DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 18-04

Estimation of regression discontinuities: a generalized reflection approach

Sihong Xie
University of Colorado Boulder

April 2018
Revised October 2018

Department of Economics

University of Colorado Boulder
Boulder, Colorado 80309

Estimation of jump discontinuities in regression: a generalized reflection approach[1]

Sihong Xie

Department of Economics
University of Colorado
Boulder, CO 80309-0256, USA
email: sihong.xie@colorado.edu

October, 2018

(please find the latest version here)

Abstract: We propose a new class of estimators for the size of a jump discontinuity on a nonparametric regression. While there is a vast literature in Econometrics that addresses this issue (see, *inter alia*, Hahn et al. (2001), Porter (2003), Imbens and Lemieux (2008)), the main approach in these studies is to use local polynomial (mostly local linear) approximations for the regression on both sides of the discontinuity. In this paper, we adopt a novel approach. The basic idea of our estimator is to extend the regressions on both sides of the discontinuity using the extension of Hestenes (1941). These two extended regressions are then estimated and used to estimate the jump at the discontinuity. The inspiration for our method comes from recent work by Mynbaev and Martins-Filho (2019), where a simple and elegant solution to boundary problems in density estimation is obtained using the same extension principle. Our work provides a *class* of jump estimators that are easy to construct using classical kernels and bandwidths that are constant over the entire domain of the regression. Focusing on the properties of our estimators at boundary points, we provide their bias, variance and asymptotic distributions and compare them with those of local linear (LL) estimators. We conduct extensive Monte Carlo simulations to contrast the finite sample performance of our estimators with that of the NW and LL estimators and, more importantly, to investigate the sources of bias at the cutoff point. Finally, we apply our estimators to data provided by Litschig and Morrison (2013) to illustrate their empirical applicability and demonstrate some advantages of using our nonparametric estimation over standard procedures in regression discontinuity designs.

Keywords: regression discontinuity designs; estimation of jump discontinuities; Hestenes' extension; boundary bias.

JEL codes: C13, C14.
AMS codes: 62G08, 62G20.

[1]I am very grateful to my advisor, Carlos Martins-Filho, for his immense support and inspiring guidance. He sets a great example of who I want to be and what I can become. I would like to thank Xiaodong Liu and Keith Maskus, for their longtime support in my research. I also want to thank William Ridley for his diligent proofreading of this manuscript and many other documents of mine. Last, but not least, I would like to thank my cohorts and office mates, whose support and encouragement has created a wonderful experience for my Ph.D. study at CU.

1 Introduction

As a method for empirically evaluating the effects of policy or experimental interventions, regression discontinuity (RD) designs have been widely used in Economics, Political Science and other social and behavioral sciences. See Imbens and Lemieux (2008), Lee and Lemieux (2010) and Cattaneo and Escanciano (2017) for theoretical and empirical overviews of the existing literature. RD designs are inherently nonparametric models as identification typically relies only on smoothness assumptions on the relevant conditional expectations at a known threshold or cut-off point in the set where the conditioning covariate (regressor) takes values. It is well known that traditional nonparametric kernel regression estimators, such as Nadaraya-Watson, suffer from boundary problems (see, *inter alia,* Gasser et al., 1979, Gasser and Muller, 1984, Fan, 1992, Härdle and Linton, 1994). Specifically, these estimators have slower rates of convergence for bias at boundary points than at interior points in the regression domain. Under typical assumptions on the smoothness of the regression and regressor density, the traditional Nadaraya-Watson (NW) estimator constructed with bandwidth $h > 0$ has bias of order $O(h)$ at boundary points, compared to $O(h^2)$ at interior points of the regression domain. This problem is particularly relevant for RD designs as the estimation of regression functions at boundary points is precisely the object of interest.

The problem can be aggravated in RD designs, see Porter (2003), as an estimator for the jump discontinuity at the threshold may compound the poor bias behavior of nonparametric estimators of the regression to the right and to left of the threshold. While there is a vast literature in Econometrics and Statistics that attempts to address this issue, (see, *inter alia,* Fan, 1992, Hahn et al., 2001, Porter, 2003, Imbens and Lemieux, 2008, Lee and Lemieux, 2010, Imbens and Kalyanaraman, 2012) the main approach in RD designs is to estimate local polynomial (mostly local linear) approximations for the regressions on both sides of the discontinuity and use these to produce an estimate for the jump discontinuity at the threshold. This approach is justified by Fan and Gijbels (1992) where it is shown that local linear estimators, under standard smoothness assumptions, have bias of order $O(h^2)$ at boundary points. Porter (2003) proposes RD estimators - partially linear local polynomial estimators - that can achieve smaller order biases at boundary points by using high order kernels. However, these estimators require identical regression functions (separated by a jump) on both sides of the threshold, a restriction that is not required by typical nonparametric estimators when applied to to data to the left and right sides of a point of discontinuity.

In this paper, we adopt a novel approach. The basic idea behind our estimation procedure is to extend regressions beyond the boundary of their domains to the entire real line, using an extension proposed by Hestenes (1941). These extended regressions are then estimated using a generalized reflection approach and used to estimate jump discontinuities. The inspiration for our method comes from Mynbaev and Martins-Filho (2019), where a simple and elegant solution to boundary problems in density estimation is obtained using the same extension principle. Their solution can be applied not only to densities but also to any sufficiently smooth function, such as suitably defined regressions. Building on their work, we apply Hestenes' extension to estimate regressions that have jump discontinuities and can thus be viewed as comprising two regimes with boundaries: one to the left and one to the right of the point of discontinuity. Regression functions on each side of the discontinuity can be different and, in particular, can have different degrees of smoothness. In essence, instead of using higher order polynomial functions to reduce bias, we use a generalized reflection method of extending the regression functions across discontinuity points to reduce bias.

In fact, as outlined in section 2 of this paper, our estimation strategy produces a *class* of jump discontinuity estimators that is an alternative to the commonly used procedures based on local polynomial regression. What distinguishes the elements in the class are the types of Hestenes' extension used. Our estimators are constructed based on the algebraic structure of the classical NW estimator. However, contrary to the NW estimator that suffers from the aforementioned boundary problems (slow rates of bias decay and, in some cases, inconsistency), our estimators have boundary behavior that is completely analogous to that at interior points of the regression domain. Thus, we restore bias behavior at boundary points to be the same of that at interior points. As is the case for the NW estimator, the estimators we propose are easy to construct, require no modification to commonly used kernels and allow for a common bandwidth over the entire domain of the regressions.

Focusing on properties at boundary points, we derive the bias, variance and asymptotic distribution of our estimators. In addition, we provide a theoretical comparison between our estimator and the popular local linear (LL) estimator. This estimator has the same unconditional bias of order $O(h^2)$ and variance order $O((nh)^{-1})$, where n is the sample size, but with different magnitudes for both bias and variance. Our estimators eliminate the boundary problem by circumventing partial integration of kernels, but the size of the bias and variance is a function of by the type of Hestenes (1941) extension used. Alternatively, the size of the bias and variance of LL estimators are impacted by partial integration of kernels. Hence, for a given kernel, our estimators may outperform LL estimators, since researchers have control over the selection of which type of extension is used.

We have conducted extensive Monte Carlo simulations to shed light on the finite sample behavior of our estimators. We contrast their finite sample performance with that of the NW and LL estimators, and compare bias and variance of our estimators to their theoretical values. Our simulations confirm our theoretical results. In particular, our estimators are free of boundary problems and perform better than NW estimators in all cases. Compared to LL estimators, our estimators have the same bias order and similar, or in some cases, smaller bias size.

To illustrate the applicability of our estimators in empirical settings, we apply them to data used in Litschig and Morrison (2013), where a RD design is used to examine the degree to which intergovernmental transfers impact education and poverty outcomes in Brazil. We use our estimators and the NW and LL estimators to estimate the jump of the conditional mean of the treatment and outcome variables, and the Hestenes-based density estimators of Mynbaev and Martins-Filho (2019) and LL density estimators of Cheng (1994) to estimate the jump on the density of the running variable. Litschig and Morrison (2013) use the standard least squares method to implement a local linear estimator to estimate jumps, which is the typical approach in empirical RD design studies. We demonstrate that our approach is more flexible than theirs in several ways: first, we choose bandwidths optimally while their choice of bandwidth is arbitrary, following no particular optimization criterion; second, by using a Gaussian kernel, all data influences the estimation of the jump while they only use data in the vicinity (defined by the bandwidth) of the jump.

The rest of this paper is organized as follows. In section 2, we introduce the Hestenes-based estimators for regressions with a domain that includes boundary points. Then we derive the bias, variance and asymptotic distribution of our estimator for points in (and outside) a vicinity of the boundary. We compare these properties with those of LL estimators. In section 3, we connect regression estimators to estimation of a jump discontinuity within the context of the RD design literature. We define estimators for the jump discontinuity and establish consistency and asymptotic normality of the estimator. In section 4, we conduct Monte Carlo simulations that compare our estimators to NW and LL estimators. In section 5, we give an empirical illustration of our methodology. Section 6 gives some concluding remarks and topics for future study. Supporting lemmata and proofs are collected in the Appendix.

2 A Hestenes-based regression estimator

We start by considering an independent and identically distributed sequence of random vectors in $\mathbb{R}\times[0,\infty)$, denoted by $\{(Y_i,X_i)\}_{i=1}^n$, each of which is distributed as (Y,X), with

$$E(Y|X=x)=m(x). \tag{1}$$

We assume that the marginal density f of X exists and that $m,f\in C_b^s\left([0,\infty)\right)$, where $C_b^s([0,\infty))$ the class of functions $f:\mathbb{R}\to\mathbb{R}$ whose support is $[0,\infty)$ and which is s-times differentiable with $|f^{(s)}(x)|\le C$ for some $0<C<\infty$, $s\in\mathbb{N}$ and $f^{(s)}(x)$ denotes the derivative of order $s\in\mathbb{N}$ of f. To overcome boundary problems when estimating m, we smoothly extend the function $h(x)\equiv m(x)f(x)$ from its original domain $[0,\infty)$ to $(-\infty,0)$. Its smooth extensions to $(-\infty,0)$ will be denoted by $\phi(x)$ for $x<0$ and are given below. In particular, we scale h and reflect it over from the nonnegative to the negative side of the real line up to $s+1$ times. The extension ϕ is a linear combination of the reflection of scaled h functions which satisfy sewing conditions that preserve continuity and derivatives up to order s. We will then use the observations on (Y,X), which include only nonnegative values of X to estimate $\mu(x)$, which is defined on the whole real line by piecing together h and ϕ.

Specifically, let $w_1, \ldots, w_{s+1}$ be pairwise distinct positive numbers for $s = 0, 1, \ldots$ such as, $w_i = 1/i$, or $w_i = i$, for $i = 1, \ldots, s+1$. Also, let the numbers $k_1, \ldots, k_{s+1}$ be defined by the following system of equations

$$\sum_{i=1}^{s+1} (-w_i)^j k_i = 1, \ j = 0, \ldots, s. \tag{2}$$

The determinant of this system (Vandermonde) is nonsingular, i.e.,

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ -w_1 & -w_2 & \cdots & -w_{s+1} \\ \vdots & \vdots & \vdots & \vdots \\ (-w_1)^s & (-w_2)^s & \cdots & (-w_{s+1})^s \end{vmatrix} \neq 0,$$

and consequently $k_1, \ldots, k_{s+1}$ are uniquely defined for any choice of $\{w_i\}_{i=1}^{s+1}$. Then, the Hestenes' extensions of h to $(-\infty, 0)$ are given by

$$\phi(x) = \sum_{j=1}^{s+1} k_j h(-w_j x).$$

It follows immediately that sewing conditions are satisfied due to (2), and we have

$$\phi^{(d)}(0-) = \sum_{j=1}^{s+1} (-w_j)^d k_j h^{(d)}(0+) = h^{(d)}(0+), \ d = 0, \cdots, s,$$

where for an arbitrary function $g : \mathbb{R} \to \mathbb{R}$ and $\varepsilon > 0$, $g(x+) = \lim_{\varepsilon \downarrow 0} g(x + \varepsilon)$ and $g(x-) = \lim_{\varepsilon \downarrow 0} g(x - \varepsilon)$.

Now, we define

$$\mu(x) \equiv \begin{cases} m(x)f(x), & x \geq 0 \\ \sum_{j=1}^{s+1} k_j m(-w_j x) f(-w_j x), & x < 0 \end{cases} \tag{3}$$

with $\mu(x)$ being s-times differentiable. This follows from s-times differentiability of f and m as $[m(x)f(x)]^{(d)} = \sum_{j=0}^{d} \begin{pmatrix} d \\ j \end{pmatrix} m^{(d-j)}(x) f^{(j)}(x)$. Note that in our context, differentiability of $m(x)$ at $x = 0$ must be understood as differentiability from the right, i.e., $m^{(d)}(0+)$.[1] In the following subsections, we construct estimators for $m(x)$ where $x \in [0, \infty)$ and study their bias, variance and asymptotic distributions for points in a vicinity of zero, and compare them to cases where x is an interior point, i.e., outside this vicinity.

[1] As an example, the sewing condition for $d = 1$ is satisfied as

$$\begin{aligned} \phi^{(1)}(0-) &= \lim_{\varepsilon \downarrow 0} \phi^{(1)}(0 - \varepsilon) = \lim_{\varepsilon \downarrow 0} \sum_{j=1}^{s+1} k_j h^{(1)}(-w_j(0 - \varepsilon)) \\ &= \lim_{\varepsilon \downarrow 0} \sum_{j=1}^{s+1} k_j \left[m^{(1)}(-w_j(0 - \varepsilon)) f(-w_j(0 - \varepsilon)) + m(-w_j(0 - \varepsilon)) f^{(1)}(-w_j(0 - \varepsilon)) \right] \\ &= \sum_{j=1}^{s+1} (-w_j) k_j \left[m^{(1)}(0^+) f(0) + m(0^+) f^{(1)}(0) \right] = h^{(1)}(0+). \end{aligned}$$

2.1 An infeasible Hestenes-based regression estimator

We start by defining an infeasible Hestenes-based regression estimator $m_H(x)$ for $m(x)$ when $x \geq 0$ which assumes that f is known. We will need a kernel function K that will satisfy,

Assumption 1. *K is uniformly bounded and symmetric with $\int_{\mathbb{R}} K(u)du = 1$, $\int_{\mathbb{R}} uK(u)du = 0$ and $\int_{\mathbb{R}} \left|u^i K(u)\right| du < C$ for $i = 1, 2, 3, 4$ and some $0 < C < \infty$.*

If $f(x) > 0$, we define

$$m_H(x) = \frac{1}{f(x)} \frac{1}{nh} \sum_{i=1}^{n} \left[K\left(\frac{X_i - x}{h}\right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K\left(\frac{\frac{X_i}{w_j} + x}{h}\right) \right] Y_i, \tag{4}$$

where n is taken to be the size of a random sample of observations $\{(Y_i, X_i)\}_{i=1}^n$ and $0 < h$ is a sequence of nonstochastic bandwidths that depend on n such that $h \to 0$ as $n \to \infty$.

The algebraic structure of the estimator is motivated by Mynbaev and Martins-Filho (2019) where density estimators are constructed based on Hestenes' extension. Since m_H depends on s and the sequence $\{w_j\}_{j=1}^{s+1}$, equation (4) defines a *class* of estimators whose elements are indexed by $\{w_j\}_{j=1}^{s+1}$. For instance, $w_j = 1/j$ or $w_j = j$ have been suggested in Mynbaev and Martins-Filho (2019) and will produce different estimators in the class. Once $\{w_j\}_{j=1}^{s+1}$ is chosen, the sequence $\{k_j\}_{j=1}^{s+1}$ is uniquely defined by (2) and every estimator in the class is uniquely indexed by $\{w_j\}_{j=1}^{s+1}$.

The following theorem gives an integral representation for the bias of m_H. Its proof gives the mathematical motivation for the algebraic structure of the estimator. In what follows we adopt the following notation: $\kappa_i = \int_{\mathbb{R}} u^i K(u)du$, $\mu_{i,x} = \int_{-\infty}^{x} u^i K(u)du$, $\kappa_{i,x} = \int_x^{\infty} u^i K(u)du$, $\lambda_i = \int_{\mathbb{R}} u^i K^2(u)du$, $\eta_{i,x} = \int_{-\infty}^{x} u^i K^2(u)du$ and $\lambda_{i,x} = \int_x^{\infty} u^i K^2(u)du$ for $i = 0, 1, 2, 3, 4$.

Theorem 1. *Suppose that K satisfies Assumption 1. Then,*

$$E\left(m_H(x)\right) - m(x) = \frac{1}{f(x)} \int_{\mathbb{R}} K(\psi) \left[\mu(x - h\psi) - \mu(x)\right] d\psi. \tag{5}$$

where $\mu(x)$ is as defined in equation (3). If, in addition, $f, m \in C_b^4([0, \infty))$

$$E\left(m_H(x)\right) - m(x) = h^2 \frac{\kappa_2}{2} \left[m^{(2)}(x) + \frac{2m^{(1)}(x) f^{(1)}(x)}{f(x)} + \frac{f^{(2)}(x) m(x)}{f(x)} \right] + O(h^4). \tag{6}$$

Proof. Since $\{Y_i, X_i\}_{i=1}^n$ is independent and identically distributed sequence

$$E\left(m_H(x)|X_1,\cdots,X_n\right) = \frac{1}{f(x)}\frac{1}{nh}\sum_{i=1}^{n}\left[K\left(\frac{X_i-x}{h}\right)+\sum_{j=1}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{X_i}{w_j}+x}{h}\right)\right]m(X_i)$$

$$\begin{aligned}
E\left(m_H(x)\right) &= \frac{1}{f(x)}\frac{1}{h}E\left[K\left(\frac{X_1-x}{h}\right)m(X_1)d_1\right]+\frac{1}{f(x)}\frac{1}{h}E\left[\sum_{j=1}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{X_1}{w_j}+x}{h}\right)m(X_1)\right]\\
&= \frac{1}{f(x)}\frac{1}{h}\int_0^\infty K\left(\frac{X_1-x}{h}\right)m(X_1)f(X_1)dX_1\\
&+\frac{1}{f(x)}\frac{1}{h}\int_0^\infty\sum_{j=1}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{X_1}{w_j}+x}{h}\right)m(X_1)f(X_1)dX_1.
\end{aligned}$$

Since K is symmetric, in the first integral let $\psi=\frac{x-X_1}{h}$ and in the second integral, let $\psi=\frac{\frac{X_1}{w_j}+x}{h}$. Then,

$$\begin{aligned}
E(m_H(x)) &= \frac{1}{f(x)}\int_{-\infty}^{\frac{x}{h}}K(\psi)m(x-h\psi)f(x-h\psi)d\psi\\
&+\frac{1}{f(x)}\int_{\frac{x}{h}}^{\infty}\sum_{j=1}^{s+1}k_jK(\psi)m\left(-(x-h\psi)w_j\right)f\left(-(x-h\psi)w_j\right)d\psi
\end{aligned}$$

In the first integral, we have $x-h\psi\geq 0$ and in the second integral, we have $x-h\psi<0$. Hence,

$$E\left(m_H(x)\right)=\frac{1}{f(x)}\int_{-\infty}^{\infty}K(\psi)\mu(x-h\psi)d\psi. \tag{7}$$

Since $\int_{\mathbb{R}}K(\psi)=1$,

$$E\left(m_H(x)\right)-m(x)=\frac{1}{f(x)}\int_{-\infty}^{\infty}K(\psi)\left[\mu(x-h\psi)-\mu(x)\right]d\psi \text{ for } x\geq 0. \tag{8}$$

For the second part of the theorem note that since $f,m\in C_b^4([0,\infty))$ for $x\geq 0$, we have $\mu^{(4)}(x)=\sum_{j=0}^4\begin{pmatrix}4\\j\end{pmatrix}m^{(4-j)}(x)f^{(j)}(x)$. Using diffentiability of μ we have

$$\begin{aligned}
E\left(m_H(x)\right)-m(x) &= \frac{1}{f(x)}\int_{-\infty}^{\infty}K(\psi)\left(\mu^{(1)}(x)(-h\psi)+\frac{1}{2}\mu^{(2)}(x)(-h\psi)^2\right.\\
&\left.+\frac{1}{6}\mu^{(3)}(x)(-h\psi)^3+\frac{1}{24}\mu^{(4)}(\bar{x})(-h\psi)^4\right)d\psi \text{ where } \bar{x}=\alpha x+(1-\alpha)(x-h\psi),\alpha\in[0,1]\\
&= \frac{1}{f(x)}\int_{-\infty}^{\infty}K(\psi)\left(\frac{1}{2}\mu^{(2)}(x)(-h\psi)^2+\frac{1}{24}\mu^{(4)}(\bar{x})(-h\psi)^4\right)d\psi
\end{aligned}$$

since K is symmetric. Now, $\left|\int_{-\infty}^{\infty}K(\psi)\psi^4\mu^{(4)}(\bar{x})d\psi\right|\leq\int_{-\infty}^{\infty}|K(\psi)|\,\psi^4\left|\mu^{(4)}(\bar{x})\right|d\psi\leq C$ for some $C<\infty$, provided $|\mu^{(4)}(x)|<C<\infty$, and $\int_{-\infty}^{\infty}|K(\psi)|\,\psi^4d\psi<C$. Consequently,

$$\begin{aligned}
E\left(m_H(x)\right)-m(x) &= \left(\frac{\mu^{(2)}(x)}{2f(x)}\int_{-\infty}^{\infty}K(\psi)\psi^2d\psi\right)h^2+O(h^4)=\frac{\mu^{(2)}(x)}{2f(x)}\kappa_2h^2+O(h^4)\\
&= \frac{h^2}{2}\left[m^{(2)}(x)+\frac{2m^{(1)}(x)f^{(1)}(x)}{f(x)}+\frac{f^{(2)}(x)m(x)}{f(x)}\right]\kappa_2h^2+O(h^4)
\end{aligned}$$

□

The integral representation given for the bias of $m_H(x)$ given in equation (5) is the key insight in constructing the estimator. It shows that the representation obtained for traditional estimators, such as NW, can be obtained with the function μ in place of the regressor density. The algebraic structure of m_H permits the unification of the bias representations for x in and outside a neighborhood of 0. As a direct consequence of the method of proof of Theorem 1, if, as usually done in the nonparametric kernel literature, the kernel K is of order s and $m, f \in C_b^s([0,\infty))$ then the bias of m_H is $O(h^s)$.

At the boundary point $x = 0$ we have,

$$E\left(m_H(0)\right) - m(0) = \frac{h^2}{2}\left[m^{(2)}(0) + \frac{2m^{(1)}(0)f^{(1)}(0)}{f(0)} + \frac{f^{(2)}(0)m(0)}{f(0)}\right]\kappa_2 + O(h^4).$$

It is instructive to compare this expression to that of the bias for an infeasible NW estimator given by $m_{NW}(x) = \frac{1}{nhf(x)}\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)Y_i$. The following corollary follows directly from Theorem 1.

Corollary 1. *Under the assumptions of Theorem 1, for $x \geq 0$, the bias of $m_{NW}(x)$ is given by*

$$\begin{aligned} E(m_{NW}(x)) - m(x) = &-m(x)\kappa_{0,\frac{x}{h}} - h\left(\frac{m(x)f^{(1)}(x)}{f(x)} + m^{(1)}(x)\right)\mu_{1,\frac{x}{h}} \\ &+ h^2\left(\frac{m(x)f^{(2)}(x)}{2f(x)} + \frac{m^{(1)}(x)f^{(1)}(x)}{f(x)} + \frac{m^{(2)}(x)}{2}\right)\mu_{2,\frac{x}{h}} + O(h^3). \end{aligned}$$

The slower order of the remainder term results from the fact that the symmetry of K can no longer be used to eliminate the term of order h^3. Hence, at $x = 0$

$$\begin{aligned} E(m_{NW}(0)) - m(0) \quad = \quad &-\frac{1}{2}m(0) - h\left(\frac{m(0)f^{(1)}(0)}{f(0)} + m^{(1)}(0)\right)\mu_{1,0} \\ &+\frac{h^2}{2}\left(m^{(2)}(0) + \frac{2m^{(1)}(0)f^{(1)}(0)}{f(0)} + \frac{f^{(2)}(0)m(0)}{f(0)}\right)\mu_{2,0} + O(h^3). \end{aligned}$$

We note that this expression suggests that m_{NW} is inconsistent at the boundary. Although the coefficient on the term of order h^2 in the bias of m_{NW} is half the size of the corresponding term in m_H, it has little impact on bias magnitude since m_{NW} has two extra terms that are of larger magnitude.

The following theorem provides approximations for the variance of m_H.

Theorem 2. *Suppose $E((Y - m(X))^2|X = x) = \sigma^2$, $m, f \in C_b^4([0,1)$ and K satisfies Assumption 1. Then, if $nh \to \infty$ and $x \geq 0$, the variance of $m_H(x)$ is given by*

$$V(m_H(x)) = \begin{cases} \frac{1}{nh}\frac{m^2(0)+\sigma^2}{f(0)}\int_0^\infty\left[\sum_{i=0}^{s+1}\frac{k_i}{w_i}K\left(\frac{u}{w_i}\right)\right]^2 du + o((nh)^{-1}), & x = 0 \\ \frac{1}{nh}\frac{m^2(x)+\sigma^2}{f(x)}\lambda_0 + o((nh)^{-1}), & x > 0. \end{cases}$$

Proof. Write $m_H(x) = \frac{\hat{g}(x)}{f(x)}$ where $f(x) \neq 0$ and $\hat{g}(x) = \frac{1}{nh}\sum_{i=1}^n\left[K\left(\frac{X_i-x}{h}\right) + \sum_{j=1}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{X_i}{w_j}+x}{h}\right)\right]Y_i$ and put $u_i = \frac{1}{h}\left[K\left(\frac{X_i-x}{h}\right) + \sum_{j=1}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{X_i}{w_j}+x}{h}\right)\right]Y_i$. Then letting $w_0 = -1$, and $k_0 = -1$ we can write

$u_i = \frac{1}{h}\sum_{j=0}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{X_i}{w_j}+x}{h}\right)Y_i$ and $\hat{g}(x) = \frac{1}{n}\sum_{i=1}^{n}u_i$. Consequently, $V(m_H(x)) = \frac{1}{nf^2(x)}\left(Eu_1^2 - E(u_1)^2\right)$. Now,

$$E(u_1^2) = E\left[\sum_{j=0}^{s+1}\frac{1}{h}\frac{k_j}{w_j}K\left(\frac{\frac{X_1}{w_j}+x}{h}\right)Y_1\right]^2$$

$$= \frac{1}{h^2}\sum_{i,j=0}^{s+1}\frac{k_i}{w_i}\frac{k_j}{w_j}E\left[K\left(\frac{\frac{X_1}{w_i}+x}{h}\right)K\left(\frac{\frac{X_1}{w_j}+x}{h}\right)Y_1^2\right]$$

$$= \frac{1}{h^2}\sum_{i,j=0}^{s+1}\frac{k_i}{w_i}\frac{k_j}{w_j}E\left[K\left(\frac{\frac{X_1}{w_i}+x}{h}\right)K\left(\frac{\frac{X_1}{w_j}+x}{h}\right)(m(X_1)+\epsilon_1)^2\right]$$

where $\epsilon_1 = Y_1 - m(X_1)$.

$$= \frac{1}{h^2}\sum_{i,j=0}^{s+1}\frac{k_i}{w_i}\frac{k_j}{w_j}\left\{E\left[K\left(\frac{\frac{X_1}{w_i}+x}{h}\right)K\left(\frac{\frac{X_1}{w_j}+x}{h}\right)m^2(X_1)\right] + \sigma^2 E\left[K\left(\frac{\frac{X_1}{w_i}+x}{h}\right)K\left(\frac{\frac{X_1}{w_j}+x}{h}\right)\right]\right\}$$

$$= \frac{1}{h^2}\sum_{i,j=0}^{s+1}\frac{k_i}{w_i}\frac{k_j}{w_j}\int_0^\infty K\left(\frac{\frac{t}{w_i}+x}{h}\right)K\left(\frac{\frac{t}{w_j}+x}{h}\right)m^2(t)f(t)dt$$

$$+ \frac{\sigma^2}{h^2}\sum_{i,j=0}^{s+1}\frac{k_i}{w_i}\frac{k_j}{w_j}\int_0^\infty K\left(\frac{\frac{t}{w_i}+x}{h}\right)K\left(\frac{\frac{t}{w_j}+x}{h}\right)f(t)dt$$

$$= T_1 + T_2$$

We first study T_1 and consider two cases $x > 0$ and $x = 0$.
Case ($x > 0$): Letting $I_{ij} = \frac{1}{h}\int_0^\infty K\left(\frac{\frac{t}{w_i}+x}{h}\right)K\left(\frac{\frac{t}{w_j}+x}{h}\right)m^2(t)f(t)dt$, note that $hT_1 = I_{00} + \sum_{i+j>0}^{s+1} I_{ij}$. Now,

$$\left|I_{00} - m^2(x)f(x)\int_{\mathbb{R}}K^2(u)du\right| = \left|\int_{-\infty}^{\frac{x}{h}}K^2(u)m^2(x-hu)f(x-hu)du - m^2(x)f(x)\int_{\mathbb{R}}K^2(u)du\right|$$

$$= \left|\int_{\mathbb{R}}K^2(u)\left[m^2(x-hu)f(x-hu) - m^2(x)f(x)\right]du - \int_{\frac{x}{h}}^{\infty}K^2(u)m^2(x-hu)f(x-hu)du\right|$$

$$\leq \left|\int_{\mathbb{R}}K^2(u)\left[m^2(x-hu)f(x-hu) - m^2(x)f(x)\right]du\right| + \left|\int_{\frac{x}{h}}^{\infty}K^2(u)m^2(x-hu)f(x-hu)du\right|$$

$$\leq \left|\int_{|u|\leq C}K^2(u)\left[m^2(x-hu)f(x-hu) - m^2(x)f(x)\right]du\right| + \left|\int_{|u|>C}K^2(u)\left[m^2(x-hu)f(x-hu) - m^2(x)f(x)\right]du\right| + \left|\int_{\frac{x}{h}}^{\infty}K^2(u)m^2(x-hu)f(x-hu)du\right|$$

Let $\bar{p}(\delta, x) = \underset{|y|\le\delta}{sup}|f(x-y)-f(x)|$, and since $f \in C_b^4([0,\infty))$ we have $f(x-hu)-f(x) \le \bar{p}(Ch,x)$. Thus,

$$\left|\int_{|u|\le C} K^2(u)\left[m^2(x-hu)f(x-hu)-m^2(x)f(x)\right]du\right| \le C\bar{p}(Ch,x)\int_{|u|\le C}K^2(u)du.$$

Consequently, since $m \in C_b^4([0,\infty))$

$$\left|I_{00} - m^2(x)f(x)\int_{\mathbb{R}}K^2(u)du\right| \le C\bar{p}(Ch,x)\int_{|u|\le C}K^2(u)du + C\int_{|u|>C}K^2(u)du + C\int_{\frac{x}{h}}^{\infty}K^2(u)du$$

For C be sufficiently large and h, ϵ sufficiently small, by continuity of f, $\bar{p}(Ch,x) < \epsilon$. Since $\int_R |K(u)|^2 du < C$, $\int_{|u|>C} K^2(u)du < \epsilon$ and $\int_{\frac{x}{h}}^{\infty} K^2(u)du < \epsilon$. Therefore, for all $\epsilon > 0$,

$$\left|I_{00} - m^2(x)f(x)\int_{\mathbb{R}}K^2(u)du\right| \le \epsilon. \tag{9}$$

Now we turn attention to I_{ij} where $i+j>0$, and without loss of generality take $w_i > 0$. Changing variables by setting $u = \frac{1}{h}(t + xw_i)$, $I_{ij} = \frac{1}{h}\int_{\frac{xw_i}{h}}^{\infty} K\left(\frac{u}{w_i}\right) K\left(\frac{x}{h}(1-\frac{w_i}{w_j}) + \frac{u}{w_j}\right) m^2(hu - xw_i)f(hu-xw_i)du$. Given the uniform boundedness of $K(x)$, $f(x)$ and $m^2(x)$,

$$I_{ij} \le Cw_i \int_{\frac{x}{h}}^{\infty} K(u)\,du \tag{10}$$

where $\int_{\frac{x}{h}}^{\infty} K(u)\,du < \epsilon$ for sufficiently small h. Consequently, inequalities (9) and (10) give

$$\left|hT_1 - m^2(x)f(x)\int_{\mathbb{R}}K^2(u)du\right| \le \epsilon.$$

We now turn to T_2. Let $J_{ij} = \frac{1}{h}\int_0^{\infty} K\left(\frac{\frac{t}{w_i}+x}{h}\right) K\left(\frac{\frac{t}{w_j}+x}{h}\right) f(t)dt$, then $hT_2 = \sigma^2(J_{00} + \sum_{i+j>0}^{s+1} J_{ij})$. Using arguments similar to those for I_{00}, we have $\left|J_{00} - f(x)\int_{\mathbb{R}}K^2(u)du\right| \le \epsilon$. Again, similar to the case of I_{ij}, we have $J_{ij} \le Cw_i \int_{\frac{x}{h}}^{\infty} K(u)\,du < \epsilon$ for sufficiently small h. Thus,

$$\left|hT_2 - \sigma^2 f(x)\int_{\mathbb{R}}K^2(u)du\right| \le \epsilon \tag{11}$$

Since $Eu_1^2 = T_1 + T_2$, we have

$$\left|hEu_1^2 - \left(m^2(x)+\sigma^2\right)f(x)\int_{\mathbb{R}}K^2(u)du\right| \le \epsilon \tag{12}$$

From Theorem 1, $hE(u_1)^2 = o(1)$, consequently

$$V(m_H(x)) = \frac{1}{nh}\left\{\frac{m^2(x)+\sigma^2}{f(x)}\int_{\mathbb{R}}K^2(u)du + o(1)\right\}.$$

Case $x=0$: First, we consider T_1. $I_{00} = \frac{1}{h}\int_0^\infty K^2(\frac{0-t}{h})m^2(t)f(t)dt = \int_{-\infty}^0 K^2(u)m^2(-hu)f(-hu)du$ and

$$\begin{aligned}
\left|I_{00} - m^2(0)f(0)\int_{-\infty}^0 K^2(u)du\right| &= \left|\int_{-\infty}^0 K^2(u)\left[m^2(-hu)f(-hu)du - m^2(0)f(0)\right]du\right| \\
&= \left|\int_{-\infty}^{-C} K^2(u)\left[m^2(-hu)f(-hu) - m^2(0)f(0)\right]du\right. \\
&\quad \left. + \int_{-C}^0 K^2(u)\left[m^2(-hu)f(-hu) - m^2(0)f(0)\right]du\right| \\
&\le C\int_{-\infty}^{-C} K^2(u)du + \bar{p}(Ch,0)\int_0^C K^2(u)du
\end{aligned}$$

For C sufficiently large and ϵ, h sufficiently small we have

$$\left|I_{00} - m^2(0)f(0)\int_{-\infty}^0 K^2(u)du\right| \le \epsilon. \tag{13}$$

Now we consider the case where $i+j>0$. Note that $I_{ij} = \int_0^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)m^2(hu)f(hu)du$.

$$\begin{aligned}
|I_{ij} &- m^2(0)f(0)\int_0^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)du\Big| = \left|\int_0^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)\left[m^2(hu)f(hu) - m^2(0)f(0)\right]du\right| \\
&= \left|\int_0^C K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)\left[m^2(hu)f(hu) - m^2(0)f(0)\right]du\right. \\
&\quad \left. + \int_C^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)\left[m^2(hu)f(hu) - m^2(0)f(0)\right]du\right| \\
&\le \bar{p}(hC,0)\int_0^C \left|K\left(\frac{u}{w_i}\right)\right|\left|K\left(\frac{u}{w_j}\right)\right|du + C\int_C^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)du
\end{aligned}$$

where for sufficiently large C, and for all $\epsilon > 0$, $\left|\int_C^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)du\right| < \epsilon$ and for sufficient small h, $\bar{p}(hC,0) < \epsilon$. Thus,

$$\left|I_{ij} - m^2(0)f(0)\int_0^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)du\right| < \epsilon \tag{14}$$

Consequently, (13) and (14) give

$$\left|hT_1 - m^2(0)f(0)\sum_{i,j=0}^{s+1}\frac{k_i}{w_i}\frac{k_j}{w_j}\int_0^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)du\right| < \epsilon.$$

Turning to the term T_2.

$$hT_2 = \frac{\sigma^2}{h}\int_0^\infty K^2\left(\frac{t}{h}\right)f(t)dt + \frac{\sigma^2}{h}\sum_{i+,j>0}^{s+1}\frac{k_i}{w_i}\frac{k_j}{w_j}\int_{-\infty}^\infty K\left(\frac{t}{w_ih}\right)K\left(\frac{t}{w_jh}\right)f(t)dt.$$

The first term $\frac{\sigma^2}{h}\int_0^\infty K^2\left(\frac{t}{h}\right)f(t)dt = \sigma^2\int_0^\infty K^2(u)f(hu)du = \sigma^2 f(0)\int_0^\infty K^2(u)du + o(1)$. From the second term, $\frac{1}{h}\int_0^\infty K\left(\frac{t}{w_ih}\right)K\left(\frac{t}{w_jh}\right)f(t)dt = \int_0^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)f(hu)du$. Now,

$$\begin{aligned}
&\left|\int_0^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)f(hu)du - f(0)\int_0^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)du\right| \le \\
&\left|\int_0^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)\left[f(hu) - f(0)\right]du\right| < \epsilon.
\end{aligned}$$

by the continuity of f and the dominated convergence theorem. Thus,

$$\left| hT_2 - \sigma^2 f(0) \sum_{i,j=0}^{s+1} \frac{k_i}{w_i} \frac{k_j}{w_j} \int_0^\infty K\left(\frac{u}{w_i}\right) K\left(\frac{u}{w_j}\right) du \right| \le \epsilon \tag{15}$$

Since $Eu_1^2 = T_1 + T_2$, we have

$$\left| hEu_1^2 - (m^2(0) + \sigma^2) f(0) \sum_{i,j=0}^{s+1} \frac{k_i}{w_i} \frac{k_j}{w_j} \int_0^\infty K\left(\frac{u}{w_i}\right) K\left(\frac{u}{w_j}\right) du \right| \le \epsilon \tag{16}$$

Thus,

$$V(m_H(0)) = \frac{1}{nhf^2(0)} \left(hEu_1^2 - hE^2 u_1 \right) = \frac{1}{nh} \left\{ \frac{m^2(0) + \sigma^2}{f(0)} \sum_{i,j=0}^{s+1} \frac{k_i}{w_i} \frac{k_j}{w_j} \int_0^\infty K\left(\frac{u}{w_i}\right) K\left(\frac{u}{w_j}\right) du + o(1) \right\}.$$

Combining the two cases, we have

$$V(m_H(x)) = \begin{cases} \frac{1}{nh} \left\{ \frac{m^2(0)+\sigma^2}{f(0)} \int_0^\infty \left[\sum_{i=0}^{s+1} \frac{k_i}{w_i} K\left(\frac{u}{w_i}\right) \right]^2 du + o(1) \right\}, & x = 0 \\ \frac{1}{nh} \left\{ \frac{m^2(x)+\sigma^2}{f(x)} \int_{\mathbb{R}} K^2(u) du + o(1) \right\}, & x > 0. \end{cases}$$

□

The expressions for the variance of $m_H(x)$ given in Theorem 2 are analogous to those obtained in Mynbaev and Martins-Filho (2019) (see their equations (10) and (11)). The following corollary to Theorem 2 gives an expression for the variance of the infeasible NW estimator.

Corollary 2. *Suppose $E((Y - m(X))^2 | X = x) = \sigma^2$, $m, f \in C_b^4([0, 1)$ and K satisfies Assumption 1. Then, if $nh \to \infty$ and $x \ge 0$, the variance of $m_{NW}(x)$ is given by*

$$V(m_{NW}(x)) = \frac{1}{nh} \frac{m^2(x) + \sigma^2}{f(x)} \eta_{0,x/h} + o((nh)^{-1}).$$

Note that m_H and m_{NW} have the same variance at interior points, but different variances at the boundary point ($x = 0$). With suitable choice of w_i it may be possible to have the leading term of the expression in $V(m_H(0)) \le V(m_{NW}(0))$.

Remark 1. *An optimal plug-in bandwidth h_{pi} for $m_H(0)$ can be obtained by minimizing asymptotic mean squared error (AMSE) at the boundary $x = 0$. As such, consider the asymptotic mean squared error (AMSE) given by*

$$AMSE(h) = \left\{ \frac{h^2}{2} \left[m^{(2)}(0) + \frac{2m^{(1)}(0) f^{(1)}(0)}{f(0)} + \frac{f^{(2)}(0) m(0)}{f(0)} \right] \kappa_2 \right\}^2 + \frac{1}{nh} \left\{ \frac{m^2(0) + \sigma^2}{f(0)} \gamma \right\} + s.o.$$

where $\gamma = \int_0^\infty \left[\sum_{i=0}^{s+1} \frac{k_i}{w_i} K\left(\frac{u}{w_i}\right) \right]^2 du$ and and s.o. denotes terms of smaller order. Routine optimization of the leading terms with respect to h gives

$$h_{pi} = n^{-\frac{1}{5}} \left\{ \frac{m^2(0) + \sigma^2}{f(0)} \gamma \right\}^{\frac{1}{5}} \left[m^{(2)}(0) + \frac{2m^{(1)}(0) f^{(1)}(0)}{f(0)} + \frac{f^{(2)}(0) m(0)}{f(0)} \right]^{-\frac{2}{5}} \kappa_2^{-\frac{2}{5}}.$$

2.2 Feasible Hestenes regression estimator $\hat{m}_H^+$

The regression model that motivated the infeasible m_H had regressors taking values in $[0, \infty)$. It is apparent that an identical estimator can be defined for the case where regressors take values in, and the support of m is, $(-\infty, 0]$. When the regressor takes values in $\mathbb{R}$ and there are potentially two regressions, one to the right and one to the left of a discontinuity at $x = 0$, two infeasible Hestenes estimators can be constructed. The first, for the regression to the right of the point of discontinuity,

$$m_H^+(x) \equiv \frac{1}{f(x)} \frac{1}{nh} \sum_{i=1}^{n} \left[K\left(\frac{X_i - x}{h}\right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K\left(\frac{\frac{X_i}{w_j} + x}{h}\right) \right] Y_i d_i \text{ for } x \geq 0 \tag{17}$$

where $d_i = I_{X_i \geq 0}$, and the second for the regression to the left of the point of discontinuity

$$m_H^-(x) \equiv \frac{1}{f(x)} \frac{1}{nh} \sum_{i=1}^{n} \left[K\left(\frac{X_i - x}{h}\right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K\left(\frac{\frac{X_i}{w_j} + x}{h}\right) \right] Y_i(1 - d_i) \text{ for } x \geq 0. \tag{18}$$

In a RD model where the point of discontinuity is $x = 0$, an estimator for the jump at 0, denoted by $J(0)$, is naturally given by

$$J_H(0) = m_H^+(0) - m_H^-(0). \tag{19}$$

Since these are infeasible estimators due to the fact that f is unknown, we define their feasible versions by replacing f with the Rosenblatt-Parzen estimator $\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$. The feasible versions of these estimators are denoted by $\hat{m}_H^+(x)$, $\hat{m}_H^-(x)$ and $\hat{J}_H(0)$.

We note that a critical assumption for identification in RD models is that the density f of the regressor (running variable) be continuous at the point of potential discontinuity, $x = 0$ in this case. As such, $\hat{m}_H^+(x)$ and $\hat{m}_H^-(x)$ are defined with the estimator $\hat{f}$ using all observations available, to the right and to the left of the point of potential discontinuity.

Theorem 3. *Suppose that K satisfies Assumption 1, $f \in C_b^4(\mathbb{R})$ and $m^+ : [0, \infty) \to \mathbb{R}$ is such that $m^+ \in C_b^4([0, \infty))$. Then, for $x \geq 0$, the bias of $\hat{m}_H^+(x)$ is given by*

$$E(\hat{m}_H^+(x)) - m^+(x) = \left[\frac{1}{2} m^{+(2)}(x) + \frac{m^{+(1)}(x) f^{(1)}(x)}{f(x)} \right] \kappa_2 h^2 + O(h^4 + (nh)^{-1}) \tag{20}$$

Proof. Given that K satisfies Assumption 1 and $f(x) > 0$, for n sufficiently large $\hat{f}(x) > 0$ and $E(\hat{f}(X)) > 0$. Thus, using Taylor's Theorem, we expand $\hat{m}_H^+(x) \equiv q(\hat{g}(x), \hat{f}(x)) = \frac{\hat{g}(x)}{\hat{f}(x)}$ at the point $(E(\hat{g}(x)), E(\hat{f}(x)))$ and obtain,

$$\begin{aligned} \hat{m}_H^+(x) = & \frac{E(\hat{g}(x))}{E(\hat{f}(x))} + \frac{1}{E(\hat{f}(x))}(\hat{g}(x) - E(\hat{g}(x))) - \frac{E(\hat{g}(x))}{E(\hat{f}(x))^2}(\hat{f}(x) - E(\hat{f}(x))) \\ & - \left\{ \frac{1}{E(\hat{f}(x))^2} \left(\hat{f}(x) - E(\hat{f}(x))\right)(\hat{g}(x) - E(\hat{g}(x))) - \frac{E(\hat{g}(x))}{E(\hat{f}(x))^3} \left(\hat{f}(x) - E(\hat{f}(x)\right)^2 \right\} + Z_n \end{aligned}$$

where $\hat{g}(x) = \frac{1}{nh}\sum_{i=1}^{n}\left[K\left(\frac{X_i - x}{h}\right) + \sum_{j=1}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{X_i}{w_j}+x}{h}\right)\right]Y_i d_i$ and

$$Z_n(x) = 3(\hat{g}(x) - E\hat{g}(x))(\hat{f}(x) - E\hat{f}(x))^2\int_0^1 (1-t)^2\frac{1}{[E\hat{f}(x) + t(\hat{f}(x) - E\hat{f}(x))]^3}dt$$
$$- 3(\hat{f}(x) - E\hat{f}(x))^3\int_0^1 (1-t)^2\frac{E\hat{g}(x) + t(\hat{g}(x) - E\hat{g}(x))}{\left[E\hat{f}(x) + t(\hat{f}(x) - E\hat{f}(x))\right]^4}dt.$$

Taking the expectations on both sides of the expression for $\hat{m}_H^+(x)$ gives,

$$E(\hat{m}_H^+(x)) = \frac{E(\hat{g}(x))}{E(\hat{f}(x))} - \frac{1}{E^2(\hat{f}(x))}Cov\left(\hat{g}(x), \hat{f}(x)\right) + \frac{E(\hat{g}(x))}{E^3(\hat{f}(x))}V(\hat{f}(x)) + E(Z_n(x)).$$

Now, from Theorem 1, we have

$$E(\hat{g}(x)) = f(x)E(m_H^+(x)) = f(x)m^+(x) + \frac{h^2}{2}\kappa_2\left[f(x)m^{+(2)}(x) + 2m^{(1)}(x)f^{(1)}(x) + f^{(2)}(x)m^+(x)\right] + O(h^4).$$

From standard results for kernel density estimators (see Li and Racine, 2007) $E(\hat{f}(x)) = f(x) + \frac{h^2}{2}\kappa_2 f^{(2)}(x) + O(h^4)$. Thus,

$$\frac{E(\hat{g}(x))}{E(\hat{f}(x))} = m^+(x) + \frac{h^2}{2}\kappa_2\left[m^{+(2)}(x) + 2\frac{m^{+(1)}(x)f^{(1)}(x)}{f(x)}\right] + O(h^4).$$

Now, from Lemma 1, $Cov(\hat{f}(x), \hat{g}(x)) = O(\frac{1}{nh})$, and from standard results for kernel density estimators Li and Racine, 2007, $Var(\hat{f}(x)) = O(\frac{1}{nh})$. Lastly, from Lemma 2, $E(Z_n(x)) = O\left((nh)^{-3/2}\right)$. Thus, $E(\hat{m}_H^+(x)) - m^+(x) = \left[\frac{1}{2}m^{+(2)}(x) + \frac{m^{+(1)}(x)f^{(1)}(x)}{f(x)}\right]\kappa_2 h^2 + O(h^4 + (nh)^{-1})$. □

An identical expression can be found for the bias of $\hat{m}_H^-(x)$, so that for $x \leq 0$

$$E(\hat{m}_H^-(x)) - m^-(x) = \left[\frac{1}{2}m^{-(2)}(x) + \frac{m^{-(1)}(x)f^{(1)}(x)}{f(x)}\right]\kappa_2 h^2 + O(h^4 + (nh)^{-1}). \tag{21}$$

We now provide expressions for the variance of $\hat{m}_H^+(x)$.

Theorem 4. *Suppose that K satisfies Assumption 1, $f \in C_b^4(\mathbb{R})$ and $m^+ : [0,\infty) \to \mathbb{R}$ is such that $m^+ \in C_b^4([0,\infty))$. Then, for $x \geq 0$, the variance of $\hat{m}_H^+(x)$ is given by*

$$V(\hat{m}_H^+(x)) = \begin{cases} \frac{1}{nh}\left\{\frac{m^{+2}(0)+2\sigma^2}{f(0)}\sum_{i,j=0}^{s+1}\frac{k_i}{w_i}\frac{k_j}{w_j}\int_0^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)du\right. & \\ \left. + \frac{m^{+2}(0)}{f(0)}\int_{\mathbb{R}}K^2(u)du + \frac{2m^2(0)}{f(0)}\sum_{j=0}^{s+1}\frac{k_j}{w_j}\int_0^\infty K(u)K\left(\frac{u}{w_j}\right)du + o(1)\right\}, & if\ x = 0 \\ \frac{1}{nh}\left\{\frac{4m^2(x)+\sigma^2}{f(x)}\int_{\mathbb{R}}K^2(u)du + o(1)\right\}, & if\ x > 0. \end{cases}$$

Proof. Write $a = E\left(\hat{g}(x)\right)$ and $b = E\left(\hat{f}(x)\right)$. Then, $\hat{m}_H^+(x) = \frac{\hat{g}(x)}{\hat{f}(x)} = \frac{a}{b} + \frac{1}{b}(\hat{g}(x) - a) - \frac{a}{b^2}(\hat{f}(x) - b)) + S_n(x)$ where

$$S_n(x) = 2(\hat{g}(x) - a)(\hat{f}(x) - b)\int_0^1 (1-t)(-1)\frac{1}{\left[b + t(\hat{f}(x) - b)\right]^2}dt + (\hat{f}(x) - b)^2\int_0^1 (1-t)\frac{2\left[a + t(\hat{g}(x) - a)\right]}{\left[b + t(\hat{f}(x) - b)\right]^3}dt$$

Then, $E\left(\hat{m}^+(x)\right) = \frac{a}{b} + E(S_n)$ and $V(\hat{m}^+(x)) = \frac{1}{b^2}V(\hat{g}(x)) + \frac{a^2}{b^4}V(\hat{f}(x)) - \frac{2a}{b^3}Cov(\hat{g}(x), \hat{f}(x)) + W_n(x)$ where $W_n(x) = V(S_n) + \frac{2}{b}Cov(\hat{g}(x), S_n) - \frac{2a}{b^2}Cov(\hat{f}(x), S_n)$. Now,

$$\frac{1}{b^2}V(\hat{g}(x)) = V(m_H(x)) = \begin{cases} \frac{1}{nh}\left\{\frac{m^{+2}(0)+2\sigma^2}{f(0)}\sum_{i,j=0}^{s+1}\frac{k_i}{w_i}\frac{k_j}{w_j}\int_0^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)du + o(1)\right\}, & x=0 \\ \frac{1}{nh}\left\{\frac{m^{+2}(x)+\sigma^2}{f(x)}\lambda_0 + o(1)\right\}, & x>0 \end{cases}$$

From the properties of the Rosenblatt-Parzen estimator $\hat{f}$, we have $\frac{a^2}{b^4}V(\hat{f}(x)) = \frac{1}{nh}\left\{\frac{m^{+2}(x)}{f(x)}\lambda_0 + o(1)\right\}$. From Lemma 1,

$$Cov(\hat{g}(x), \hat{f}(x)) = \begin{cases} \frac{1}{nh}\left\{m^+(0)f(0)\sum_{,j=0}^{s+1}\frac{k_j}{w_j}\int_0^\infty K(u)K\left(\frac{u}{w_j}\right)du + o(1)\right\}, & x=0 \\ \frac{1}{nh}\left\{m^+(x)f(x)\int_{\mathbb{R}}K^2(u)du + o(1)\right\}, & x>0 \end{cases}$$

and consequently

$$-\frac{2a}{b^3}Cov(\hat{g}(x), \hat{f}(x)) = \begin{cases} \frac{1}{nh}\left\{\frac{2m^2(0)}{f(0)}\sum_{,j=0}^{s+1}\frac{k_j}{w_j}\int_0^\infty K(u)K\left(\frac{u}{w_j}\right)du + o(1)\right\}, & x=0 \\ \frac{1}{nh}\left\{\frac{2m^2(x)}{f(x)}\int_{\mathbb{R}}K^2(u)du + o(1)\right\}, & x>0. \end{cases}$$

Finally, using Lemma 2 we obtain $W_n(x) = O\left((\frac{1}{nh})^{\frac{3}{2}}\right)$. Thus,

$$V(\hat{m}_H^+(x)) = \begin{cases} \frac{1}{nh}\left(\frac{m^{+2}(0)+2\sigma^2}{f(0)}\sum_{i,j=0}^{s+1}\frac{k_i}{w_i}\frac{k_j}{w_j}\int_0^\infty K\left(\frac{u}{w_i}\right)K\left(\frac{u}{w_j}\right)du + \frac{m^{+2}(0)}{f(0)}\lambda_0\right. \\ \left. + \frac{2m^2(0)}{f(0)}\sum_{,j=0}^{s+1}\frac{k_j}{w_j}\int_0^\infty K(u)K\left(\frac{u}{w_j}\right)du + o(1)\right), & x=0 \\ \frac{1}{nh}\left(\frac{4m^2(x)+\sigma^2}{f(x)}\lambda_0 + o(1)\right), & x>0 \end{cases} \tag{22}$$

□

As in the case of bias, an identical expression for the variance of $\hat{m}_H^-(x)$ can be obtained with the only change being that $x \le 0$. The next theorem gives asymptotic normality of $\hat{m}_H^+(x)$ for $x \ge 0$.

Theorem 5. *Suppose that K satisfies Assumption 1, $f \in C_b^4(\mathbb{R})$ and $m^+ : [0,\infty) \to \mathbb{R}$ is such that $m^+ \in C_b^4([0,\infty))$. If $E\left(|(Y_i - m^+(X_i))d_i|^{2+\delta}|X\right) < \infty$. Then, for $x \ge 0$,*

$$(nh)^{\frac{1}{2}}\left(\hat{m}_H^+(x) - m^+(x) - \left[\frac{1}{2}m^{+(2)}(x) + \frac{m^{+(1)}(x)f^{(1)}(x)}{f(x)}\kappa_2\right]h^2\right) \xrightarrow{d} N(0, c/f^2(x))$$

where $c = \begin{cases} \sigma^2 f(x)\int_{\mathbb{R}}K^2(u)du, & \text{if } x>0 \\ \sigma^2 f(0)\int_0^\infty\left[\sum_{i=0}^{s+1}\frac{k_i}{w_i}K\left(\frac{u}{w_i}\right)\right]^2 du, & \text{if } x=0. \end{cases}$

Proof. Let $w_0 = -1$, $k_0 = -1$, $u_i = \sum_{j=0}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{X_i}{w_j}+x}{h}\right)$, and $K(\frac{X_i - x}{h}) = K_i$. Then, $\hat{m}_H^+(x) = \frac{\hat{g}(x)}{\hat{f}(x)} = \frac{\sum_{i=1}^n u_i Y_i d_i}{\sum_{i=1}^n K_i}$. $\hat{f}(x) \xrightarrow{p} f(x)$, thus we are concerned with the convergence in distribution of $\hat{g}(x)$. Note that $E(\hat{g}(x)|X_1, \cdots, X_n) = (nh)^{-1}\sum_{i=1}^n u_i m^+(X_i)d_i$ and $\hat{g}(x) - E(\hat{g}(x)|X_1, \cdots, X_n) = (nh)^{-1}\sum_{i=1}^n u_i(Y_i - m^+(X_i))d_i$. Let $Z_{in} = \frac{u_i(Y_i - m^+(X_i))d_i}{nh}$ and note that $E(Z_{in}) = 0$ and

$$V(Z_{in}) = E(Z_{in}^2) = \frac{\sigma^2}{(nh)^2}E(u_i^2 d_i) = \frac{\sigma^2}{(nh)^2}\int_0^\infty\left[\sum_{j=0}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{\alpha}{w_j}+x}{h}\right)\right]^2 f(\alpha)d\alpha.$$

Now, let $S_n^2 = \sum_{i=1}^n E(Z_{in}^2) = \frac{1}{nh}\frac{\sigma^2}{h}\int_0^\infty \left[\sum_{j=0}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{\alpha}{w_j}+x}{h}\right)\right]^2 f(\alpha)d\alpha$ and

$$X_{in} = \frac{Z_{in}}{S_n} = \frac{u_i(Y_i - m^+(X_i))d_i}{(nh)^{\frac{1}{2}}\left(\frac{\sigma^2}{h}\int_0^\infty \left[\sum_{j=0}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{\alpha}{w_j}+x}{h}\right)\right]^2 f(\alpha)d\alpha\right)^{\frac{1}{2}}}.$$

Consequently, $\sum_{i=1}^n X_{in} = 1$ and by Liapounov's Central Limit Theorem $\sum_{i=1}^n X_{in} \xrightarrow{d} N(0,1)$ provided that $lim_{n\to\infty}\sum_{i=1}^n E(|X_{in}|^{2+\delta}) = 0$ for some $\delta > 0$. Note that $|X_{in}| = \frac{|u_i(Y_j - m^+(X_j))d_i|}{(nh)^{\frac{1}{2}}c(n)^{\frac{1}{2}}}$ with $c(n) = \frac{\sigma^2}{h}\int_0^\infty \left[\sum_{j=0}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{\alpha}{w_j}+x}{h}\right)\right]^2 f(\alpha)d\alpha$ and $|X_{in}|^{2+\delta} = \frac{|u_i(Y_j - m^+(X_j))d_i|^{2+\delta}}{(nh)^{\frac{2+\delta}{2}}c(n)^{\frac{2+\delta}{2}}}$. $c(n)$ is non-stochastic, therefore

$$E(|X_{in}|^{2+\delta}) = (nhc(n))^{-1-\frac{\delta}{2}}E(|u_i|^{2+\delta}\left|\left[Y_i - m^+(X_i)\right]d_i\right|^{2+\delta})$$

and $\sum_{i=1}^n E(|X_{in}|^{2+\delta}) = (nhc(n))^{-1-\frac{\delta}{2}}\sum_{i=1}^n E(|u_i|^{2+\delta}\left|[Y_i - m^+(X_i)]d_i\right|^{2+\delta})$. Now, if

$$E\left(\left|\left[Y_i - m^+(X_i)\right]d_i\right|^{2+\delta}|X_i\right) < C < \infty$$

then

$$\begin{aligned} E(|u_i|^{2+\delta}\left|\left[Y_i - m^+(X_i)\right]d_i\right|^{2+\delta}) &= E\left[|u_id_i|^{2+\delta}E\left(\left|\left[Y_i - m^+(X_i)\right]\right|^{2+\delta}|X_i\right)\right] \\ &\le C\int_0^\infty \left[\sum_{j=0}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{\alpha}{w_j}+x}{h}\right)\right]^{2+\delta} f(\alpha)d\alpha \end{aligned}$$

Consequently,

$$\sum_{i=1}^n E(|X_{in}|^{2+\delta}) \le (nh)^{-\frac{\delta}{2}}(c(n))^{-1-\frac{\delta}{2}}C\frac{1}{h}\int_0^\infty \left|\sum_{j=0}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{\alpha}{w_j}+x}{h}\right)\right|^{2+\delta} f_X(\alpha)d\alpha.$$

Note that $c(n) = nT_2$ in Theorem 2, thus we have for $x > 0$, $c(n) \to \sigma^2 f(x)\int_{\mathbb{R}} K^2(u)du$ from (11). For $x = 0$, $c(n) \to \sigma^2 f(0)\int_0^\infty \left[\sum_{i=0}^{s+1}\frac{k_i}{w_i}K\left(\frac{u}{w_i}\right)\right]^2 du$ from (15). By the c_r-Inequality

$$\begin{aligned} E\left|\sum_{j=0}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{\alpha}{w_j}+x}{h}\right)\right|^{2+\delta} &\le (s+2)^{1+\delta}\sum_{j=0}^{s+1}E\left|\frac{k_j}{w_j}K\left(\frac{\frac{\alpha}{w_j}+x}{h}\right)\right|^{2+\delta} = (s+2)^{1+\delta}h\sum_{j=0}^{s+1}\frac{1}{h}E\left|\frac{k_j}{w_j}K\left(\frac{\frac{\alpha}{w_j}+x}{h}\right)\right|^{2+\delta} \\ &= (s+2)^{1+\delta}h\sum_{j=0}^{s+1}\frac{1}{h}\int_0^\infty \left|\frac{k_j}{w_j}K\left(\frac{\frac{\alpha}{w_j}+x}{h}\right)\right|^{2+\delta} f_X(\alpha)d\alpha \end{aligned}$$

Changing variable by setting $u = \frac{\frac{\alpha}{w_j}+x}{h}$,

$$\frac{1}{h}E\left|\frac{k_j}{w_j}K\left(\frac{\frac{\alpha}{w_j}+x}{h}\right)\right|^{2+\delta} = |k_j|^{2+\delta}\int_{\frac{x}{h}}^\infty |K(u)|^{2+\delta} f(w_j(hu-x))du.$$

For $x > 0$, since f is bounded, K satisfies assumption 1

$$\int_{\frac{x}{h}}^{\infty} |K(u)|^{2+\delta} f(w_j(hu-x))du \leq C \int_{\frac{x}{h}}^{\infty} |K(u)|^{2+\delta} du \leq \epsilon$$

for sufficiently small h. For $x = 0$, and $C > 0$

$$\begin{aligned}
\left| \int_0^{\infty} |K(u)|^{2+\delta} f(w_j hu)du - f(0) \int_0^{\infty} |K(u)|^{2+\delta} du \right| &= \left| \int_0^{C} |K(u)|^{2+\delta} [f(w_j hu) - f(0)] du \right. \\
&\quad \left. + \int_C^{\infty} |K(u)|^{2+\delta} [f(w_j hu) - f(0)] du \right| \\
&\leq \bar{p}(w_j hC, 0) \int_0^{C} |K(u)|^{2+\delta} du + 2sup(f) \int_C^{\infty} |K(u)|^{2+\delta} du \\
&\leq \epsilon
\end{aligned}$$

for sufficiently small h. Now, given that $\int_0^{\infty} |K(u)|^{2+\delta} f(w_j hu)du \to f(0) \int_0^{\infty} |K(u)|^{2+\delta} du$ since $nh_n \to \infty$ we have that $lim_{n\to\infty} \sum_{i=1}^{n} E(|X_{in}|^{2+\delta}) = 0$. Hence,

$$\frac{(nh)^{\frac{1}{2}}(\hat{g}(x) - E(\hat{g}(x)|X_1, \cdots, X_n)}{c(n)^{\frac{1}{2}}} \xrightarrow{d} N(0,1),$$

which implies that $(nh_n)^{\frac{1}{2}} (\hat{g}(x) - E(\hat{g}(x)|X_1, \cdots, X_n) \xrightarrow{d} N(0, c)$ where

$$c = \lim_{n\to\infty} c(n) = \begin{cases} \sigma^2 f(x) \int_{\mathbb{R}} K^2(u)du, & x > 0 \\ \sigma^2 f(0) \int_0^{\infty} \left[\sum_{i=0}^{s+1} \frac{k_i}{w_i} K\left(\frac{u}{w_i}\right)\right]^2 du, & x = 0 \end{cases}.$$

It follows immediately that

$$\sqrt{nh} \left\{ \left(\hat{m}_H^+(x)) - m^+(x)\right) - \left[\frac{1}{2} m^{+(2)}(x) + \frac{m^{+(1)}(x) f^{(1)}(x)}{f(x)} \kappa_2\right] h^2 - O_p(h^4 + (nh)^{-1}) \right\} \xrightarrow{d} N(0, c/f^2(x)).$$

□

An optimal bandwidth can be obtained by minimizing asymptotic weighted mean integrated squared error (AWMISE). We will define an optimal h_{pi} for $\hat{m}_H^+(0)$ which is obtained by minimizing

$$AWMISE = \int_0^{\infty} \left\{ \left[\frac{1}{2}\left(f(x)m^{+(2)}(x) + 2m^{+(1)}(x)f^{(1)}(x)\right)\kappa_2 h^2\right]^2 + \frac{1}{nh}\left[\left(4m^{+2}(x) + \sigma^2\right)\lambda_0\right] \right\} dx + s.o.$$

where $\kappa_i = \int_{-\infty}^{\infty} K(u)u^i du$; $\lambda_i = \int_{-\infty}^{\infty} K^2(u)u^i du$, for $i = 0, 1, 2...$ a full kernel for $x > 0$, and $s.o.$ denotes terms of smaller orders. Then, routine optimization gives

$$h_{pi} = \left(\frac{\lambda_0}{2\kappa_2 n}\right)^{\frac{1}{5}} \left\{ \frac{\int_0^{\infty} \left(4m^{+2}(x) + \sigma^2\right) dx}{\int_0^{\infty} \left[f(x)m^{+(2)}(x) + 2m^{+(1)}(x)f^{(1)}(x)\right] dx} \right\}^{\frac{1}{5}} = n^{-\frac{1}{5}} C$$

where $C = \left(\frac{\lambda_0}{2\kappa_2}\right)^{\frac{1}{5}} \left\{ \frac{\int_0^{\infty}\left(4m^{+2}(x)+\sigma^2\right)dx}{\int_0^{\infty}\left[f(x)m^{+(2)}(x)+2m^{+(1)}(x)f^{(1)}(x)\right]dx} \right\}^{\frac{1}{5}}$.

2.3 A comparison with local linear (LL) estimators

Local linear (LL) estimators are the most commonly used estimators for nonparametric regression in RD models. Estimation is normally conducted by selecting a uniform kernel K and a bandwidth that in effect constrains the estimation to subsamples of $\{Y_i, X_i\}_{i=1}^n$ to the right ($X_i \geq 0$) and to the left ($X_i < 0$) of the point of discontinuity $x = 0$. Hence, two local linear estimators are obtained $\hat{m}_{LL}^+(x)$ and $\hat{m}_{LL}^-(x)$. Letting $Z_i(x) = \begin{pmatrix} 1 & X_i - x \end{pmatrix}$, $Z(x)' = \begin{pmatrix} Z_1(x)' & \ldots & Z_n(x)' \end{pmatrix}$, $K_{ix} = K(\frac{X_i - x}{h})$, $K(x) = diag\{K_{ix}\}_{i=1}^n$, and $Y' = \begin{pmatrix} Y_1 & \cdots & Y_n \end{pmatrix}$, only the observations $\{(X_i, Y_i) : X_i \geq 0\}_{i=1}^n$ are used to to estimate $m^+(x)$ for $x \geq 0$. Observations $\{(X_i, Y_i) : X_i < 0\}_{i=1}^n$ are used to estimate $m^-(x)$ where $x < 0$. For $x \geq 0$, the local linear estimator is given by

$$\hat{m}_{LL}^+(x) = \begin{pmatrix} 1 & 0 \end{pmatrix} \left(Z(x)^{'} K(x) Z(x) \right)^{-1} Z(x)^{'} K(x) Y \tag{23}$$

where X_i takes values in $[0, +\infty)$. Similarly,

$$\hat{m}_{LL}^-(x) = \begin{pmatrix} 1 & 0 \end{pmatrix} \left(Z(x)^{'} K(x) Z(x) \right)^{-1} Z(x)^{'} K(x) Y, \tag{24}$$

in which X_i takes values in $(-\infty, 0]$. Expressions for the conditional bias and variance of $\hat{m}_{LL}^+(x)$ at boundary points were obtained by Fan and Gijbels (1992) and are given by

$$E\left(\hat{m}_{LL}^+(x) - m^+(x) | X_1, \cdots, X_n\right) = \frac{h^2}{2} m^{(2)}(x) \left(\frac{\mu_{2,\frac{x}{h}}^2 - \mu_{1,\frac{x}{h}} \mu_{3,\frac{x}{h}}}{\mu_{0,\frac{x}{h}} \mu_{2,\frac{x}{h}} - \mu_{1,\frac{x}{h}}^2} \right) + o_p(h^2)$$

and

$$V\left(\hat{m}_{LL}^+(x) | X_1, \cdots, X_n\right) = \frac{\mu_{2,\frac{x}{h}}^2 \eta_{0,\frac{x}{h}} - 2\mu_{2,\frac{x}{h}} \mu_{1,\frac{x}{h}} \eta_{1,\frac{x}{h}} + \mu_{1,\frac{x}{h}}^2 \eta_{2,\frac{x}{h}}}{\left(\mu_{0,\frac{x}{h}} \mu_{2,\frac{x}{h}} - \mu_{2,\frac{x}{h}}^2 \right)^2} \frac{\sigma^2}{nhf(x)} + o_p((nh)^{-1}).$$

Compared to the bias of LL estimators at interior points, given by $E(\hat{m}_{LL}^+(x) | X_1, \cdots, X_n) - m^+(x) = \frac{h^2}{2} m^{+(2)}(x) \kappa_2 + o_p(h^2)$, we see that their leading terms have the same order h^2 but different magnitude.[2] One way to interpret this is that LL estimators adapt to the boundary by adjusting a regular kernel to an effective kernel, substituting κ_2 with

$$\left(\frac{\mu_{2,\frac{x}{h}}^2 - \mu_{1,\frac{x}{h}} \mu_{3,\frac{x}{h}}}{\mu_{0,\frac{x}{h}} \mu_{2,\frac{x}{h}} - \mu_{1,\frac{x}{h}}^2} \right).$$

Comparing the bias of the Hestenes estimators at the boundary in equation (20) with that of the LL estimator we can see that both have leading terms of the same order h^2 but different magnitudes. The bias size of Hestenes-based estimator depends on the chosen coefficients of Hestenes' extension, whereas the bias size of LL estimators is impacted by the partial kernels at the boundary point.

[2] Expressions for the unconditional bias and variance of the LL estimator when the regressors take values in $\mathbb{R}$ were given by Fan (1993). In particular, he finds that $E(\hat{m}_{LL}(x)) - m(x) = \frac{h^2}{2} m^{(2)}(x) \kappa_2 + o(h^2)$ and $V(\hat{m}_{LL}(x)) = \frac{\sigma^2}{nh} f^{-1}(x) \lambda_0 + o((nh)^{-1})$.

An asymptotic approximation for the conditional MSE of $\hat{m}^+_{LL}(x)$ can easily be obtained as is given by

$$MSE(\hat{m}^+_{LL}(x)|X_1,\cdots X_n) = \frac{h^4}{4}\left(m^{+(2)}(x)\right)^2\left(\frac{\mu^2_{2,\frac{x}{h}}-\mu_{1,\frac{x}{h}}\mu_{3,\frac{x}{h}}}{\mu_{0,\frac{x}{h}}\mu_{2,\frac{x}{h}}-\mu^2_{1,\frac{x}{h}}}\right)^2 + \frac{\mu^2_{2,\frac{x}{h}}\eta_{0,\frac{x}{h}}-2\mu_{2,\frac{x}{h}}\mu_{1,\frac{x}{h}}\eta_{1,\frac{x}{h}}+\mu^2_{1,\frac{x}{h}}\eta_{2,\frac{x}{h}}}{\left(\mu_{0,\frac{x}{h}}\mu_{2,\frac{x}{h}}-\mu^2_{2,\frac{x}{h}}\right)^2}\frac{\sigma^2}{nhf(x)}+o_p\left(h^4+\frac{1}{nh}\right)$$

and an optimal bandwidth h_{pi} for $\hat{m}^+_{LL}$ can be obtained by minimizing the leading terms in this expression,

$$h_{pi} = n^{-\frac{1}{5}}t_{1n}^{\frac{1}{5}}t_{1n}^{-\frac{2}{5}}$$

where $t_{1n} = \left(m^{(2)}\right)^2\left(\frac{\mu^2_{2,\frac{x}{h}}-\mu_{1,\frac{x}{h}}\mu_{3,\frac{x}{h}}}{\mu_{0,\frac{x}{h}}\mu_{2,\frac{x}{h}}-\mu^2_{1,\frac{x}{h}}}\right)^2$ and $t_{2n} = \frac{\mu^2_{2,\frac{x}{h}}\eta_{0,\frac{x}{h}}-2\mu_{2,\frac{x}{h}}\mu_{1,\frac{x}{h}}\eta_{1,\frac{x}{h}}+\mu^2_{1,\frac{x}{h}}\eta_{2,\frac{x}{h}}}{\left(\mu_{0,\frac{x}{h}}\mu_{2,\frac{x}{h}}-\mu^2_{2,\frac{x}{h}}\right)^2}\frac{\sigma^2}{f(x)}$. Although a direct comparison the bias and variance expression for the Hestenes-based and LL estimators is made difficult by the complexity of these expressions, our simulations will provide additional evidence on their relative magnitudes.

3 Estimators for a jump discontinuity

The Hestenes-based estimators $\hat{m}^+_H(0)$ and $\hat{m}^-_H(0)$ can be used to estimate the jump at $x = 0$, denote by $J(0)$ by $\hat{J}_H(0) = \hat{m}^+_H(0) - \hat{m}^-_H(0)$. Hahn et al. (2001) establishes the identification of the RD model and uses the jump discontinuity of the expected outcome at that point to measure an average treatment effect. If Y is the outcome variable and let X is the running variable, when $X \in \mathbb{R}$ is above a threshold $x = 0$, the individual gets the treatment and $D = 1$, otherwise the individual does not get the treatment and $D = 0$. The regression jump is

$$J(0) = \frac{\lim_{x\downarrow 0} m^+(x) - \lim_{x\uparrow 0} m^-(x)}{\lim_{x\downarrow c} E(D|X=x) - \lim_{x\uparrow c} E(D|X=x)} \tag{25}$$

and, in particular, for a sharp RD design

$$J(0) = \lim_{x\downarrow 0} m^+(x) - \lim_{x\uparrow 0} m^-(x)$$

because $\lim_{x\downarrow c} E(D|X=x) - \lim_{x\uparrow c} E(D|X=x) = 1$. An estimator for $J(0)$

$$\hat{J}_H(0) = \lim_{x\downarrow 0}\hat{m}^+_H(x) - \lim_{x\uparrow 0}\hat{m}^-_H(x) = \hat{m}^+_H(0) - \hat{m}^-_H(0) \tag{26}$$

where $\hat{m}^+_H(0) = \lim_{x\downarrow 0}\hat{m}^+_H(x)$ and $\hat{m}^-_H(0) = \lim_{x\uparrow 0}\hat{m}^-_H(x)$. From Theorem 3, we get the unconditional biases of $\hat{m}^+_H$ and $\hat{m}^-_H$,

$$E(\hat{m}^+_H(0)) - m^+(0) = \left[\frac{1}{2}m^{+(2)}(0) + \frac{m^{+(1)}(0)f^{(1)}(0)}{f(0)}\right]\kappa_2 h^2 + O(h^4 + (nh)^{-1})$$

and

$$E(\hat{m}^-_H(0)) - m^-(0) = \left[\frac{1}{2}m^{-(2)}(0) + \frac{m^{-(1)}(0)f^{(1)}(0)}{f(0)}\right]\kappa_2 h^2 + O(h^4 + (nh)^{-1}).$$

It follows immediately that $E(\hat{J}_H(0)) - J(0) = B_n(0) + O(h^4 + (nh)^{-1})$, where

$$B_n(0) = \left(\frac{1}{2}(m^{+(2)}(0) - m^{-(2)}(0)) + \left(\frac{m^{+(1)}(0)f^{(1)}(0)}{f(0)} - \frac{m^{-(1)}(0)f^{(1)}(0)}{f(0)}\right)\kappa_2\right)h^2$$

In addition, from Theorem 5, we have

$$\sqrt{nh}\left(\hat{m}_H^+(0) - m^+(0) - \left[\frac{1}{2}m^{+(2)}(0) + \frac{m^{+(1)}(0)f^{(1)}(0)}{f(0)}\kappa_2\right]h^2\right) \xrightarrow{d} N\left(0, cf^{-2}(0)\right)$$

where $c = \sigma^2 \int_0^\infty \left[\sum_{i=0}^{s+1} \frac{k_i}{w_i} K\left(\frac{u}{w_i}\right)\right]^2 du$ and an equivalent expression holding for for $\hat{m}_H^-(0)$.

Consequently, we obtain the asymptotic distribution of $\hat{J}_H(0)$ as

$$\sqrt{nh}(\hat{J}_H(0) - J(0) - B_n) \xrightarrow{d} N\left(0, 2cf^{-2}(0)\right).$$

4 Simulations

in this section we compare the finite sample performance of the Hestenes (H), Nadaraya-Watson (NW) and local linear (LL) estimators. Our simulation investigates the sources of bias at discontinuity points by exploring three scenarios related to the H and NW estimators: infeasible estimators using the true density function f, feasible estimators using either the whole sample or partial samples (one sided) for the estimation of f. In addition, we examine how the performance of these estimators is affected by considering different regressions, densities for the regressors, method of obtaining the bandwidth or choice of kernel functions. Although estimates at interior points are considered, the primary focus of our comparisons is the estimates at the point of discontinuity.

Comparisons are based on the evaluation of bias, standard deviation (SD), and root mean squared error (RMSE) for the discontinuity point as well as average mean squared error (AMSE) for all points including interior points and the boundary point.

Let $x = 0$ be the discontinuity point, $\hat{m}(x)$ be an estimate for the regression function $m(x)$ at x, and $\hat{J}(0)$ be an estimate for the jump $J(0)$. We generate M samples calculate these statistics ofor $\hat{m}(x)$ and $\hat{J}(0)$

$$Bias(\hat{\theta}) = \frac{\sum_{m=1}^M \left(\hat{\theta}_m - \theta\right)}{M},\ SD(\hat{\theta}) = \sqrt{\frac{\sum_{m=1}^M \left(\hat{\theta}_m - \bar{\hat{\theta}}\right)^2}{M-1}},\ RMSE(\hat{\theta}) = \sqrt{\frac{\sum_{m=1}^M \left(\hat{\theta}_m - \theta\right)^2}{M}}$$

where $\bar{\hat{\theta}} = \frac{1}{M}\sum_{m=1}^M \hat{\theta}_m$, θ denotes $J(0)$, $m^+(0)$, or $m^-(0)$, and $\hat{\theta}$ denotes $\hat{J}(0)$, $\hat{m}^+(0)$ or $\hat{m}^-(0)$. We also calculate

$$R_{RMSE}(\hat{\theta}) = \frac{RMSE(\hat{\theta})}{\min\limits_{\hat{\theta}} RMSE(\hat{\theta})}.$$

We calculate these statistics of $\hat{m}(x)$ at all points of evaluation of the regression. For each sample, and each estimator, we calculate a root average squared error ($RASE$) over K evaluation points,

$$RASE_{\hat{m},m} = \sqrt{\frac{\sum_{k=1}^K (\hat{m}(x_k) - m(x_k))^2}{K}} \text{ for sample } m$$

we then calculate an average of $RASE_{\hat{m},m}$ across all generated samples,

$$AMSE_{\hat{m}} = \frac{\sum_{m=1}^{M} RASE_{\hat{m},m}}{M}.$$

We also calculate R_{AMSE} as $R_{AMSE_{\hat{m}}} = \frac{AMSE_{\hat{m}}}{\min_{\hat{m}} AMSE_{\hat{m}}}$. We compare three types of estimators: NW estimators, LL estimators, and Hestenes estimators. We denote Hestenes estimators as HXX, such as H00, H10, H11, H20, or H21, where the first digit stands for the degree of smoothness of the composite function $\mu(x)$, $s = 0, 1, 2$, and the second digit denotes which sequence of w_i is used: $b = 0$ means the sequence $w_i = 1/i$ is used while $b = 1$ means the sequence $w_i = i$ is used, where $i = 1, 2, \cdots, n$.

4.1 Estimating different types of jumps

We start by studying how estimators behave when the true regression functions have different types of jumps at the discontinuity point. For instance, regression functions that have jumps or drops from a concave function to a convex function or vice versa. We consider the following four regression functions:

$$m(x) = \begin{cases} (x+1)^2, & x < 0 \\ \sin(2\pi x + 0.1\pi) & x \geq 0 \end{cases} \tag{27}$$

$$m(x) = \begin{cases} \sin(2\pi x + 0.1\pi), & x < 0 \\ -(x-1)^2 + 2 & x \geq 0 \end{cases} \tag{28}$$

$$m(x) = \begin{cases} -(x-1)^2 + 2, & x < 0 \\ \frac{1}{x+1} - 1 & x \geq 0 \end{cases} \tag{29}$$

$$m(x) = \begin{cases} (x-1)^2, & x < 0 \\ \frac{1}{x+1} - 1 & x \geq 0 \end{cases}. \tag{30}$$

We construct $\hat{m}_{NW}$, the NW estimators, using all observations to the right of the discontinuity point to estimate the regression function on the right and use all observations to the left of the discontinuity point to estimate the regression function on the left, and then calculate the jump at the discontinuity point.

For LL estimators, we use Fan (1992)'s modified version of the LL estimator to avoid singularities.[3] We construct $\hat{m}_{HXX}$, the H estimators, according to the expressions given in section 2, where we use the entire sample to estimate f so that we can take advantage of the assumption of RDD that the density function f is continuous.

Figure 1 shows four regression functions and four estimators, NW, LL, H10, and H21, that approximate the true regression functions. We choose two H estimators with different degrees of smoothness in sewing conditions and different series of w to show variations in the H-estimators. As expected, estimators differ

[3]As proved by Fan (1992), this estimators has the same asymptotic properties as the regular LL estimator described in equations (23) and (24), so we can use properties derived from the latter estimators for comparison.

mostly at the vicinity of the discontinuity point. NW estimators have a significant bias while the rest of the estimators stay close to each other and are close to the true regression functions at the boundary. There is no visual discrepancy between the two H-estimators.

4.2 Performance at the discontinuity point

Graphics can give us an intuitive impression of estimator's behavior, but to precisely evaluate performance across estimators, we rely on large sample simulations to show the distribution of the estimators. In the following simulations, we compare the performance of three types of estimators in approximating the four regression functions using 10000 repeated samples with a sample size of 2000.

To investigate the sources of bias at the boundary point, we explore three scenarios with respect to the NW and H estimators. First, we construct infeasible estimators with the true density function f in the denominator. Then, we construct feasible estimators with the whole sample to estimate f. Lastly, we construct feasible estimators using partial samples (one sided) for the estimation of f: using only observations to the right of the discontinuity point to estimate $f(0+)$ and only observations to the left of the discontinuity point to estimate $f(0-)$. For all three scenarios, LL estimators are constructed in the same way as described in section 4.1. In generating samples, we let X have a standard normal distribution with the peak of the density occurring at $x = 0$, and use the four regression functions described above to generate Y with standard normal error terms. In constructing estimators, we use Gaussian kernels and obtain optimal bandwidth h through plug-in methods.

Table 1 shows the results of the first scenario. We construct infeasible NW estimators m_{NW} and infeasible H estimators $\tilde{m}_{HXX}$, where the true density function f is used in the denominator. The first part of the results shows bias, variance, RMSE, and R_{RMSE} of estimates of the jump at the discontinuity point and AMSE and R_{AMSE} of estimates of the regression function at all evaluation points. To explore the estimate of the jump, we examine the estimates of two regressions to the right and left of the discontinuity point in the second part of the report. We show bias and variance both from samples and quantified results from our theory. It is not surprising that the bias and variance of the jump is the sum of the bias and variance of estimates on regressions from two sides. Most importantly, estimates from samples match the asymptotic bias and variance from our theory very well, which further confirms our theoretical findings. It is evident that NW estimators have the largest bias at the discontinuity point. In most cases, we can always find one set of Hestenes estimators that have smaller bias than LL estimators.

Table 2 shows the results of the second scenario. We construct feasible NW estimators as $\hat{m}^{+}_{NW}(x) = \frac{(nh)^{-1}\sum_{i=1}^{n} K(\frac{X_i - x}{h}) Y_i d_i}{(nh)^{-1}\sum_{i=1}^{n} K(\frac{X_i - x}{h})}$ for $x \geq 0$ and $\hat{m}^{-}_{NW}(x) = \frac{(nh)^{-1}\sum_{i=1}^{n} K(\frac{X_i - x}{h}) Y_i (1-d_i)}{(nh)^{-1}\sum_{i=1}^{n} K(\frac{X_i - x}{h})}$ for $x < 0$, and feasible H estimators as defined in section 2. The results are similar to those in the first scenario. What is worth mentioning is that the NW estimators appear to have the smallest bias on the jump, but this conclusion does not withstand further scrutiny. As shown in part two, the NW estimators on regressions have a large bias. Not surprisingly, when we switch to using only half of the sample to estimate f , the large bias of NW estimators on the jump occurs again.

Table 3 shows the results of the third scenario. We construct feasible NW estimators where the density estimator in the denominator uses only data corresponding $X_i \geq 0$ and $X_i < 0$ for $\hat{m}^+_{NW}$ and $\hat{m}^-_{NW}$, respectively. Similarly, for H estimators we consider

$$\hat{m}^+_H(x) = \frac{\frac{1}{nh}\sum_{i=1}^n \left[K\left(\frac{X_i - x}{h}\right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K\left(\frac{\frac{X_i}{w_j} + x}{h}\right)\right] Y_i d_i}{\frac{1}{nh}\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) d_i}$$

for $x \geq 0$ and

$$\hat{m}^-_H(x) = \frac{(nh)^{-1}\sum_{i=1}^n K(\frac{X_i - x}{h}) Y_i \left(1 - d_i\right)}{(nh)^{-1}\sum_{i=1}^n K(\frac{X_i - x}{h}) \left(1 - d_i\right)}$$

for $x < 0$, which we construct here solely for comparison with these NW estimators. The simulation outcomes confirm the earlier notion: H estimators constructed in this way have large bias. Compared to the last table, the bias of NW estimators depends on the regression function. When the regression function has a large intersection away from zero, the NW estimators constructed with half of the sample have a smaller bias; when the regression function has a steep slope, the NW estimators constructed with the whole sample have a smaller bias, which accords with our theory.

From our theory, we know that the bias of NW estimators comes from the incomplete kernel at the boundary points. Therefore, when the whole sample is used to estimate f, the leading term of bias is a constant term. When half of the sample is used to estimate f, the leading term of bias is a term of order h. As is often described in the literature (e.g.,Imbens and Lemieux (2008), Porter (2003)), NW estimators are constructed to use half of the sample to estimate density f and the bias is described by a term of order h. This statement is not wrong but it understates two drawbacks of this method: one, compared to the method of using the whole sample to estimate f, this estimator has larger variance; two, it contains the h term twice as large as the former method when regression m has a slope and symmetric kernels are used – both situations are common. A large sample size does not help to alleviate the problem here because h converges to 0 slowly: the optimal h yielded by cross-validation and plug-in methods is $h = n^{-\frac{1}{5}}$. H estimators fix the problem by making the kernels on the numerator complete, which reduces bias from the constant term – the h term – at the same time, using the whole sample to estimate f in the denominator to avoid introducing a fraction at the bottom which amplifies bias on the top.

Table 4 shows the performance of these estimators by focusing on one regression function, regression (28) and increasing the sample size from 1000 to 4000 by 1000 units. As sample size increases, the bias and variance of each of the estimators becomes smaller, but they converge at a different speed. NW estimators converge slower than LL and H estimators, which agrees with the theoretical result that the bias of NW estimators decreases in the order of h while LL and H estimators decrease in the order of h^2.

4.3 Choosing bandwidth and kernels

Choosing bandwidth is critical in non-parametric estimation. In the simulation, we use plug-in and cross-validation methods to obtain bandwidths. For the plug-in method, we choose locally optimal bandwidth by optimizing the leading terms of mean squared error of the regression estimator at the boundary point; for the cross-validation method, we choose a globally optimal bandwidth by minimizing the sum of the squared deviations between the observed outcomes and the regression estimates. In both methods, we take advantage of the fact that we know the true function of the density and regression. For most simulations, we use plug-in methods to find optimal bandwidth because this method is fast while the cross-validation method is computationally time consuming. Using the plug-in method, h is calculated by the formulas described in section 2 for the LL and H estimators. Table 5 shows that using different methods of the bandwidth does not change results.

We compare two types of kernels: kernels without compact support, such as Gaussian kernels, and kernels with compact support, such as Epanechnikov Kernels. Table 6 shows kernels make little impacts on the performance of estimators.

4.4 Changing density function f

The simulations conducted above use standard normal distributions in generating X with the peak of the density at the discontinuity point. This splits observations almost evenly between two sides of the discontinuity point, which we know in reality will not always be the case. For example, in the case of using students' class grade as the running variable to decide whether a third-year student can be promoted to the next grade or to be held back for one more year, the cutoff grade, let us say 60 out of 100, is often located away from the peak of the grade distribution, say 80 out 100. Therefore, it is useful to check how estimators detect the jump when the cutoff point is off the peak of the distribution of the running variable. Table 7 shows there are few differences in performance when the peak of the density function f is shifted away from the cutoff point.

In summary, our theory has provided good predictions for bias and variance in finite samples for H estimators in different setups. Hestenes estimators perform better than NW estimators in all cases and, in most situations, one can always find a set of H estimators that have smaller bias than LL estimators.

5 Empirical illustration

To illustrate the applicability of our estimators in empirical settings, we use data collected by Litschig and Morrison (2013) who use a RD model to examine the impact of intergovernmental transfer programs on education and poverty reduction outcomes. We begin with a discussion of the assumptions underlying RD designs and their implications for empirical modeling. We then describe how a typical empirical RD model is estimated to verify these assumptions and address some implementation issues.

5.1 Assumptions on RD designs and their implications for empirical studies

Identification of RD designs depends on several assumptions, which have important implications in empirical studies. The most important assumption is on the regression function associated with the outcome variable. Instead of assuming a specific functional form for the regression function, identification of RD models assumes the existence of a smooth regression at the vicinity of the discontinuity point. Contrast this with the difference-in-difference (DID) design where an "equal trend" is required: regression functions before and after an intervention must be the same. The reason for this difference is that RD designs assess a local average treatment effect (LATE).

The second assumption relates to how a treatment is assigned in association with a running variable. Around the discontinuity point, individuals are similar but receive different treatments based on whether their associated running variable values are slightly above or below the threshold, which determines the treatment group and control group. We assume that the jump in the regression of outcomes is actually caused by a treatment variable rather than other covariates and the running variable has a continuous density. To verify the empirical validity of these assumptions, researchers check that all other covariates across the discontinuity point are continuous to ensure that the running variable, rather than other covariates, are associated with the treatment effect. They normally follow McCrary (2008)'s recommendation to check that the density of the running variable around the discontinuity point to ensure that no individual endogenously manipulates the running variable. This also trivially satisfies an assumption that sample data exist on both sides of the discontinuity point. This is an important theoretical assumption for regression estimators that use one-sided data (NW or LL) to ensure that the denominator, which consists of a kernel density estimator, is not equal to zero.

Lastly, an important assumption for identification is knowledge of the point of discontinuity – that is, the discontinuity of treatment status when the running variable crosses the discontinuity point is known to the econometrician: in sharp RD designs, the jump on the probability of getting treated is one while in fuzzy RD designs this jump is between 0 and 1. Without this underlying assumption, the jump in the expected outcome will not be assigned to any treatment. Moreover, as shown in equation (25), the assumption ensures the denominator of the estimator is not zero.

One common feature of empirical RD models is that, instead of using all data to perform a nonparametric estimation, researchers often conduct estimation with a compactly supported kernel by arbitrarily restricting a running variable to a small range of values around the threshold, then demonstrating, as a robustness test, that the parameter of interest is not sensitive to different ranges induced by the bandwidth in the vicinity of the point of discontinuity. The rationale is that using all data could give undue influence to data far away from the threshold. However, this procedure has at least two drawbacks. First, when we have a high volume of data in a small range of discrete running variable values (often running variables are discrete, such as age, test scores, or the number of individuals), the variance of the estimate is large. Second, data away from the threshold have information that can influence the estimated regression and, therefore, affect the jump size estimate.

5.2 A typical empirical RD model estimation procedure

The running variable X is not an object of direct interest, but it is of interest insofar as the expected value of the outcome variable Y has a jump at a particular value of X. Normally, economists are interested in regression slopes, but not in this case, where the primary interest in the regression jump. At the discontinuity point, another variable D – the treatment variable – experiences a jump, and the jump in the expectation of Y is thought of as the average treatment effect (ATE) of the treatment D on the outcome Y under the assumptions discussed in the last subsection. Empirical work using RD designs often involve the the following procedures:

1. Use a scatter plot to visually check if there is a jump in the regression function of the outcome variable and treatment variable with respect to the running variable at the discontinuity point.
2. Perform regression estimation on the outcome and treatment variable.
3. Conduct a robustness test to ensure that no other covariates have a jump at the discontinuity point.
4. Conduct a test suggested by McCrary (2008) for checking that the density of the running variable is continuous at the discontinuity point.
5. Repeat estimation with a truncated sample to ensure that points far away from the discontinuity point do not exert undue influence.

Abstracting from specific empirical context or designs – sharp or fuzzy – estimations in RD models can be simply categorized into regression and density discontinuity estimation at the discontinuity point. The regression discontinuity estimation includes estimation of the regression discontinuity of the outcome variables and the treatment variable with respect to the running variable, while the density estimation includes density discontinuity of the running variable. In many studies (for example, Litschig and Morrison (2013), Matsudaira (2008)), a local least squares method is used for regression estimation and a histogram or McCrary's procedure is used for checking density discontinuity.

5.3 An empirical example

Litschig and Morrison (2013) exploit an opportunity provided by the passing of Decree 188181, a federal funds transfer plan in Brazil, which stipulates that federal funds – FPM (the federal Fundo de Participação dos Municípios) – must be distributed to local communities according to municipality population. The intergovernmental transfers jump in per capita spending at several population thresholds, which constitutes a sharp RD design with multiple thresholds. The treatment variable here is per capita spending and the running variable is the population of the municipality. The impacts on education and poverty are measured by outcome variables such as years of schooling, literacy rate, poverty rate and political party reelection rate. The study estimates the jump in the conditional means of the outcome variables using local least squares and estimate jump in the density of the running variable at the threshold using the method recommended by McCrary (2008). To show the robustness of their results, Litschig and Morrison vary the choice of bandwidth, percentage of population away from the discontinuity point, from 2%, to 3% and 4%, and try different functional forms ranging from linear, quadratic, cubic and quartic. Compared to their approach, our approach is more flexible. Instead of arbitrarily choosing the bandwidth, we choose an optimal bandwidth based on the sample; instead of trying different functional forms, we do not specify any functional form.

A direct comparison between our estimation results with Litschig and Morrison's is not possible because we choose different parameters, such as bandwidth. However, we try to produce comparable results using our estimators by considering two situations: estimate the jump in regression of the running variable with and without extra covariates. As mentioned in their paper, extra covariates are included to guard against misspecification of the model and to increase the precision of the estimates.

For estimation without extra covariates, we consider

$$Y = \begin{cases} m^+(X) + \epsilon, & X \geq 0 \\ m^-(X) + \epsilon, & X < 0 \end{cases},$$

where X is the running variable: the population of a municipality and Y represents either a treatment or the outcome variable. The jump size is estimated by $\hat{J}(0) = \lim_{x\downarrow 0} \hat{m}^+(x) - \lim_{x\uparrow 0} \hat{m}^-(x)$.

For estimation with extra continuous covariates, we propose the following additive model and use the marginal integration method to estimate regression functions.

$$Y = \begin{cases} m_1^+(X_1) + m_2(X_2) + \cdots + m_p(X_p) + \epsilon, & X_1 \geq 0 \\ m_1^-(X_1) + m_2(X_2) + \cdots + m_p(X_p) + \epsilon, & X_1 < 0 \end{cases},$$

where X_1 is the running variable and $X_2, \cdots, X_p$ are continuous covariates. Different from regular additive models, the first regression function in this model is discontinuous. To make the jump identifiable, there is no intercept term, and we assume that the regression functions other than m_1 have mean zeros, i.e., $E(m_i(X_i)) = 0$, for $i = 2, \cdots, p$.

We use two methods to obtain an optimal h: a plug-in rule of thumb and a jackknife cross-validation method. For rule of thumb, $h = n^{-1/5} std(X)$. For the jackknife cross-validation method, we choose h to minimize the sum of squared leave-one-out residuals using observations to the right and to the left of the threshold respectively.

Estimation procedure is straightforward. We use Hestenes estimators to estimate the jump of the regressions of the treatment variables and outcome variables and we use Hestenes density estimators proposed by Mynbaev and Martins-Filho (2019) to estimate the jump of the density of the running variable.

Figures 2 and 3, which correspond to Figures 4 and 5 in Litschig and Morrison (2013), show the estimation of the jump in the treatment and outcome variables. Each subgraph shows three estimates of regression: regression without covariates using optimal h by the rule-of- thumb method or by the cross-validation method, and regression with covariates using optimal h by the rule-of- thumb method. We can see from Figure 2 that, as a treatment variable, there is a clear increase in spending per capita while there isn't significant changes in other revenue and own revenue per capita. If anything, there is slightly increase in own revenue. This agrees with their explanation that the federal transfer causes an increase in per capita public spending without crowding out other and own per capita revenue. As a consequence, Figure 3 shows that there is a increase in years of schooling for the 19-28 age group, a decrease in the illiteracy rate and poverty rates, and an increase in the reelection rate of the incumbent party. These results agree with their results.

Figure 2 represents the discontinuity in regressions of the treatment variables: total spending per capital, own revenue, and other revenue, and Figure 3 graphically represents the discontinuity in regressions of outcome variables: schooling, literacy, poverty, and party reelection. Visually, we can see that our results agree with their results: the federal transfer causes an increase in per capita public spending without crowding out other and own per capita revenue, which in turn is responsible for an increase in years of schooling for the 19–28 year-old group, a decrease in the illiteracy rate and poverty rates, and an increase in the reelection rate of the incumbent party.

All jump estimates above are shown in Table 8. To compare with their results, we show their corresponding estimation results using similar bandwidth. When the rule-of-thumb method is used, a bandwidth is chosen based on the size and standard deviation of the population sample, so we have two bandwidths for all 16 regressions, one for the regression to the left and the other for the regression to the right of the threshold. When the cross-validation method is used, a bandwidth is chosen based on the residuals of the sample, so we have 16 bandwidths for 16 regressions. Although Hestenes estimation uses different bandwidth and method from the OLS estimation that Litschig and Morrison use, the estimates are amazingly similar.

Finally, we check the continuity of the density of the population running variable. Figure 6, which corresponds to their online appendix Figure 1, graphically presents estimates from Cheng (1994)'s LL density estimators and Mynbaev and Martins-Filho (2019)'s Hestenes density estimators. The two estimates almost completely overlap. Since McCrary (2008)'s jump density estimators are based on Cheng (1994)'s LL density estimators, it is not surprising that Figure 4 looks very similar to their online appendix Figure 1. Table 9 compares our density estimates with theirs. The estimates on the jump from the three estimators are slightly different, but they all lead to the acceptance of the null hypothesis that there are no discontinuities at all six cutoffs points. The discrepancy between the estimates from the LL and McCrary estimators could be explained by different bin size and bandwidth chosen for each implementation.

In summary, our estimation for both regression and density mostly agree with the results Litschig and Morrison (2013). One thing we want to point out is that they have made a mistake in implementing their empirical specification described in equation (1) in their paper. One way to run this OLS regression is to regress outcome variables with respect to all other explanatory variables which include the running variable, other pretreatment variables, and state fixed effects. Alternatively, if one wants to remove the states effects first, one can apply the Frisch-Waugh-Lovell theorem to get residuals of the outcome variables, the running variable and the pretreatment variable against the state fixed effects, then regress the residues of the outcome variable with respect to the residuals of explanatory variables. Instead, they first get the residual from the regression of the outcome variables with respect to the state fixed effects and then run the residual from the first regression against the rest of explanatory variables rather than their residuals. This mistake does not qualitatively change the result but affects the size of the jumps.

6 Conclusions

In this paper, we provided a new nonparametric estimator for regression discontinuity based on the extension proposed by Hestenes (1941). Constructed using the same algebraic structure of NW estimators, our estimators restore the bias at the boundary points to the same as that of the interior points in both order and magnitude. A theoretical comparison between our estimators and the popular local linear approach shows that these two types of estimators have the same unconditional bias order of $O(h^2)$ and variance order $O(\frac{1}{nh})$. In Monte Carlo simulations, we show that our estimators are free of boundary problems and perform better than NW estimators in all cases. Compared to LL estimators, our estimators have the same bias order and similar, or in some cases, smaller bias size. By applying our estimators to an empirical study by Litschig and Morrison (2013), we show our estimators are easy to use and provide more flexibility than standard OLS estimation procedure used for RDD.

For future study, there is an opportunity to extend our method to other estimators, such as LL estimators. Boundary problems actually exist in LL estimators, even though they do not affect the performance of LL estimators as much as they do NW estimators: the bias of LL estimators at the boundary points has the same order but different magnitude from biases at interior points. While bias order is important for asymptotic properties, the size of the bias is also of great concern for researchers who deal with finite samples in empirical studies. This means that applying Hestenes (1941)'s extension to modify LL estimators could reduce bias size for LL estimators on boundary points.

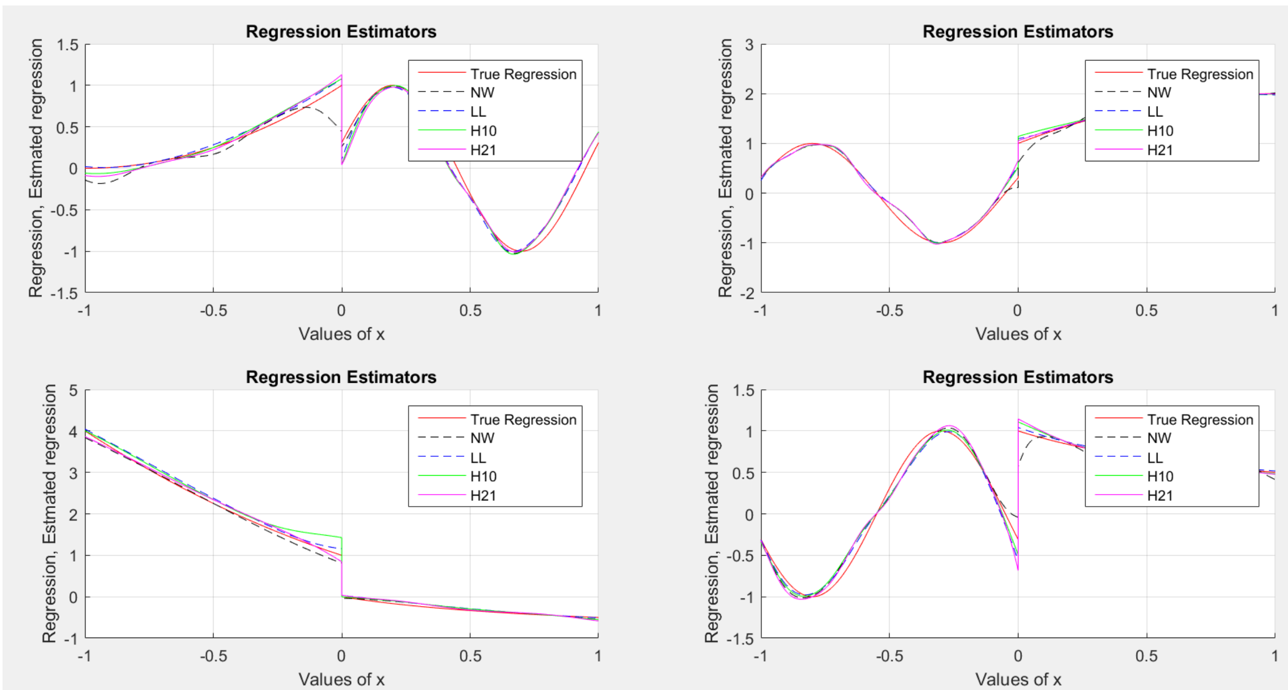

Figure 1: Four regression functions estimated by four estimators

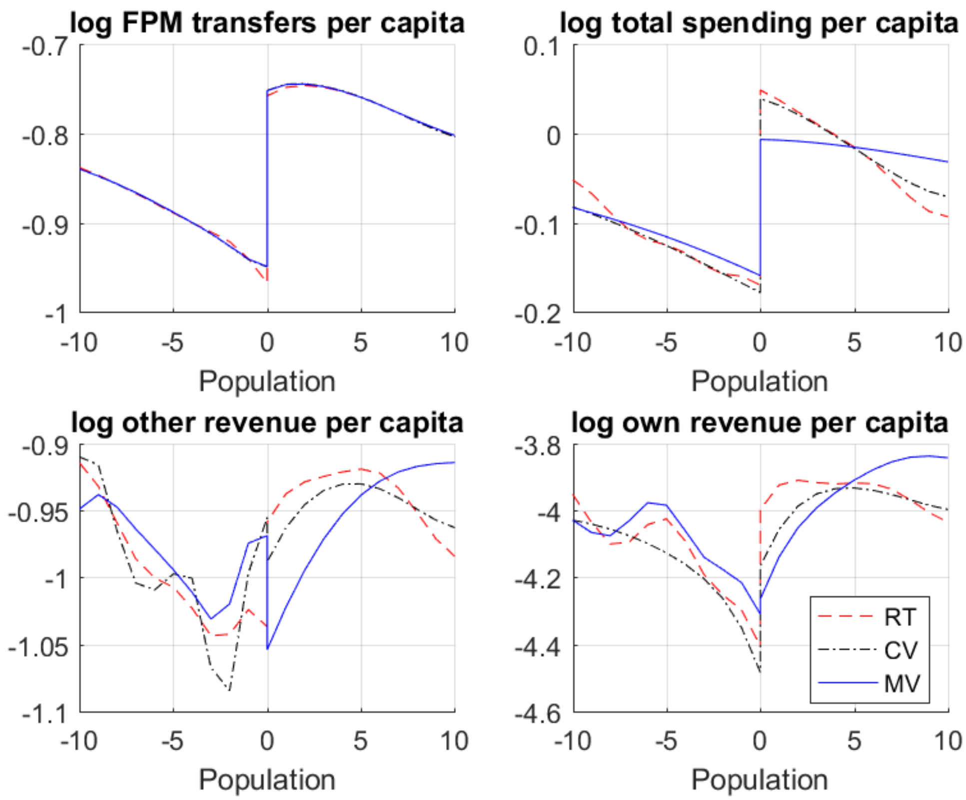

Figure 2: Impacts on Total Spending, Other Revenue, and Own Revenue

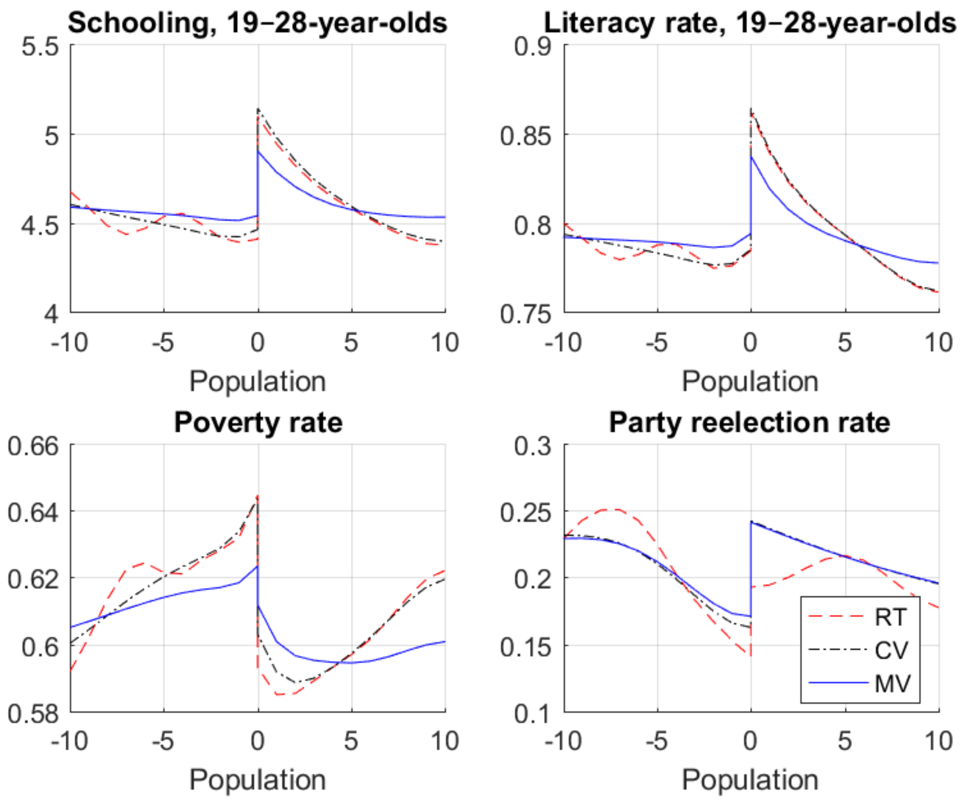

Figure 3: Impacts on Schooling, Literacy, Poverty, and Party Reelection

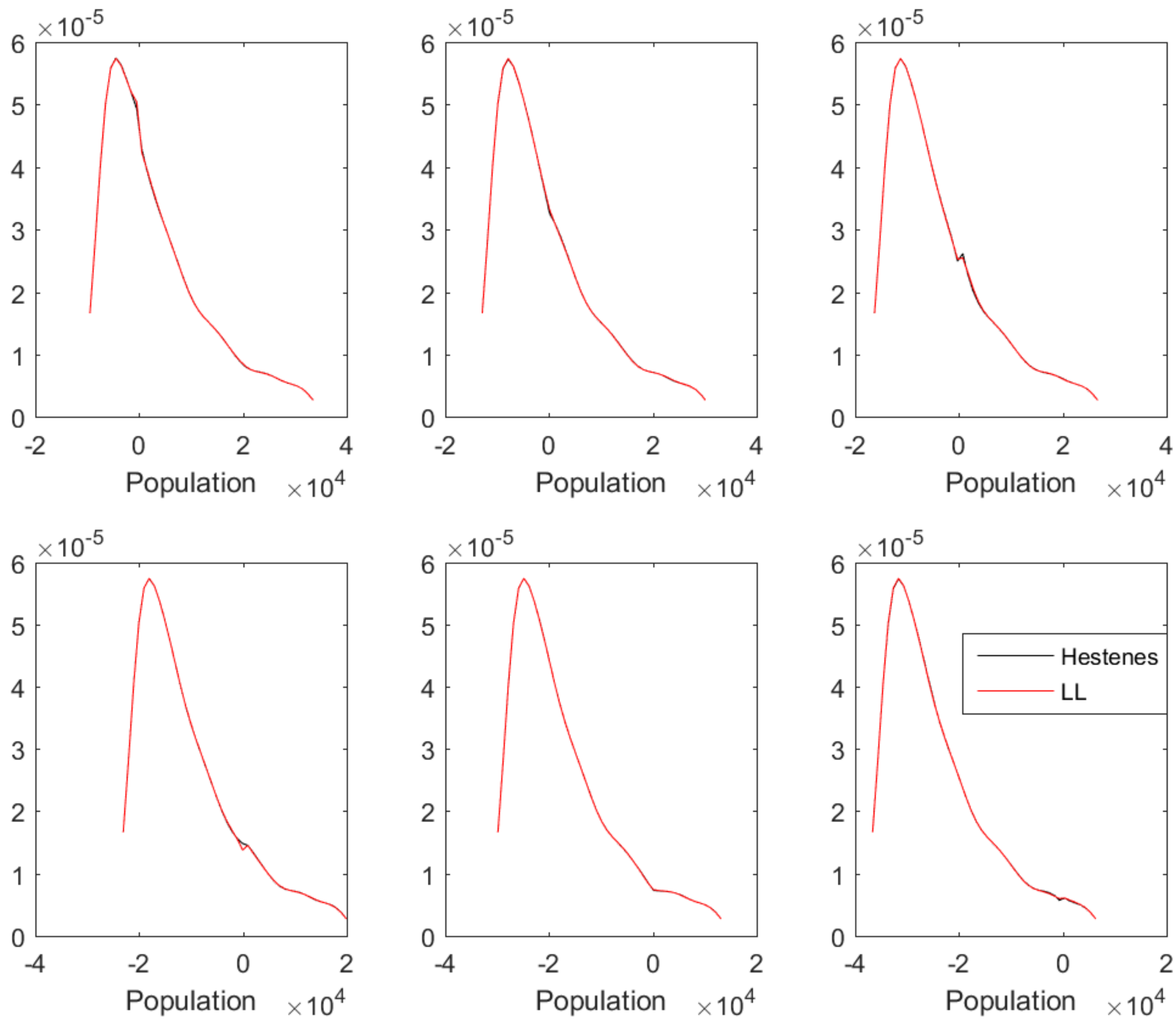

Figure 4: Population Density estimated by LL estimator (black) and Hestenes estimator (red)

Table 1: Using the true density to estimate four different regression functions

Jump estimators

	Experiment	NW	LL	H00	H10	H11	H21
Bias	1	0.6857	0.1310	0.6804	0.1086	0.4759	0.1008
Variance	1	0.0039	0.0345	0.0155	0.0455	0.0262	0.0394
RMSE	1	0.6885	0.2272	0.6917	0.2393	0.5026	0.2227
REL_RMSE	1	3.0912	1.0202	3.1056	1.0742	2.2567	1.0000
AMSE	1	0.1803	0.0967	0.1265	0.1591	0.1309	0.1552
REL_AMSE	1	1.8647	1.0000	1.3081	1.6448	1.3536	1.6049
Bias	2	-0.0154	0.0425	0.6602	0.1044	0.3706	-0.1294
Variance	2	0.0038	0.0389	0.0152	0.0521	0.0278	0.0480
RMSE	2	0.0635	0.2017	0.6716	0.2510	0.4063	0.2545
REL_RMSE	2	1.0000	3.1745	10.5720	3.9514	6.3963	4.0061
AMSE	2	0.1371	0.0954	0.1157	0.1379	0.1167	0.1353
REL_AMSE	2	1.4380	1.0000	1.2130	1.4455	1.2237	1.4187
Bias	3	0.2005	0.0364	-0.5990	-0.1118	-0.2915	0.1288
Variance	3	0.0023	0.0217	0.0092	0.0310	0.0172	0.0315
RMSE	3	0.2062	0.1517	0.6066	0.2085	0.3197	0.2193
REL_RMSE	3	1.3587	1.0000	3.9978	1.3741	2.1069	1.4453
AMSE	3	0.1064	0.0635	0.1369	0.1199	0.1075	0.1178
REL_AMSE	3	1.6752	1.0000	2.1559	1.8884	1.6932	1.8548
Bias	4	-0.9424	-0.0492	-0.5757	0.0100	-0.2095	0.2029
Variance	4	0.0032	0.0346	0.0128	0.0375	0.0215	0.0325
RMSE	4	0.9441	0.1924	0.5867	0.1938	0.2557	0.2715
REL_RMSE	4	4.9060	1.0000	3.0489	1.0071	1.3285	1.4109
AMSE	4	0.1700	0.0920	0.1093	0.1360	0.1113	0.1338
REL_AMSE	4	1.8468	1.0000	1.1880	1.4774	1.2099	1.4535

Sample Bias and Variance of regresion estimators

		NW-	NW+	LL-	LL+	H00-	H00+	H10-	H10+	H11-	H11+	H21-	H21+
Bias	1	-0.6734	0.0124	-0.0495	0.0815	-0.3467	0.3337	-0.0064	0.1021	-0.1560	0.3199	0.0122	0.1130
Variance	1	0.0009	0.0025	0.0086	0.0261	0.0036	0.0099	0.0134	0.0293	0.0065	0.0170	0.0099	0.0271
Bias	2	-0.3626	-0.3780	-0.0202	0.0223	-0.4161	0.2440	-0.0299	0.0745	-0.2020	0.1686	0.1357	0.0063
Variance	2	0.0022	0.0018	0.0264	0.0131	0.0087	0.0071	0.0256	0.0245	0.0150	0.0129	0.0239	0.0221
Bias	3	-0.2582	-0.0577	-0.0511	-0.0146	0.4836	-0.1154	0.0897	-0.0221	0.2329	-0.0586	-0.1316	-0.0028
Variance	3	0.0014	0.0010	0.0087	0.0128	0.0056	0.0040	0.0185	0.0124	0.0108	0.0069	0.0212	0.0106
Bias	4	0.3633	-0.5791	0.0220	-0.0272	0.4176	-0.1581	0.0315	0.0416	0.2042	-0.0052	-0.1324	0.0704
Variance	4	0.0022	0.0009	0.0259	0.0086	0.0086	0.0038	0.0251	0.0140	0.0150	0.0067	0.0239	0.0104

Asymptotic Bias and Variance from theory

		NW-	NW+	LL-	LL+	H00-	H00+	H10-	H10+	H11-	H11+	H21-	H21+
Bias	1	-0.6967	0.0290	-0.0545	0.0346	-0.3934	0.3670	-0.0314	0.0409	-0.0951	0.1238	0.0583	-0.0759
Variance	1	0.0013	0.0022	0.0083	0.0258	0.0053	0.0089	0.0157	0.0267	0.0090	0.0153	0.0136	0.0231
Bias	2	-0.3852	-0.3842	0.0346	0.0226	-0.4615	0.2317	0.0409	0.0391	0.1238	0.1184	-0.0759	-0.0726
Variance	2	0.0022	0.0020	0.0258	0.0129	0.0089	0.0082	0.0267	0.0244	0.0153	0.0140	0.0231	0.0211
Bias	3	-0.2671	-0.0551	-0.0545	-0.0237	0.4658	-0.1102	-0.0314	-0.0273	-0.0951	-0.0828	0.0583	0.0507
Variance	3	0.0013	0.0010	0.0083	0.0126	0.0053	0.0040	0.0157	0.0119	0.0090	0.0068	0.0136	0.0103
Bias	4	0.3852	-0.5893	-0.0346	-0.0545	0.4615	-0.1786	-0.0409	-0.0314	-0.1238	-0.0951	0.0759	0.0583
Variance	4	0.0022	0.0013	0.0258	0.0083	0.0089	0.0053	0.0267	0.0157	0.0153	0.0090	0.0231	0.0136

Table 2: Using the whole sample to estimate four different regression functions

Jump estimators

	Experiment	NW	LL	H00	H10	H11	H21
Bias	1	0.0289	0.1857	0.7488	0.1535	0.5147	0.0061
Variance	1	0.0067	0.0489	0.0268	0.0924	0.0490	0.0853
RMSE	1	0.0868	0.2888	0.7665	0.3404	0.5602	0.2921
REL_RMSE	1	1.0000	3.3253	8.8264	3.9203	6.4512	3.3637
AMSE	1	0.1626	0.1724	0.1472	0.1792	0.1539	0.1742
REL_AMSE	1	1.1044	1.1711	1.0000	1.2172	1.0451	1.1833
Bias	2	-0.0155	0.1242	0.6599	0.1075	0.3719	-0.1262
Variance	2	0.0039	0.0281	0.0156	0.0523	0.0285	0.0492
RMSE	2	0.0643	0.2087	0.6716	0.2528	0.4084	0.2552
REL_RMSE	2	1.0000	3.2474	10.4489	3.9323	6.3543	3.9711
AMSE	2	0.1378	0.1371	0.1162	0.1384	0.1174	0.1359
REL_AMSE	2	1.1859	1.1797	1.0000	1.1913	1.0105	1.1698
Bias	3	-0.0367	0.1018	0.6176	0.0897	0.3100	-0.1620
Variance	3	0.0029	0.0206	0.0114	0.0381	0.0209	0.0356
RMSE	3	0.0648	0.1761	0.6268	0.2147	0.3420	0.2487
REL_RMSE	3	1.0000	2.7150	9.6661	3.3111	5.2737	3.8351
AMSE	3	0.1256	0.1188	0.1013	0.1184	0.0998	0.1166
REL_AMSE	3	1.2590	1.1908	1.0153	1.1863	1.0000	1.1683
Bias	4	-0.0537	0.0844	0.5836	0.0755	0.2663	-0.1806
Variance	4	0.0023	0.0160	0.0090	0.0305	0.0164	0.0280
RMSE	4	0.0717	0.1519	0.5912	0.1902	0.2955	0.2462
REL_RMSE	4	1.0000	2.1195	8.2481	2.6540	4.1223	3.4344
AMSE	4	0.1185	0.1075	0.0918	0.1064	0.0891	0.1051
REL_AMSE	4	1.3296	1.2066	1.0303	1.1938	1.0000	1.1790

Sample Bias and Variance of regresion estimators

		NW-	NW+	LL-	LL+	H00-	H00+	H10-	H10+	H11-	H11+	H21-	H21+
Bias	1	-0.3908	-0.3619	-0.1255	0.0602	-0.4725	0.2762	-0.0540	0.0995	-0.2928	0.2219	0.0402	0.0464
Variance	1	0.0039	0.0031	0.0337	0.0160	0.0156	0.0125	0.0453	0.0438	0.0269	0.0226	0.0428	0.0388
Bias	2	-0.3626	-0.3782	-0.0796	0.0446	-0.4163	0.2437	-0.0319	0.0755	-0.2030	0.1689	0.1335	0.0073
Variance	2	0.0022	0.0017	0.0192	0.0092	0.0089	0.0070	0.0260	0.0238	0.0155	0.0128	0.0246	0.0221
Bias	3	-0.3493	-0.3860	-0.0634	0.0385	-0.3897	0.2280	-0.0258	0.0639	-0.1653	0.1447	0.1560	-0.0059
Variance	3	0.0016	0.0013	0.0138	0.0066	0.0063	0.0052	0.0188	0.0174	0.0108	0.0095	0.0172	0.0162
Bias	4	-0.3381	-0.3919	-0.0498	0.0345	-0.3673	0.2163	-0.0189	0.0566	-0.1375	0.1288	0.1678	-0.0127
Variance	4	0.0012	0.0010	0.0106	0.0052	0.0049	0.0041	0.0146	0.0140	0.0086	0.0073	0.0136	0.0124

Asymptotic Bias and Variance from theory

		NW-	NW+	LL-	LL+	H00-	H00+	H10-	H10+	H11-	H11+	H21-	H21+
Bias	1	-0.4236	-0.3708	0.0895	0.0623	-0.5381	0.2584	0.0539	0.0515	0.1634	0.1563	-0.1002	-0.0958
Variance	1	0.0039	0.0035	0.0320	0.0155	0.0155	0.0142	0.0465	0.0425	0.0267	0.0244	0.0402	0.0367
Bias	2	-0.3852	-0.3842	0.0678	0.0472	-0.4615	0.2317	0.0409	0.0391	0.1238	0.1184	-0.0759	-0.0726
Variance	2	0.0022	0.0020	0.0184	0.0089	0.0089	0.0082	0.0267	0.0244	0.0153	0.0140	0.0231	0.0211
Bias	3	-0.3656	-0.3916	0.0577	0.0402	-0.4221	0.2169	0.0347	0.0332	0.1053	0.1007	-0.0646	-0.0617
Variance	3	0.0016	0.0015	0.0133	0.0065	0.0064	0.0059	0.0193	0.0177	0.0111	0.0101	0.0167	0.0152
Bias	4	-0.3527	-0.3966	0.0514	0.0358	-0.3964	0.2068	0.0310	0.0296	0.0939	0.0897	-0.0575	-0.0550
Variance	4	0.0013	0.0012	0.0106	0.0051	0.0051	0.0047	0.0153	0.0140	0.0088	0.0081	0.0132	0.0121

Table 3: Using half of the sample to estimate four different regression functions

Jump estimators

		NW	LL	H00	H10	H11	H21
Bias	1	0.7496	0.1858	2.1901	0.9927	1.7160	0.6928
Variance	1	0.0229	0.0490	0.0916	0.3245	0.1668	0.2986
RMSE	1	0.7647	0.2889	2.2110	1.1445	1.7639	0.8824
REL_RMSE	1	2.6470	1.0000	7.6531	3.9617	6.1058	3.0543
AMSE	1	0.1510	0.1725	0.3214	0.3081	0.3135	0.2865
REL_AMSE	1	1.0000	1.1422	2.1288	2.0403	2.0763	1.8976
Bias	2	0.6605	0.1242	2.0121	0.9036	1.4322	0.4325
Variance	2	0.0132	0.0281	0.0527	0.1836	0.0959	0.1695
RMSE	2	0.6704	0.2087	2.0251	1.0001	1.4653	0.5972
REL_RMSE	2	3.2121	1.0000	9.7025	4.7913	7.0205	2.8610
AMSE	2	0.1205	0.1371	0.2808	0.2666	0.2670	0.2480
REL_AMSE	2	1.0000	1.1378	2.3309	2.2134	2.2161	2.0584
Bias	3	0.6176	0.1017	1.9263	0.8680	1.3082	0.3617
Variance	3	0.0095	0.0206	0.0379	0.1331	0.0691	0.1212
RMSE	3	0.6253	0.1760	1.9361	0.9416	1.3344	0.5020
REL_RMSE	3	3.5531	1.0000	11.0017	5.3505	7.5824	2.8526
AMSE	3	0.1057	0.1188	0.2610	0.2466	0.2452	0.2298
REL_AMSE	3	1.0000	1.1239	2.4684	2.3321	2.3191	2.1732
Bias	4	0.5836	0.0844	1.8581	0.8402	1.2215	0.3261
Variance	4	0.0074	0.0160	0.0298	0.1055	0.0542	0.0949
RMSE	4	0.5899	0.1519	1.8661	0.9007	1.2434	0.4486
REL_RMSE	4	3.8830	1.0000	12.2835	5.9290	8.1848	2.9529
AMSE	4	0.0962	0.1075	0.2477	0.2341	0.2316	0.2189
REL_AMSE	4	1.0000	1.1178	2.5753	2.4334	2.4072	2.2752

Sample Bias and Variance of regresion estimators

		NW-	NW+	LL-	LL+	H00-	H00+	H10-	H10+	H11-	H11+	H21-	H21+
Bias	1	-0.4736	0.2760	-0.1255	0.0603	-0.6382	1.5519	0.2022	1.1949	-0.2758	1.4402	0.3942	1.0870
Variance	1	0.0158	0.0072	0.0337	0.0160	0.0633	0.0290	0.1808	0.1459	0.1079	0.0587	0.1754	0.1213
Bias	2	-0.4168	0.2437	-0.0795	0.0447	-0.5246	1.4874	0.2454	1.1490	-0.0962	1.3360	0.5787	1.0112
Variance	2	0.0090	0.0042	0.0192	0.0092	0.0361	0.0170	0.1033	0.0806	0.0621	0.0342	0.1007	0.0698
Bias	3	-0.3901	0.2276	-0.0633	0.0384	-0.4711	1.4551	0.2581	1.1262	-0.0208	1.2875	0.6237	0.9853
Variance	3	0.0063	0.0030	0.0138	0.0066	0.0253	0.0121	0.0747	0.0583	0.0436	0.0250	0.0707	0.0506
Bias	4	-0.3676	0.2160	-0.0498	0.0345	-0.4261	1.4321	0.2714	1.1115	0.0346	1.2561	0.6463	0.9724
Variance	4	0.0050	0.0024	0.0106	0.0052	0.0199	0.0097	0.0578	0.0466	0.0343	0.0194	0.0554	0.0388

Asymptotic Bias and Variance from theory

		NW-	NW+	LL-	LL+	H00-	H00+	H10-	H10+	H11-	H11+	H21-	H21+
Bias	1	-0.5062	0.2981	0.0895	0.0623	-0.5381	0.2584	0.0539	0.0515	0.1634	0.1563	-0.1002	-0.0958
Variance	1	0.0039	0.0035	0.0320	0.0155	0.0155	0.0142	0.0465	0.0425	0.0267	0.0244	0.0402	0.0367
Bias	2	-0.4373	0.2618	0.0678	0.0472	-0.4615	0.2317	0.0409	0.0391	0.1238	0.1184	-0.0759	-0.0726
Variance	2	0.0022	0.0020	0.0184	0.0089	0.0089	0.0082	0.0267	0.0244	0.0153	0.0140	0.0231	0.0211
Bias	3	-0.4016	0.2425	0.0577	0.0402	-0.4221	0.2169	0.0347	0.0332	0.1053	0.1007	-0.0646	-0.0617
Variance	3	0.0016	0.0015	0.0133	0.0065	0.0064	0.0059	0.0193	0.0177	0.0111	0.0101	0.0167	0.0152
Bias	4	-0.3781	0.2296	0.0514	0.0358	-0.3964	0.2068	0.0310	0.0296	0.0939	0.0897	-0.0575	-0.0550
Variance	4	0.0013	0.0012	0.0106	0.0051	0.0051	0.0047	0.0153	0.0140	0.0088	0.0081	0.0132	0.0121

Table 4: Choosing Sample size from 1000, 2000, 3000, to 4000

Jump estimators

		NW	LL	H00	H10	H11	H21
Bias	1	0.0154	0.1858	0.7219	0.1223	0.4861	-0.0175
Variance	1	0.0076	0.0490	0.0305	0.0913	0.0546	0.0857
RMSE	1	0.0887	0.2889	0.7427	0.3260	0.5393	0.2933
REL_RMSE	1	1.0000	3.2562	8.3711	3.6739	6.0790	3.3058
AMSE	1	0.1858	0.1725	0.1675	0.1974	0.1733	0.1939
REL_AMSE	1	1.1090	1.0294	1.0000	1.1783	1.0346	1.1577
Bias	2	-0.0250	0.1242	0.6410	0.0833	0.3512	-0.1445
Variance	2	0.0044	0.0281	0.0176	0.0520	0.0314	0.0491
RMSE	2	0.0709	0.2087	0.6546	0.2429	0.3934	0.2644
REL_RMSE	2	1.0000	2.9458	9.2384	3.4277	5.5516	3.7318
AMSE	2	0.1552	0.1371	0.1323	0.1535	0.1336	0.1521
REL_AMSE	2	1.1736	1.0361	1.0000	1.1608	1.0098	1.1497
Bias	3	-0.0449	0.1017	0.6011	0.0690	0.2924	-0.1776
Variance	3	0.0032	0.0206	0.0127	0.0377	0.0226	0.0353
RMSE	3	0.0720	0.1760	0.6116	0.2061	0.3288	0.2586
REL_RMSE	3	1.0000	2.4425	8.4888	2.8611	4.5635	3.5887
AMSE	3	0.1400	0.1188	0.1148	0.1315	0.1138	0.1306
REL_AMSE	3	1.2299	1.0439	1.0088	1.1555	1.0000	1.1470
Bias	4	-0.0609	0.0844	0.5693	0.0573	0.2510	-0.1944
Variance	4	0.0025	0.0160	0.0100	0.0303	0.0178	0.0278
RMSE	4	0.0788	0.1519	0.5780	0.1833	0.2842	0.2561
REL_RMSE	4	1.0000	1.9282	7.3358	2.3262	3.6069	3.2502
AMSE	4	0.1310	0.1075	0.1040	0.1185	0.1020	0.1178
REL_AMSE	4	1.2839	1.0540	1.0193	1.1612	1.0000	1.1547

Sample Bias and Variance of regresion estimators

		NW-	NW+	LL-	LL+	H00-	H00+	H10-	H10+	H11-	H11+	H21-	H21+
Bias	1	-0.3899	-0.3745	-0.1255	0.0603	-0.4708	0.2511	-0.0569	0.0654	-0.2936	0.1925	0.0347	0.0172
Variance	1	0.0038	0.0039	0.0337	0.0160	0.0153	0.0155	0.0444	0.0468	0.0264	0.0279	0.0407	0.0438
Bias	2	-0.3622	-0.3872	-0.0795	0.0447	-0.4155	0.2255	-0.0338	0.0496	-0.2038	0.1475	0.1297	-0.0147
Variance	2	0.0022	0.0022	0.0192	0.0092	0.0088	0.0088	0.0258	0.0259	0.0152	0.0160	0.0236	0.0252
Bias	3	-0.3490	-0.3939	-0.0633	0.0384	-0.3890	0.2121	-0.0274	0.0416	-0.1662	0.1263	0.1527	-0.0250
Variance	3	0.0015	0.0016	0.0138	0.0066	0.0062	0.0066	0.0187	0.0190	0.0107	0.0117	0.0166	0.0184
Bias	4	-0.3379	-0.3988	-0.0498	0.0345	-0.3668	0.2024	-0.0205	0.0368	-0.1383	0.1126	0.1649	-0.0296
Variance	4	0.0012	0.0013	0.0106	0.0052	0.0049	0.0051	0.0145	0.0153	0.0085	0.0091	0.0131	0.0142

Asymptotic Bias and Variance from theory

		NW-	NW+	LL-	LL+	H00-	H00+	H10-	H10+	H11-	H11+	H21-	H21+
Bias	1	-0.4236	-0.3708	0.0895	0.0623	-0.5381	0.2584	0.0539	0.0515	0.1634	0.1563	-0.1002	-0.0958
Variance	1	0.0039	0.0035	0.0320	0.0155	0.0155	0.0142	0.0465	0.0425	0.0267	0.0244	0.0402	0.0367
Bias	2	-0.3852	-0.3842	0.0678	0.0472	-0.4615	0.2317	0.0409	0.0391	0.1238	0.1184	-0.0759	-0.0726
Variance	2	0.0022	0.0020	0.0184	0.0089	0.0089	0.0082	0.0267	0.0244	0.0153	0.0140	0.0231	0.0211
Bias	3	-0.3656	-0.3916	0.0577	0.0402	-0.4221	0.2169	0.0347	0.0332	0.1053	0.1007	-0.0646	-0.0617
Variance	3	0.0016	0.0015	0.0133	0.0065	0.0064	0.0059	0.0193	0.0177	0.0111	0.0101	0.0167	0.0152
Bias	4	-0.3527	-0.3966	0.0514	0.0358	-0.3964	0.2068	0.0310	0.0296	0.0939	0.0897	-0.0575	-0.0550
Variance	4	0.0013	0.0012	0.0106	0.0051	0.0051	0.0047	0.0153	0.0140	0.0088	0.0081	0.0132	0.0121

Table 5: Choosing Bandwidth by Cross-Validation vs. Plug-in

		Plug-in Method to obtain optimal h (with 10,000 repetitions)						Cross-Validation to obtain optimal h(with 2,000 repetitions)					
		NW	LL	H00	H10	H11	H21	NW	LL	H00	H10	H11	H21
Bias	1	0.6848	0.2144	0.6785	0.1038	0.4724	0.0956	0.5063	0.0569	0.4014	0.0053	0.1390	-0.1031
Variance	1	0.0038	0.0277	0.0152	0.0447	0.0257	0.0386	0.0099	0.0478	0.0236	0.0787	0.0424	0.0578
RMSE	1	0.6875	0.2715	0.6897	0.2355	0.4988	0.2186	0.5160	0.2259	0.4297	0.2805	0.2483	0.2616
REL_RMSE	1	3.1457	1.2422	3.1554	1.0774	2.2824	1.0000	2.2845	1.0000	1.9026	1.2417	1.0992	1.1582
AMSE	1	0.1805	0.1294	0.1264	0.1589	0.1307	0.1550	0.1099	0.0789	0.0861	0.0854	0.0811	0.0838
REL_AMSE	1	1.4273	1.0236	1.0000	1.2569	1.0335	1.2260	1.3935	1.0000	1.0911	1.0829	1.0286	1.0627
Bias	2	-0.0149	0.1249	0.6612	0.1077	0.3734	-0.1246	-0.1490	0.0294	0.4208	0.0307	0.1291	-0.1029
Variance	2	0.0039	0.0287	0.0156	0.0529	0.0285	0.0496	0.0092	0.0516	0.0212	0.0787	0.0381	0.0562
RMSE	2	0.0642	0.2104	0.6729	0.2539	0.4098	0.2551	0.1772	0.2290	0.4452	0.2821	0.2339	0.2582
REL_RMSE	2	1.0000	3.2798	10.4885	3.9578	6.3868	3.9760	1.0000	1.2923	2.5121	1.5916	1.3197	1.4570
AMSE	2	0.1376	0.1369	0.1162	0.1384	0.1173	0.1357	0.1082	0.0786	0.0873	0.0857	0.0812	0.0837
REL_AMSE	2	1.1846	1.1780	1.0000	1.1913	1.0096	1.1681	1.3759	1.0000	1.1104	1.0895	1.0327	1.0647
Bias	3	0.2007	0.0191	-0.5985	-0.1111	-0.2905	0.1308	0.3167	-0.0154	-0.2870	-0.0200	-0.0908	0.0423
Variance	3	0.0023	0.0180	0.0090	0.0302	0.0168	0.0307	0.0082	0.0148	0.0104	0.0330	0.0180	0.0224
RMSE	3	0.2063	0.1357	0.6060	0.2063	0.3182	0.2188	0.3294	0.1225	0.3047	0.1828	0.1619	0.1554
REL_RMSE	3	1.5204	1.0000	4.4667	1.5207	2.3451	1.6123	2.6885	1.0000	2.4864	1.4917	1.3215	1.2684
AMSE	3	0.1060	0.0597	0.1365	0.1192	0.1068	0.1172	0.0822	0.0503	0.0655	0.0627	0.0585	0.0609
REL_AMSE	3	1.7750	1.0000	2.2858	1.9968	1.7881	1.9625	1.6341	1.0000	1.3029	1.2471	1.1620	1.2107
Bias	4	-0.9433	-0.1065	-0.5776	0.0049	-0.2124	0.1989	-0.7948	-0.0368	-0.3599	0.0460	-0.0331	0.1382
Variance	4	0.0032	0.0284	0.0129	0.0384	0.0217	0.0330	0.0079	0.0402	0.0162	0.0550	0.0296	0.0398
RMSE	4	0.9450	0.1993	0.5887	0.1961	0.2586	0.2693	0.7997	0.2038	0.3818	0.2390	0.1751	0.2427
REL_RMSE	4	4.8196	1.0165	3.0022	1.0000	1.3186	1.3732	4.5664	1.1635	2.1798	1.3645	1.0000	1.3857
AMSE	4	0.1704	0.1333	0.1097	0.1362	0.1117	0.1339	0.1135	0.0756	0.0830	0.0809	0.0771	0.0799
REL_AMSE	4	1.5523	1.2142	1.0000	1.2409	1.0179	1.2205	1.5018	1.0000	1.0986	1.0703	1.0205	1.0578

Table 6: Choosing Kernels with or without Compact Support

		a standard Gussian kernel						an Epanechnikov kernel					
		NW	LL	H00	H10	H11	H21	NW	LL	H00	H10	H11	H21
Bias	1	0.6952	0.1304	0.6995	0.1610	0.5092	0.1515	0.7163	0.0951	0.7416	0.1599	0.5423	0.0891
Variance	1	0.0037	0.0347	0.0147	0.0426	0.0250	0.0374	0.0036	0.0392	0.0143	0.0443	0.0263	0.0434
RMSE	1	0.6978	0.2274	0.7099	0.2617	0.5332	0.2458	0.7188	0.2196	0.7512	0.2642	0.5660	0.2266
REL_RMSE	1	3.0692	1.0000	3.1220	1.1512	2.3450	1.0810	3.2727	1.0000	3.4202	1.2031	2.5770	1.0316
AMSE	1	0.1864	0.0967	0.1327	0.1558	0.1332	0.1515	0.1944	0.0981	0.1375	0.1645	0.1393	0.1693
REL_AMSE	1	1.9275	1.0000	1.3721	1.6112	1.3774	1.5668	1.9816	1.0000	1.4009	1.6766	1.4195	1.7249
Bias	2	-0.0249	0.0438	0.6412	0.0803	0.3509	-0.1456	-0.0068	0.0251	0.6775	0.0766	0.3522	-0.2820
Variance	2	0.0043	0.0396	0.0173	0.0524	0.0309	0.0486	0.0044	0.0438	0.0174	0.0555	0.0341	0.0630
RMSE	2	0.0704	0.2038	0.6546	0.2425	0.3926	0.2642	0.0663	0.2107	0.6902	0.2478	0.3977	0.3776
REL_RMSE	2	1.0000	2.8954	9.3022	3.4465	5.5783	3.7543	1.0000	3.1786	10.4103	3.7372	5.9987	5.6954
AMSE	2	0.1549	0.0958	0.1318	0.1532	0.1330	0.1516	0.1577	0.0960	0.1334	0.1592	0.1355	0.1670
REL_AMSE	2	1.6167	1.0000	1.3762	1.5997	1.3889	1.5825	1.6434	1.0000	1.3896	1.6584	1.4116	1.7397
Bias	3	0.2274	0.0356	-0.5451	-0.0489	-0.2345	0.1771	0.2142	0.0289	-0.5717	-0.0428	-0.2260	0.2789
Variance	3	0.0025	0.0212	0.0101	0.0308	0.0190	0.0315	0.0026	0.0244	0.0102	0.0333	0.0218	0.0465
RMSE	3	0.2329	0.1501	0.5543	0.1821	0.2720	0.2506	0.2200	0.1588	0.5806	0.1875	0.2700	0.3526
REL_RMSE	3	1.5522	1.0000	3.6939	1.2135	1.8129	1.6702	1.3856	1.0000	3.6560	1.1805	1.7003	2.2202
AMSE	3	0.1394	0.0635	0.1424	0.1504	0.1272	0.1574	0.1421	0.0644	0.1462	0.1581	0.1304	0.1793
REL_AMSE	3	2.1937	1.0000	2.2422	2.3680	2.0028	2.4779	2.2077	1.0000	2.2717	2.4557	2.0257	2.7858
Bias	4	-0.9565	-0.0493	-0.6040	-0.0515	-0.2544	0.1421	-0.9722	-0.0336	-0.6354	-0.0481	-0.2502	0.2641
Variance	4	0.0032	0.0348	0.0129	0.0389	0.0220	0.0333	0.0031	0.0404	0.0123	0.0403	0.0226	0.0393
RMSE	4	0.9582	0.1931	0.6146	0.2038	0.2945	0.2313	0.9738	0.2038	0.6450	0.2065	0.2919	0.3302
REL_RMSE	4	4.9632	1.0000	3.1833	1.0556	1.5253	1.1979	4.7787	1.0000	3.1653	1.0132	1.4325	1.6202
AMSE	4	0.1788	0.0919	0.1163	0.1342	0.1134	0.1316	0.1857	0.0925	0.1184	0.1402	0.1158	0.1460
REL_AMSE	4	1.9456	1.0000	1.2652	1.4608	1.2337	1.4325	2.0069	1.0000	1.2793	1.5150	1.2509	1.5776

Table 7: Choosing Density Functions

		Density of x is a standard normal distribution						Density of x is a normal with mu = 0.5					
		NW	LL	H00	H10	H11	H21	NW	LL	H00	H10	H11	H21
Bias	1	0.6848	0.2144	0.6785	0.1038	0.4724	0.0956	0.7033	0.2322	0.7156	0.2091	0.6604	0.4858
Variance	1	0.0038	0.0277	0.0152	0.0447	0.0257	0.0386	0.0046	0.0313	0.0186	0.0530	0.0316	0.0466
RMSE	1	0.6875	0.2715	0.6897	0.2355	0.4988	0.2186	0.7066	0.2920	0.7285	0.3109	0.6839	0.5316
REL_RMSE	1	3.1457	1.2422	3.1554	1.0774	2.2824	1.0000	2.4202	1.0000	2.4953	1.0649	2.3425	1.8207
AMSE	1	0.1805	0.1294	0.1264	0.1589	0.1307	0.1550	0.1787	0.1351	0.1439	0.1939	0.1667	0.1916
REL_AMSE	1	1.4273	1.0236	1.0000	1.2569	1.0335	1.2260	1.3229	1.0000	1.0651	1.4349	1.2342	1.4184
Bias	2	-0.0149	0.1249	0.6612	0.1077	0.3734	-0.1246	0.1976	0.1342	1.0862	0.3080	0.9110	0.3606
Variance	2	0.0039	0.0287	0.0156	0.0529	0.0285	0.0496	0.0030	0.0307	0.0118	0.0407	0.0213	0.0368
RMSE	2	0.0642	0.2104	0.6729	0.2539	0.4098	0.2551	0.2050	0.2207	1.0916	0.3681	0.9226	0.4084
REL_RMSE	2	1.0000	3.2798	10.4885	3.9578	6.3868	3.9760	1.0000	1.0769	5.3262	1.7960	4.5014	1.9927
AMSE	2	0.1376	0.1369	0.1162	0.1384	0.1173	0.1357	0.1712	0.1424	0.1962	0.2179	0.2026	0.2125
REL_AMSE	2	1.1846	1.1780	1.0000	1.1913	1.0096	1.1681	1.2020	1.0000	1.3774	1.5297	1.4227	1.4919
Bias	3	0.2007	0.0191	-0.5985	-0.1111	-0.2905	0.1308	0.2460	0.0163	-0.5081	-0.1763	-0.4098	-0.1335
Variance	3	0.0023	0.0180	0.0090	0.0302	0.0168	0.0307	0.0021	0.0214	0.0086	0.0288	0.0155	0.0268
RMSE	3	0.2063	0.1357	0.6060	0.2063	0.3182	0.2188	0.2503	0.1471	0.5164	0.2448	0.4283	0.2113
REL_RMSE	3	1.5204	1.0000	4.4667	1.5207	2.3451	1.6123	1.7020	1.0000	3.5119	1.6647	2.9124	1.4367
AMSE	3	0.1060	0.0597	0.1365	0.1192	0.1068	0.1172	0.2101	0.0624	0.1755	0.2502	0.2076	0.2412
REL_AMSE	3	1.7750	1.0000	2.2858	1.9968	1.7881	1.9625	3.3671	1.0000	2.8136	4.0110	3.3277	3.8664
Bias	4	-0.9433	-0.1065	-0.5776	0.0049	-0.2124	0.1989	-0.9714	-0.1100	-0.6339	0.0305	-0.3286	0.0202
Variance	4	0.0032	0.0284	0.0129	0.0384	0.0217	0.0330	0.0027	0.0312	0.0108	0.0314	0.0183	0.0273
RMSE	4	0.9450	0.1993	0.5887	0.1961	0.2586	0.2693	0.9728	0.2081	0.6424	0.1799	0.3553	0.1664
REL_RMSE	4	4.8196	1.0165	3.0022	1.0000	1.3186	1.3732	5.8474	1.2510	3.8610	1.0812	2.1356	1.0000
AMSE	4	0.1704	0.1333	0.1097	0.1362	0.1117	0.1339	0.2127	0.1386	0.1500	0.2165	0.1769	0.2162
REL_AMSE	4	1.5523	1.2142	1.0000	1.2409	1.0179	1.2205	1.5344	1.0000	1.0822	1.5613	1.2761	1.5592

Table 8: Regression Discontinuity Estimates on Treatment and Outcome Variables

	Hestenes Estimation			From Litschig & Morrison	
Bandwidth	Rule-of-thumb	Cross-validation	Cross-validation	Fixed	Fixed
Pretreatment covariates	No	No	Yes	No	Yes
log FPM transfers per capita	0.210	0.198	0.197		
log total spending per capita	0.218	0.217	0.155	0.0197	0.167
log other revenue per capita	0.078	-0.033	-0.052		
log own revenue per capita	0.407	0.316	0.128		
schooling, 19 to 28 age group in 1991	0.686	0.678	0.453	0.516	0.301
literacy rate, 19 to 28 age group in 1991	0.078	0.079	0.053	0.062	0.049
poverty rate in 1991	-0.051	-0.040	-0.018	-0.06	-0.051
party reelection rate	0.052	0.079	0.058	0.086	0.106

	Hestenes Estimation						From Litschig & Morrison	
Bandwidth	Rule-of-thumb		Cross-validation		Cross-validation		Fixed	Fixed
log FPM transfers per capita	1.58	2.37	2.08	2.14	2.07	2.14		
log total spending per capita	1.58	2.37	5.38	3.53	4.79	6.42	4	4
log other revenue per capita	1.58	2.37	1.09	3.28	1.47	3.75		
log own revenue per capita	1.58	2.37	4.04	3.30	3.84	2.88		
schooling, 19 to 28 age group in 1991	1.58	2.37	3.48	2.78	3.11	2.43	3	3
literacy rate, 19 to 28 age group in 1991	1.58	2.37	3.46	2.46	3.24	2.29	3	3
poverty rate in 1991	1.58	2.37	3.40	2.70	3.33	2.24	3	3
party reelection rate	1.58	2.37	3.56	7.42	3.75	7.63	4	4

Table 9: Density Estimates on Population

The Cutoffs	Local Linear	Hestenes	McCrary	(SE of McCrary)
10188	-0.0288	0.0456	-0.0720	0.095
13584	0.0331	0.0291	0.0110	0.111
16980	-0.4656	-0.3222	0.1800	0.136
23772	0.1955	0.0393	0.0540	0.174
30564	0.1336	-0.1626	-0.0110	0.269
37356	-0.6823	-0.5325	0.3500	0.357

A Supporting lemmas

Lemma 1. *Suppose the assumptions of Theorem 3 are holding and $h \to 0$ and $nh \to \infty$ as $n \to \infty$. The covariance of $\hat{g}(x)$ and $\hat{f}(x)$ for $x \geq 0$ has the following representation*

$$Cov(\hat{g}(x), \hat{f}(x)) = \begin{cases} \frac{1}{nh}\left\{m^{+}(0)f(0)\sum_{,j=0}^{s+1}\frac{k_j}{w_j}\int_0^{\infty}K(u)K\left(\frac{u}{w_j}\right)du + o(1)\right\}, & x = 0 \\ \frac{1}{nh}\left\{m^{+}(x)f(x)\int_{\mathbb{R}}K^2(u)du + o(1)\right\}, & x > 0 \end{cases}$$

Proof. Let $\hat{g}(x) = \frac{1}{nh}\sum_{i=1}^{n}\left[K\left(\frac{X_i - x}{h}\right) + \sum_{j=1}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{X_i}{w_j}+x}{h}\right)\right]Y_i d_i$, $w_0 = -1$, $k_0 = -1$, $u_i = \sum_{j=0}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{X_i}{w_j}+x}{h}\right)$ and $K(\frac{X_i - x}{h}) = K_i$. Then, since $\{X_i\}_{i=1,2,\cdots}$ forms an i.i.d. sequence and $E(Y|X=x) = m^{+}(x)$ for $x \geq 0$

$$\begin{aligned} Cov(\hat{g}(x), \hat{f}(x)) &= \frac{n}{(nh)^2}E\left[m^{+}(X_1)u_1 d_1 K_1\right] + \frac{n(n-1)}{(nh)^2}E\left[m^{+}(X_1)u_1 d_1\right]E\left[K_1\right] - \frac{1}{h^2}E\left[m^{+}(X_1)u_1 d_1\right]E\left[K_1\right] \\ &= \frac{1}{nh^2}E\left[m^{+}(X_1)u_1 d_1 K_1\right] - \frac{1}{nh^2}E\left[m^{+}(X_1)u_1 d_1\right]E\left[K_1\right] \end{aligned}$$

Put $\frac{1}{h}E\left[m^{+}(X_1)u_1 d_1 K_1\right] = T_1$ and $\frac{1}{h^2}E\left[m^{+}(X_1)u_1 d_1\right]E\left[K_1\right] = T_2$, then $Cov(\hat{g}(x), \hat{f}(x)) = \frac{1}{nh}T_1 - \frac{1}{n}T_2$. As with the variance, the covariance will be different for $x = 0$ and $x > 0$.

Case ($x > 0$):

$$T_1 = \frac{1}{h}\int_0^{\infty}K^2\left(\frac{X_1 - x}{h}\right)m^{+}(X_1)f(X_1)dX_1 + \frac{1}{h}\int_0^{\infty}K\left(\frac{X_1 - x}{h}\right)\sum_{j=1}^{s+1}\frac{k_j}{w_j}K\left(\frac{\frac{X_1}{w_j}+x}{h}\right)m^{+}(X_1)f(X_1)dX_1$$

letting $\dfrac{x - X_1}{h} = u$ in the first term and $\dfrac{\frac{X_1}{w_j}+x}{h} = u$ in the second

$$\begin{aligned} &= \int_{-\infty}^{\frac{x}{h}}K^2(u)m^{+}(x - hu)f(x - hu)du \\ &+ \sum_{j=1}^{s+1}k_j\int_{\frac{x}{h}}^{\infty}K\left(w_j u - (w_j+1)\frac{x}{h}\right)K(u)m^{+}(w_j(hu - x))f(w_j(hu - x))du = I_{00} + \sum_{j=1}^{s+1}k_j I_{0j} \end{aligned}$$

where $I_{00} = \int_{-\infty}^{\frac{x}{h}}K^2(u)m^{+}(x - hu)f(x - hu)du$ and $I_{0j} = \int_{\frac{x}{h}}^{\infty}K\left(w_j u - (w_j+1)\frac{x}{h}\right)K(u)m^{+}(w_j(hu - x))f(w_j(hu - x))du$. Now,

$$\begin{aligned} \left|I_{00} - m^{+}(x)f(x)\int_{\mathbb{R}}K^2(u)du\right| &= \left|\int_{\mathbb{R}}K^2(u)\left[m^{+}(x - hu)f(x - hu) - m^{+}(x)f(x)\right]du - \int_{\frac{x}{h}}^{\infty}K^2(u)m^{+}(x - hu)f(x - hu)du\right| \\ &\leq \left|\int_{|u|\leq C}K^2(u)\left[m^{+}(x - hu)f(x - hu) - m^{+}(x)f(x)\right]du\right| \\ &+ \left|\int_{|u|>C}K^2(u)\left[m^{+}(x - hu)f(x - hu) - m^{+}(x)f(x)\right]du\right| \\ &+ \left|\int_{\frac{x}{h}}^{\infty}K^2(u)m^{+}(x - hu)f(x - hu)du\right|, \text{ for } C > 0 \\ &\leq C\bar{p}(Ch, x)\int_{|u|\leq C}K^2(u)du + C\int_{|u|>C}K^2(u)du + C\int_{\frac{x}{h}}^{\infty}K^2(u)du \end{aligned}$$

where the last equality follows from the uniform boundedness of K, f and m^+ and where $\bar{p}(Ch, x)$ is as defined in the proof of Theorem 2. By continuity of f and the fact that $\int K^2(u)du < C$, for all $\epsilon > 0$,

$$\left| I_{00} - m^+(x)f(x)\int_{\mathbb{R}} K^2(u)du \right| \leq \epsilon. \tag{31}$$

Similarly,

$$|I_{0j}| \leq C\int_{\frac{x}{h}}^{\infty} K(u)\,du < \epsilon \text{ for all } n \text{ sufficiently large.} \tag{32}$$

Consequently, $\left|T_1 - m^+(x)f(x)\int_{\mathbb{R}} K^2(u)du\right| \leq \epsilon$.

Turning to T_2, we first observe that from standard properties of $\hat{f}$ we have

$$\frac{1}{h}E\left(K\left(\frac{X_1 - x}{h}\right)\right) \to f(x). \tag{33}$$

Now, letting $\frac{x-X_1}{h} = u$ and $\frac{\frac{X_1}{w_j}+x}{h} = u$ we have

$$\begin{aligned}
h^{-1}E\left[m^+(X_1)u_1d_1\right] = h^{-1}E\left(m^+(X_1)u_1\right) &= \frac{1}{h}\int_0^{\infty} K\left(\frac{X_1 - x}{h}\right)m^+(X_1)f(X_1)dX_1 \\
&+ \frac{1}{h}\sum_{j=1}^{s+1}\frac{k_j}{w_j}\int_0^{\infty} K\left(\frac{\frac{X_1}{w_j}+x}{h}\right)m^+(X_1)f(X_1)dX_1 \\
&= \int_{-\infty}^{\frac{x}{h}} K(u)m^+(x-hu)f(x-hu)du + \sum_{j=1}^{s+1}k_j\int_{\frac{x}{h}}^{\infty} K(u)m^+(w_j(hu-x))f(w_j(hu-x))du \\
&= I_1 + \sum_{j=1}^{s+1} I_{2j}
\end{aligned}$$

Using arguments similar to those used in the study of T_1 we have

$$\begin{aligned}
\left|I_1 - m^+(x)f(x)\int_{\mathbb{R}} K(u)du\right| &\leq C\bar{p}(Ch, x)\int_{|u|\leq C} K(u)du + 2C\int_{|u|>C} K(u)du \\
&+ C\int_{\frac{x}{h}}^{\infty} K(u)du
\end{aligned}$$

By continuity of $f(x)$ we have, for all $\epsilon > 0$, $\left|I_1 - m^+(x)f(x)\int_{\mathbb{R}} K^2(u)du\right| \leq \epsilon$. Similarly,

$$|I_{2j}| \leq C\sum_{j=0}^{s+1}k_j\int_{\frac{x}{h}}^{\infty}|K(u)|\,du \leq \epsilon$$

for sufficiently large n. Therefore,

$$\left|\frac{1}{h}E\left(m^+(X_1)u_1\right) - m^+(x)f(x)\int_{\mathbb{R}} K^2(u)du\right| \leq \epsilon. \tag{34}$$

Thus for $x > 0$, $Cov(\hat{g}(x), \hat{f}(x)) = \frac{1}{nh}T_1 - \frac{1}{n}T_2 = \frac{1}{nh}\left(m^+(x)f(x)\int_{\mathbb{R}} K^2(u)du + o(1)\right)$.

Case ($x = 0$): Repeating the change in variables used above, we have

$$T_1 = I_{00} + \sum_{j=1}^{s+1}\frac{k_j}{w_j}I_{0j}$$

where $I_{00} = \int_{-\infty}^{0} K^2(u)m^+(-hu)f(-hu)du$ and $I_{0j} = \int_0^\infty K\left(\frac{u}{w_0}\right)K\left(\frac{u}{w_j}\right)m^+(hu)f(hu)du$. Now,

$$\begin{aligned}\left|I_{00} - m^+(0)f(0)\int_{-\infty}^{0} K^2(u)du\right| &= \left|\int_{-\infty}^{-C} K^2(u)\left[m^+(-hu)f(-hu) - m^+(0)f(0)\right]du\right. \\ &\quad \left. + \int_{-C}^{0} K^2(u)\left[m^+(-hu)f(-hu) - m^+(0)f(0)\right]du\right| \\ &\le C\int_{-\infty}^{-C} K^2(u)du + \bar{p}(Ch,0)\int_0^C K^2(u)du\end{aligned}$$

For n sufficiently large and all $\epsilon > 0$,

$$\left|I_{00} - m^+(0)f(0)\int_{-\infty}^{0} K^2(u)du\right| \le \epsilon. \tag{35}$$

Similar arguments give,

$$\left|I_{0j} - m^+(0)f(0)\int_0^\infty K(u)\,K\left(\frac{u}{w_j}\right)du\right| < \epsilon. \tag{36}$$

Consequently, (35) and (36) give

$$\left|T_1 - m^+(0)f(0)\sum_{,j=0}^{s+1}\frac{k_j}{w_j}\int_0^\infty K(u)\,K\left(\frac{u}{w_j}\right)du\right| < \epsilon.$$

Turning to $T_2 = \frac{1}{h^2}E\left[m^+(X_1)u_1d_1\right]E\left[K(\frac{X_1}{h})\right]$ we have from the properties of $\hat{f}$ that $\frac{1}{h}E\left(K\left(\frac{X_1}{h}\right)\right) \to f(0)$. Now, again changing variables,

$$\begin{aligned}\frac{1}{h}E\left(m^+(X_1)u_1d_1\right) &= \frac{1}{h}\int_0^\infty K\left(\frac{X_1}{h}\right)m^+(X_1)f(X_1)dX_1 + \frac{1}{h}\sum_{j=1}^{s+1}\frac{k_j}{w_j}\int_0^\infty K\left(\frac{X_1}{w_jh}\right)m^+(X_1)f(X_1)dX_1 \\ &= I_1 + \sum_{j=1}^{s+1}\frac{k_j}{w_j}I_{2j}\end{aligned}$$

where $I_1 = \int_{-\infty}^{0} K(u)m^+(-hu)f(-hu)du$ and $I_{2j} = \int_0^\infty K(u)m^+(w_jhu)f(w_jhu)du$. Using, the same arguments as in the first case ($x > 0$) we have,

$$\left|I_1 - m^+(0)f(0)\int_{-\infty}^{0} K(u)du\right| \le C\int_{-\infty}^{-C} K(u)du + \bar{p}(Ch,0)\int_0^C K(u)du.$$

For h be sufficiently small and continuity of f we have, for all $\epsilon > 0$, $\left|I_1 - m^+(0)f(0)\int_{-\infty}^{0} K(u)du\right| \le \epsilon$.

$$\begin{aligned}\left|I_2 - m^+(0)f(0)\int_0^\infty K\left(\frac{u}{w_j}\right)du\right| &= \left|\int_0^\infty K\left(\frac{u}{w_j}\right)\left[m^+(hu)f(hu) - m^+(0)f(0)\right]du\right| \\ &= \left|\int_0^C K\left(\frac{u}{w_j}\right)\left[m^+(hu)f(hu) - m^+(0)f(0)\right]du\right. \\ &\quad \left. + \int_C^\infty K\left(\frac{u}{w_j}\right)\left[m^+(hu)f(hu) - m^+(0)f(0)\right]du\right| \\ &\le \bar{p}(hC,0)\int_0^C \left|K\left(\frac{u}{w_j}\right)\right|du + C\int_C^\infty K\left(\frac{u}{w_0}\right)K\left(\frac{u}{w_j}\right)du\end{aligned}$$

where for all $\epsilon > 0$, $\left|\int_C^\infty K\left(\frac{u}{w_j}\right)du\right| < \epsilon$ and for sufficiently small h, $\bar{p}(hC, 0) < \epsilon$. Thus,

$$\left|I_2 - m^+(0)f(0)\int_0^\infty K\left(\frac{u}{w_j}\right)du\right| < \epsilon.$$

Consequently,

$$\left|\frac{1}{h}E\left(m^+(X_1)u_1\right) - m^+(0)f(0)\sum_{j=0}^{s+1}\frac{k_j}{w_j}\int_0^\infty K\left(\frac{u}{w_j}\right)du\right| < \epsilon. \tag{37}$$

Thus for $x = 0$, $Cov(\hat{g}(x), \hat{f}(x)) = \frac{1}{nh}T_1 - \frac{1}{n}T_2 = \frac{1}{nh}\left[m^+(0)f(0)\sum_{,j=0}^{s+1}\frac{k_j}{w_j}\int_0^\infty K(u)K\left(\frac{u}{w_j}\right)du + o(1)\right]$. In summary, we have

$$Cov(\hat{g}(x), \hat{f}(x)) = \begin{cases} \frac{1}{nh}\left\{m^+(0)f(0)\sum_{,j=0}^{s+1}\frac{k_j}{w_j}\int_0^\infty K(u)K\left(\frac{u}{w_j}\right)du + o(1)\right\}, & x = 0 \\ \frac{1}{nh}\left\{m^+(x)f(x)\int_{\mathbb{R}}K^2(u)du + o(1)\right\}, & x > 0. \end{cases}$$

□

Lemma 2. *Under the assumptions of Theorem 3,* $E(|Z_n(x)|) = O\left((\frac{1}{nh})^{\frac{3}{2}}\right)$.

Proof.

$$Z_n(x) = 3(\hat{g}(x) - E\hat{g}(x))(\hat{f}(x) - E\hat{f}(x))^2\int_0^1\frac{(1-t)^2}{[E\hat{f}(x) + t(\hat{f}(x) - E\hat{f}(x))]^3}dt$$
$$- 3(\hat{f}(x) - E\hat{f}(x))^3\int_0^1(1-t)^2\frac{[E\hat{g}(x) + t(\hat{g}(x) - E\hat{g}(x))]}{[E\hat{f}(x) + t(\hat{f}(x) - E\hat{f}(x))]^4}dt = 3(J_1 - J_2).$$

Letting, $s_n = E(\hat{g}(x) - E\hat{g}(x)|X_1\cdots, X_n) = \frac{1}{nh}\sum_{i=1}^n(z_{i,n} - E(z_{i,n}))$ where $z_{i,n} = K\left(\frac{X_i - x}{h}\right)I_{\{X_i\geq 0\}}m^+(x_i)$, we have that

$$E(J_1) = \int s_n(\hat{f}(x) - E\hat{f}(x))^2\int_0^1\frac{(1-t)^2}{[E\hat{f}(x) + t(\hat{f}(x) - E\hat{f}(x))]^3}dtf(X)dX$$

and

$$E(J_2) = (\hat{f}(x) - E\hat{f}(x))^3E(\hat{g}(x))\int_0^1\frac{(1-t)^2}{[E\hat{f}(x) + t(\hat{f}(x) - E\hat{f}(x))]^4}dtf(X)dX$$
$$+ s_n(\hat{f}(x) - E\hat{f}(x))^3\int_0^1\frac{t(1-t)^2}{[E\hat{f}(x) + t(\hat{f}(x) - E\hat{f}(x))]^4}dtf(x)dx.$$

By the Cauchy-Schwartz inequality

$$|EJ_1| \leq \left(\int s_n^2(\hat{f}(x) - E\hat{f}(x))^4f(X)dX\right)^{\frac{1}{2}}\left(\int\left(\int_0^1\frac{(1-t)^2}{[E\hat{f}(x) + t(\hat{f}(x) - E\hat{f}(x))]^3}dt\right)^2f(X)dX\right)^{\frac{1}{2}}.$$

Now, since $(1-t)^2 \leq 1$ for $0 \leq t \leq 1$, letting $d = (1-t)E\hat{f}(x) + t\hat{f}(x)$ we see that $d > 0$ since $E(\hat{f}(x)) \geq 0$. Consequently, $\int_0^1\frac{1}{d^3}(1-t)^2dt \leq \int_0^1\frac{1}{d^3}dt$. But, $\int_0^1\frac{1}{d^3}dt = \frac{\hat{f}(x) + E\hat{f}(x)}{2\hat{f}(x)^2\left(E\hat{f}(x)\right)^2} = \frac{1}{2\hat{f}(x)E\hat{f}(x)^2} + \frac{1}{2\hat{f}(x)^2E\hat{f}(x)}$. Now, since $0 < B \leq E\hat{f}(x)$ we have $\int_0^1 \frac{1}{d^3} dt \leq \frac{1}{B} \max\left\{\frac{1}{\hat{f}(x)^2 B}, \frac{1}{\hat{f}(x)^2}\right\}$. Taking $\max\left\{\frac{1}{\hat{f}(x)B}, \frac{1}{\hat{f}(x)}\right\} = \frac{1}{\hat{f}(x)B}$ we have

$$|EJ_1| \leq \left(\int s_n^2(\hat{f}(x) - E\hat{f}(x))^4 f(X)dX\right)^{\frac{1}{2}} \left(\frac{1}{B^2}\int \left(\frac{1}{\hat{f}(x)}\right)^2 f(X)dX\right)^{\frac{1}{2}}.$$

Now,

$$\begin{aligned}
|E(J_2)| &\leq |J_{2,1}| + |J_{2,2}| \\
&\leq \left(\int \left((\hat{f}(x) - E\hat{f}(x))^3 E\hat{g}(x)\right)^2 f(X)dX\right)^{\frac{1}{2}} \left(\int \left(\int_0^1 \frac{(1-t)^2}{[E\hat{f}(x) + t(\hat{f}(x) - E\hat{f}(x))]^4} dt\right)^2 f(X)dX\right)^{\frac{1}{2}} \\
&+ \left(\int \left(s_n(\hat{f}(x) - E\hat{f}(x))^3\right)^2 f(X)dX\right)^{\frac{1}{2}} \left(\int \left(\int_0^1 \frac{t(1-t)^2}{[E\hat{f}(x) + t(\hat{f}(x) - E\hat{f}(x))]^4} dt\right)^2 f(X)dX\right)^{\frac{1}{2}}
\end{aligned}$$

Now, $t(1-t)^2 \leq t \leq 1$, hence

$$\begin{aligned}
\int_0^1 \frac{1}{[E\hat{f}(x) + t(\hat{f}(x) - E\hat{f}(x))]^4} dt &= \frac{(E\hat{f}(x))^2 + E\hat{f}(x)\hat{f}(x) + \hat{f}(x)^2}{3E\hat{f}(x)^3\hat{f}(x)^3} = \frac{1}{3E\hat{f}(x)\hat{f}(x)^3} + \frac{1}{3E\hat{f}(x)^2\hat{f}(x)^2} \\
&+ \frac{1}{3E\hat{f}(x)^3\hat{f}(x)} \\
&\leq \frac{1}{B} max\left\{\frac{1}{\hat{f}(x)^3}, \frac{1}{B\hat{f}(x)^2}, \frac{1}{B^2\hat{f}(x)}\right\}
\end{aligned}$$

Suppose $\max\left\{\frac{1}{\hat{f}(x)^3}, \frac{1}{B\hat{f}(x)^2}, \frac{1}{B^2\hat{f}(x)}\right\} = \frac{1}{\hat{f}(x)B^2}$. Then,

$$\begin{aligned}
|EJ_2| &\leq \left\{\left(\int \left((\hat{f}(x) - E\hat{f}(x))^3 E\hat{g}(x)\right)^2 f(x)dx\right)^{\frac{1}{2}} + \left(\int \left(s_n(\hat{f}(x) - E\hat{f}(x))^3\right)^2 f(X)dX\right)^{\frac{1}{2}}\right\} \\
&\times \left(\frac{1}{B^4}\int \left(\frac{1}{\hat{f}(x)}\right)^2 f(X)dX\right)^{\frac{1}{2}}
\end{aligned}$$

Now, $\int \left(\frac{1}{\hat{f}(x)}\right)^2 f(X)dX = \int_0^\infty \int e^{-\lambda \hat{f}(x)^2} f(X)dXd\lambda = \int_0^\infty E\left(e^{-\lambda \hat{f}(x)^2}\right)d\lambda$. Under the conditions of Theorem 3, and by Slutsky Theorem, $\hat{f}(x)^2 \xrightarrow{p} f(x)^2$. Thus, by Lebesgue's dominated convergence Theorem

$$h_n(x, \lambda) = E\left(e^{-\lambda \hat{f}(x)^2}\right) \to E\left(e^{-\lambda f(x)^2}\right) = \left(e^{-\lambda f(x)^2}\right) = h(x, \lambda).$$

since $|h_n(x, \lambda)| \leq \int |e^{-\lambda \hat{f}(x)^2}| f(x)dx \leq 1$ for all n, λ. Thus, $h_n(x, \lambda)$ is bounded and convergent on $[0, \infty)$. Then, by Arzèla's Theorem (Apostol, 1974, p. 228)

$$\lim_{n\to\infty} \int_0^\infty h_n(x, \lambda)d\lambda \to \int_0^\infty E\left(e^{-\lambda \hat{f}(x)^2}\right)d\lambda = \frac{1}{f(x)^2} \leq C$$

since $0 < B < f(x)$. Thus,

$$|E(J_1) + E(J_2)| \leq C \left(\left(\int s_n^2 (\hat{f}(x) - E\hat{f}(x))^4 f(X) dX \right)^{\frac{1}{2}} + \left(\int \left((\hat{f}(x) - E\hat{f}(x))^3 E\hat{g}(x) \right)^2 f(X) dX \right)^{\frac{1}{2}} \right.$$
$$\left. + \left(\int \left(s_n (\hat{f}(x) - E\hat{f}(x))^3 \right)^2 f(x) dx \right)^{\frac{1}{2}} \right)$$

Now, $E|\hat{f} - E\hat{f}|^3 \leq \left(E|\hat{f} - E\hat{f}|^2 \right)^{\frac{1}{2}} \left(E|\hat{f} - E\hat{f}|^4 \right)^{\frac{1}{2}} = \left(V\hat{f} \right)^{\frac{1}{2}} \left(E \left(\hat{f} - E\hat{f} \right)^4 \right)^{\frac{1}{2}} = O((nh)^{-1})^{\frac{1}{2}} (O((nh)^{-2}))^{\frac{1}{2}}$ by Hölder's Inequality and the fact that $E \left(\hat{f} - E\hat{f} \right)^4 = O((nh)^{-2})$ from Ziegler (2001). Then, $E|\hat{f} - E\hat{f}|^3 \leq (nh)^{-\frac{1}{2}} (nh)^{-1} O(1) = (nh)^{-\frac{3}{2}} O(1)$, $E\hat{g}(x) = O(1)$ and $s_n = (\hat{g}(x) - E(\hat{g}(x)|X_1, \cdots, X_n) = O_p((nh)^{-\frac{1}{2}}$. Thus, $E(Z_{n(}(x)) = O\left((\frac{1}{nh})^{\frac{3}{2}} \right)$. □

References

Apostol, T., 1974. Mathematecal Analysis. Addison Wesley, New York, NY.

Cattaneo, M., Escanciano, J. C. (Eds.), 2017. Regression discontinuity designs: theory and applications. Vol. 38 of Advances in Econometrics. Emerald Group Publishing.

Cheng, M.-Y., 1994. On boundary effects of smooth curve estimators. Ph.D. thesis, Department of Statistics, University of North Carolina at Chapel Hill.

Fan, J., 1992. Design-adaptive nonparametric regression. Journal of the American Statistical Association 87, 998–1004.

Fan, J., 1993. Local linear regression smoothers and their minimax efficiencies. The Annals of Statistics 21, 196–216.

Fan, J., Gijbels, I., 1992. Variable bandwidth and local linear regression smoothers. The Annals of Statistics 20, 2008–2036.

Gasser, T., , Muller, H.-G., 1979. Smoothing Techniques for Curve Estimation. Springer, Heidelberg, Ch. Kernel estimation of regression functions, pp. 23–68.

Gasser, T., Muller, H.-G., 1984. Estimating regression functions and their derivatives by the kernel method. Scandinavian Journal of Statistics 11, 171–185.

Hahn, J., Todd, P., Van der Klaauw, W., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69, 201–209.

Härdle, W., Linton, O. B., 1994. Handbook of Econometrics. Vol. 4. North Holland, Ch. Applied Nonparametric Methods, pp. 2295–2339.

Hestenes, M. R., 1941. Extension of the range of a differentiable function. Duke Mathematical Journal 8, 183–192.

Imbens, G., Kalyanaraman, K., 2012. Optimal bandwidth choice for the regression discontinuity estimator. The Review of Economic Studies 79, 933–959.

Imbens, G. W., Lemieux, T., 2008. Regression discontinuity designs: A guide to practice. Journal of Econometrics 142, 615–635.

Lee, D. S., Lemieux, T., 2010. Regression discontinuity designs in economics. Journal of Economic Literature 48, 281–355.

Li, Q., Racine, J. S., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press.

Litschig, S., Morrison, K. M., 2013. The impact of intergovernmental transfers on education outcomes and poverty reduction. American Economic Journal: Applied Economics 5, 206–240.

Matsudaira, J. D., 2008. Mandatory summer school and student achievement. Journal of Econometrics 142, 829–850.

McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics 142, 698–714.

Mynbaev, K., Martins-Filho, C., 2019. Unified estimation of densities on bounded and unbounded domains. Annals of the Institute of Statistical Mathematics (In press).

Porter, J., 2003. Estimation in the regression discontinuity model. Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison, 5–19.

Ziegler, K., 2001. On approximations to the bias of the nadaraya-watson regression estimator. Journal of Nonparametric Statistics 13, 583–589.