(\Gamma_p^+)^{-1}\Gamma_p^+ e_{j+1}$$
$$e_1'(\Gamma_p^-)^{-1}\int_{-1}^0 u^j r_p(u)K(u)du = e_1'(\Gamma_p^-)^{-1}\Gamma_p^- e_{j+1}$$

This means that the terms above equal 1 if $j = 0$ and 0 if $0 < j \leq p$. This implies the following must hold:

$$e_1'(\Gamma_p^+)^{-1}\mathbb{E}[G_p' W_{+,k}\eta] - e_1'(\Gamma_p^-)^{-1}\mathbb{E}[G_p' W_{-,k}\eta] = S_1(0) - S_0(0) + O(h_k^2)$$

From here, the following holds:

$$S_1(0) - S_0(0) = \int_{X:g(X)=0} 2\mathbb{E}[Y(1) - Y(0)|X]\mathbb{E}[\Gamma_n(X) - \Gamma(X)|X]f_X(x)dX$$

When $\mathbb{E}[Y(1) - Y(0)|X] \neq 0$, $\mathbb{E}[\Gamma_n(X) - \Gamma(X)|X] \xrightarrow{p} 0$. More specifically, notice the following:

$$\mathbb{E}[\Gamma_n(X) - \Gamma(X)|X] = \mathbb{P}(\Gamma_n(X) - \Gamma(X) = 1|X) - \mathbb{P}(\Gamma_n(X) - \Gamma(X) = -1|X)$$

Without loss of generality, suppose that $\tau(X) > 0$. Then the following must hold:

$$\begin{aligned}
\mathbb{P}(\Gamma_n(X) - \Gamma(X) = 1|X) - \mathbb{P}(\Gamma_n(X) - \Gamma(X) = -1|X) &= \mathbb{P}(\Gamma_n(X) = 2|X) - \mathbb{P}(\Gamma_n(X) = 0|X) \\
&= -\mathbb{P}(\Gamma_n(X) = 0|X) \\
&= -\mathbb{P}(\tau_n(X) < 0|X) \\
&= -\mathbb{P}(\tau_n(X) - \tau(X) < -\tau(X)|X) \\
&\geq -\mathbb{P}(\tau_n(X) - \tau(X) < -\tau(X)|X) \\
&\quad - \mathbb{P}(\tau_n(X) - \tau(X) > \tau(X)|X) \\
&= -\mathbb{P}(|\tau_n(X) - \tau(X)| > \tau(X)|X) \\
&\geq -\frac{\mathbb{E}[q(|\tau_n(X) - \tau(X)|)|X]}{q(\tau(X))}
\end{aligned}$$

where the last line holds by the Markov Inequality for some non-decreasing $q(\cdot) > 0$. By assumption 3, 7, and 8, $\mathbb{E}[q(|\tau_n(X) - \tau(X)|)|X]$ exists for some measurable function $q : \mathbb{R}_+ \to \mathbb{R}_+$. Let $a_h \in (0, 1)$. For any bandwidth $h = O(n^{-a_h})$ and $|\tau_n(X) - \tau(X)| = o_p(1)$, there exists a function $q(\cdot)$ such that $\sqrt{n_k h_k}\mathbb{E}[q(|\tau_n(X) - \tau(X)|)|X] \to 0$. Therefore, $\mathbb{E}[\Gamma_n(X) - \Gamma(X)|X]$ is asymptotically negligible for all $x \in \partial\Omega$, regardless of a_h.

When $\mathbb{E}[Y(1) - Y(0)|X] = 0$, the integrand above equals zero since $\mathbb{E}[\Gamma_n(X) - \Gamma(X)|X]$ is bounded. This means that $S_1(0) - S_0(0) = 0$. Therefore, we can conclude that:

$$\sqrt{n_k h_k} e_1'((G_p' W_{+,k} G_p')^{-1}(G_p' W_{+,k}\eta) - (G_p' W_{-,k} G_p')^{-1}(G_p' W_{-,k}\eta)) \xrightarrow{p} 0$$

This proves that our estimator is consistent and asymptotically normal. □

7.2.5 Proof of Theorem 5

Proof. To establish consistency and asymptotic normality, it suffices to verify these properties for each individual $\hat{\tau}_{\Lambda_k}$. Note first that the base kernels $K(\cdot)$ and $L(\cdot)$ are symmetric about zero, implying that the

quantities $\hat{g}_{i,k}$ entering the kernels may be treated as if they equal $g_{i,k}$ or $g^*_{i,k}$. Moreover, $\hat{g}^j_{i,k} = (g^*_{i,k})^j$ for any even integer j. With that, both the main kernel estimator and the corresponding bias-correction term can be shown to be consistent.

For the main kernel estimator, notice the following:

$$\begin{aligned}
\hat{f}_{g^*}(0^+) - \hat{f}_{g^*}(0^-) &= \frac{1}{n_k}\sum_{i\in I_k}\left(\alpha_1 + \alpha_2\left|\frac{\hat{g}_{i,k}}{h_k}\right|\right)K\left(\frac{\hat{g}_{i,k}}{h_k}\right)\frac{1}{h_k}(2\mathbb{1}(\hat{g}_{i,k}>0)-1) \\
&= \frac{1}{n_k}\sum_{i\in I_k}\left(\alpha_1 + \alpha_2\frac{|g_{i,k}|}{h_k}\right)K\left(\frac{g_{i,k}}{h_k}\right)\frac{1}{h_k}(2\mathbb{1}(\hat{g}_{i,k}>0)-1) \\
\mathbb{E}[\hat{f}_{g^*}(0^+) - \hat{f}_{g^*}(0^-)] &= \mathbb{E}\left[\left(\alpha_1 + \alpha_2\frac{|g_{i,k}|}{h_k}\right)K\left(\frac{g_{i,k}}{h_k}\right)\frac{1}{h_k}(2\mathbb{1}(\hat{g}_{i,k}>0)-1)\right] \\
&= \mathbb{E}\left[\left(\alpha_1 + \alpha_2\frac{|g_{i,k}|}{h_k}\right)K\left(\frac{g_{i,k}}{h_k}\right)\frac{1}{h_k}\mathbb{E}[(2\mathbb{1}(\hat{g}_{i,k}>0)-1)|X_i]\right] \\
&= \frac{1}{h_k}\int_{\text{Supp}(X)}\left(\alpha_1 + \alpha_2\frac{|g(X)|}{h_k}\right)K\left(\frac{g(X)}{h_k}\right)\mathbb{E}[(2\mathbb{1}(\hat{g}(X)>0)-1)|X]f_X(x)dX \\
&= \int_0^1(\alpha_1+\alpha_2 u)K(u)\left(\int_{X:g(X)=h_k u}\mathbb{E}[(2\mathbb{1}(\hat{g}(X)>0)-1)|X]f_X(x)dX\right)du \\
&+ \int_{-1}^0(\alpha_1-\alpha_2 u)K(u)\left(\int_{X:g(X)=h_k u}\mathbb{E}[(2\mathbb{1}(\hat{g}(X)>0)-1)|X]f_X(x)dX\right)du \\
&= \int_0^1(\alpha_1+\alpha_2 u)K(u)S_1(h_k u)du + \int_{-1}^0(\alpha_1-\alpha_2 u)K(u)S_0(h_k u)du
\end{aligned}$$

where

$$\begin{aligned}
S_1(z) &= \int_{X:g(X)=z}\mathbb{E}[(2\mathbb{1}(\Lambda_n(X)=1)-1)|X]f^+_X(x)dX \\
S_0(z) &= \int_{X:g(X)=z}\mathbb{E}[(2\mathbb{1}(\Lambda_n(X)=0)-1)|X]f^-_X(x)dX
\end{aligned}$$

which can also be rewritten as:

$$\begin{aligned}
S_1(z) &= \int_{X:g(X)=z}(2\mathbb{P}(\Lambda_n(X)=1|X)-1)f^+_X(x)dX \\
S_0(z) &= \int_{X:g(X)=z}(2\mathbb{P}(\Lambda_n(X)=0|X)-1)f^-_X(x)dX
\end{aligned}$$

Using Taylor expansions of $S_1(\cdot)$ and $S_0(\cdot)$:

$$\mathbb{E}[\hat{f}_{g^*}(0^+) - \hat{f}_{g^*}(0^-)] = (S_1(0) + S_0(0)) + \frac{h_k^2}{2}(S_1''(0) + S_0''(0)) \int_0^1 u^2(\alpha_1 + \alpha_2 u)K(u)du + O(h_k^2)$$

From here, notice the following:

$$S_1(0) + S_0(0) = \int_{X:g(X)=0} \Big[(2\mathbb{P}(\Lambda_n(X) = 1|X) - 1)f_X^+(x) + (2\mathbb{P}(\Lambda_n(X) = 0|X) - 1)f_X^-(x)\Big]dX$$

If $f_X^+(x) = f_X^-(x)$:

$$\begin{aligned} S_1(0) + S_0(0) &= \int_{X:g(X)=0} f_X(x) \left[(2\mathbb{P}(\Lambda_n(X) = 1|X) - 1) + (2\mathbb{P}(\Lambda_n(X) = 0|X) - 1)\right] \\ &= 0 \end{aligned}$$

When $f_X^+(x) \neq f_X^-(x)$, the integrand above is given as follows:

$$2\mathbb{P}(\tau_n(X) \geq 0|X)f_X^+(x) + 2\mathbb{P}(\tau_n(X) < 0|X)f_X^-(x) - (f_X^+(x) + f_X^-(x))$$

Suppose, without loss of generality, that $\tau_f(x) > 0$. Then the following must hold:

$$\begin{aligned} \mathbb{P}(\tau_n(X) < 0|X) &= \mathbb{P}(\tau_n(X) - \tau_f(x) < -\tau_f(x)|X) \\ &\leq \mathbb{P}(\tau_n(X) - \tau_f(x) < -\tau_f(x)|X) + \mathbb{P}(\tau_n(X) - \tau_f(x) > \tau_f(x)|X) \\ &= \mathbb{P}(|\tau_n(X) - \tau_f(x)| > \tau_f(x)|X) \\ &\leq \frac{\mathbb{E}[q(|\tau_n(X) - \tau_f(x)|)|X]}{q(\tau_f(x))} \end{aligned}$$

where the last step holds by the Markov Inequality for some function $q : \mathbb{R}_+ \to \mathbb{R}_+$ that is measurable and non-decreasing. Since $q(\cdot)$ is arbitrary, this probability converges to zero at an arbitrary rate that can be made faster than any fixed rate n_k^{-a} for fixed $a \in (0, 1)$. Therefore, in this case:

$$\begin{aligned} 2\mathbb{P}(\tau_n(X) \geq 0|X)f_X^+(x) + 2\mathbb{P}(\tau_n(X) < 0|X)f_X^-(x) - (f_X^+(x) + f_X^-(x)) &\to 2f_X^+(x) - f_X^+(x) - f_X^-(x) \\ &= f_X^+(x) - f_X^-(x) \end{aligned}$$

By an analogous argument, if $\tau_f(x) < 0$, then the integrand converges to $-(f_X^+(x) - f_X^-(x))$. With that, the following holds:

$$S_1(0) + S_0(0) \to \int_{x:g(x)=0} |\tau_f(x)| dX$$

Now, for the second term, notice the following:

$$S_1''(0) + S_0''(0) = \frac{d^2}{dz^2} \int_{x:g(x)=z} 2\mathbb{P}(\tau_n(X) \geq 0|X) f_X^+(x) + 2\mathbb{P}(\tau_n(X) < 0|X) f_X^-(x) - (f_X^+(x) + f_X^-(x)) dX \Big|_{z=0}$$

Using the same arguments from before, the following holds:

$$S_1''(0) + S_0''(0) \to f_{g^*}''(0^+) - f_{g^*}''(0^-)$$

To remove this curvature bias term, first notice the following:

$$\begin{aligned}
\hat{f}_{g^*}''(0^+) - \hat{f}_{g^*}''(0^-) &= \frac{1}{n_k h_k^3} \sum_{i \in I_k} L^{(2)}\left(\frac{\hat{g}_{i,k}}{h_k}\right) (2\mathbb{1}(\hat{g}_{i,k} > 0) - 1) \\
&= \frac{1}{n_k h_k^3} \sum_{i \in I_k} L^{(2)}\left(\frac{g_{i,k}}{h_k}\right) (2\mathbb{1}(\hat{g}_{i,k} > 0) - 1) \\
\mathbb{E}[\hat{f}_{g^*}''(0^+) - \hat{f}_{g^*}''(0^-)] &= \frac{1}{h_k^3} \mathbb{E}\left[L^{(2)}\left(\frac{g_{i,k}}{h_k}\right) \mathbb{E}[(2\mathbb{1}(\hat{g}_{i,k} > 0) - 1)|X_i]\right] \\
&= \frac{1}{h_k^3} \int_{\text{Supp}(X)} L^{(2)}\left(\frac{g_{i,k}}{h_k}\right) \mathbb{E}[(2\mathbb{1}(\hat{g}(X) > 0) - 1)|X] f_X(x) dX \\
&= \frac{1}{h_k^2} \int_{-1}^{1} L^{(2)}(u) \left(\int_{X:g(X)=h_k u} \mathbb{E}[(2\mathbb{1}(\hat{g}(X) > 0) - 1)|X] f_X(x)\right) du \\
&= \frac{1}{h_k^2} \int_0^1 L^{(2)}(u) S_1(h_k u) du + \frac{1}{h_k^2} \int_0^1 L^{(2)}(u) S_0(h_k u) du
\end{aligned}$$

Using Taylor expansions for $S_1(\cdot)$ and $S_0(\cdot)$ and integration by parts:

$$\mathbb{E}[\hat{f}_{g^*}''(0^+) - \hat{f}_{g^*}''(0^-)] = S_1''(0) + S_0''(0) + O(h_k)$$

which shows that the bias correction

$$\frac{h_k^2 \mathbb{E}[\hat{f}_{g^*}''(0^+) - \hat{f}_{g^*}''(0^-)]}{2} \int_0^1 \frac{u^2(\alpha_1 + \alpha_2 u) K(u)}{2!} du = \frac{h_k^2 (S_1''(0) + S_0''(0))}{2} \int_0^1 \frac{u^2(\alpha_1 + \alpha_2 u) K(u)}{2!} du + O(h_k^3)$$

is equivalent to the bias of the main estimator. Since the bias-corrected kernel density estimator can be written as the sum of bounded i.i.d random variables, asymptotic normality follows from the Central Limit Theorem. □