A new estimator of a jump discontinuity in regression

Carlos Martins-Filho[1]

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

Sihong Xie

Department of Economics
Seattle University
Seattle, WA , 98122 USA
email: sxie@seattleu.edu

and

Feng Yao

Department of Economics
West Virginia University
Morgantown, WV 26505-6025, USA
email: feng.yao@mail.wvu.edu
and
China Center for Special Economic Zone Research
Shenzhen University
Shenzhen, Guangdong 518060, China

May, 2022

Abstract. We propose a new class of estimators for a jump discontinuity on nonparametric regression. While there is a vast literature in econometrics that addresses this issue (e.g., Hahn et al., 2001; Porter, 2003; Imbens and Lemieux, 2008; Cattaneo and Escanciano, 2017), the main approach in these studies is to use local polynomial (linear) estimators on both sides of the discontinuity to produce an estimator for the jump that has desirable boundary properties. Our approach extends the regression from both sides of the discontinuity using a theorem of Hestenes (1941). The extended regressions are then estimated and used to construct an estimator for the jump discontinuity that solves the boundary problems normally associated with classical Nadaraya-Watson estimators. We provide asymptotic characterizations for the jump estimators, including bias and variance orders, and asymptotic distributions after suitable centering and normalization. Monte Carlo simulations show that our jump estimators can outperform those based on local polynomial (linear) regression.

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

JEL codes: C13, C14.

[1]Corresponding author.

1 Introduction

Regression discontinuity (RD) designs have been widely used in economics and other social and behavioral sciences. See Imbens and Lemieux (2008), Lee and Lemieux (2010) and Cattaneo and Escanciano (2017) for an overview of the literature. From a statistical perspective, the estimation of a jump discontinuity in regression is made difficult by the fact that traditional nonparametric kernel regression estimators, such as Nadaraya-Watson (NW), suffer from boundary problems (see, 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 regression and regressor density (e.g., twice continuous differentiability), the NW estimator constructed with bandwidth $h > 0$ has bias of order $O(h)$ at boundary points, compared to $O(h^2)$ at interior points. Porter (2003) observes that the problem can be aggravated in RD designs as an estimator for a jump discontinuity may compound the poor bias behavior of nonparametric estimators of the regression to the right and to left of the point of discontinuity.

While there is a vast literature in econometrics and statistics that attempts to address this issue, (see, 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 linear) approximations for the regression on both sides of the discontinuity and use these to produce an estimate for the jump at the point of discontinuity. This approach is justified by Fan (1992) where it is shown that local linear estimators, under standard smoothness assumptions, have bias of order $O(h^2)$ at interior and boundary points.

In this paper, we adopt a novel approach. The idea behind our estimation procedure is to extend regression segments from the left and from the right of the point of discontinuity to the entire real line using an extension proposed by Hestenes (1941). These extended regressions are then estimated and used to estimate the jump at the point of discontinuity. Regression segments can be different and, in particular, can have different degrees of smoothness. We are inspired 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 estimation strategy produces a *class* of jump discontinuity estimators indexed by the type of Hestenes' extension. Our estimators are simple to construct, retaining an algebraic structure that mimics that of the classical NW estimator. However, contrary to the NW estimator, our estimators have boundary behavior that is completely analogous to that at interior points of the regression domain.

This paper is organized as follows: section 2 discusses Hestenes' extension and introduces the new class of estimators; section 3 provides their asymptotic distribution and compares it to that of estimators for a jump discontinuity based on local linear regression estimators; section 4 contains Monte Carlo simulations; section 5 is a conclusion. The main theorem's proof is given in an appendix.

Throughout the paper, for $S \subseteq \mathbb{R}$ we define the class of bounded continuous functions $g : S \to \mathbb{R}$ by $\mathcal{C}_b^0(S)$. The derivative of order $s \in \mathbb{N}$ of g at $x \in S$ is denoted by $g^{(s)}(x)$ and $\mathcal{C}_b^s(S) := \{g : S \to \mathbb{R} : g^{(s)} \in \mathcal{C}_b^0(S)\}$ with $g^{(0)} := g$. The righthand derivative of order s of g at x is denoted by $g^{(s)}(x+)$ and its lefthand derivative of order s at x by $g^{(s)}(x-)$. I_S is the indicator function for the set S. The letter C represents a positive real constant that varies with the context.

2 Hestenes' extension and a new class of estimators

Let Y and X be random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ with $E(Y^2) < \infty$ and consider the regression $E(Y|X) = \mu(X)$ for some $\mu : \mathbb{R} \to \mathbb{R}$ measurable. We assume that $\lim_{x\downarrow 0} \mu(x) = \mu(0)$, $\lim_{x\uparrow 0} \mu(x) = \mu(0-) \in \mathbb{R}$ and define a jump discontinuity at 0 by

$$J_0 := \mu(0) - \mu(0-). \tag{2.1}$$

Our objective is to estimate J_0 given a random sample $\{(Y_i, X_i)\}_{i=1}^n$ of size $n \in \mathbb{N}$.[1]

Based on equation (2.1), most of the existing literature, e.g. Porter (2003), Hahn et al. (2001), Imbens and Lemieux (2008), Imbens and Kalyanaraman (2012), constructs estimators for J_0 by taking the difference between two nonparametric regression estimators at $x = 0$. For example, Hahn et al. (2001) propose

$$\check{J}_0 = \check{\mu}(0) - \check{\mu}(0-)$$

where $\check{\mu}(0)$ and $\check{\mu}(0-)$ are local linear estimators constructed using $\{(Y_i, X_i)\}_{\{i:X_i \geq 0\}}$ and $\{(Y_i, X_i)\}_{\{i:X_i<0\}}$, respectively. The widespread use of $\check{J}_0$ derives from the desirable boundary properties of local linear estimators such as $\check{\mu}(0)$ and $\check{\mu}(0-)$. In what follows we propose new estimators for J_0 based on extensions of continuously differentiable functions from bounded to unbounded domains first introduced by Hestenes (1941).

2.1 Hestenes' extension

Consider $g \in \mathcal{C}_b^s([0,\infty))$ for $s \in \{0\} \cup \mathbb{N}$. Hestenes (1941) showed that g can be extended from $[0, \infty)$ to $\mathbb{R}$ and that the extensions are elements of $\mathcal{C}_b^s(\mathbb{R})$. Formally, let $\{w_i\}_{i=1}^{s+1}$ be a sequence of positive, pairwise distinct numbers, e.g., $w_i = 1/i$ used by Hestenes (1941) or $w_i = i$. Let $\{k_i\}_{i=1}^{s+1}$ be defined by the system of equations

$$\sum_{i=1}^{s+1} (-w_i)^j k_i = 1, \quad \text{for } j = 0, \cdots, s. \tag{2.2}$$

Since this system has the Vandermonde determinant

$$\det \begin{pmatrix} 1 & 1 & ... & 1 \\ -w_1 & -w_2 & ... & -w_{s+1} \\ ... & ... & ... & \\ (-w_1)^s & (-w_2)^s & ... & (-w_{s+1})^s \end{pmatrix} \neq 0,$$

$\{k_i\}_{i=1}^{s+1}$ is uniquely determined by $\{w_i\}_{i=1}^{s+1}$. The Hestenes extension of g to $(-\infty, 0)$ is given by

$$\phi_s(x) = \sum_{j=1}^{s+1} k_j g(-w_j x) \text{ for } x < 0. \tag{2.3}$$

Due to equation (2.2) we have the following "sewing" conditions

$$\phi_s^{(m)}(0-) = \sum_{j=1}^{s+1} (-w_j)^m k_j g^{(m)}(0+) = g^{(m)}(0+) \text{ for } m = 0, 1, \cdots, s.$$

We define g_s on $\mathbb{R}$ by

$$g_s(x) = \begin{cases} g(x) & \text{for } x \geq 0 \\ \phi_s(x) & \text{for } x < 0 \end{cases}, \tag{2.4}$$

[1] The assumption that the discontinuity occurs at 0 is made without loss of generality. If it occurs at $x_d \neq 0$, it suffices to shift μ by defining $\bar{\mu}(z) := \mu(z + x_d)$ where $z = x - x_d$ with discontinuity at $z = 0$.

and note that g_s is s times differentiable at all $x \in \mathbb{R}$.[2]

Now, let $\mu(X)$ be a regression where the density f of X exists and has support $[0,\infty)$, i.e. $X \geq 0$, with $f(0) > 0$. For motivation, consider the infeasible – in that $f(0)$ is assumed to be known – NW estimator for $\mu(0)$ given by

$$\hat{\mu}(0) = \frac{1}{f(0)}\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i}{h}\right) Y_i,$$

based on a random sample $\{(Y_i, X_i)\}_{i=1}^n$, a kernel K and bandwidth $0 < h \downarrow 0$ as $n \to \infty$. For K even, we have

$$E(\hat{\mu}(0)) = \frac{1}{f(0)}\int_0^\infty \mu(uh)f(uh)K(u)du$$

and, in general, $\lim_{n\to\infty} E(\hat{\mu}(0)) \neq \mu(0)$. This asymptotic bias at 0 is a manifestation of the well known boundary bias problem associated with NW type estimators. In our context, the main idea is to construct an estimator $\tilde{\mu}(0)$ that does not carry this bias and whose expectation admits an integral representation that uses a Hestenes' extension of μf from $[0,\infty)$ to $\mathbb{R}$. For this purpose, consider

$$\tilde{\mu}(0) = \frac{1}{f(0)}\frac{1}{nh}\sum_{i=1}^{n} K_H\left(\frac{X_i}{h}\right) Y_i, \tag{2.5}$$

where $K_H(x) = K(x) + \sum_{j=1}^{s+1}\frac{k_j}{w_j}K\left(\frac{x}{w_j}\right)$, and w_j, k_j and s are as in equation (2.2). For K even, we obtain

$$\begin{aligned} E(\tilde{\mu}(0)) &= \frac{1}{f(0)}\left(\int_0^\infty \mu(uh)f(uh)K(u)du + \int_{-\infty}^0 \sum_{j=1}^{s+1} k_j\mu(-w_juh)f(-w_juh)K(u)du\right) \\ &= \frac{1}{f(0)}\int_{\mathbb{R}} g_s(uh)K(u)du \end{aligned}$$

where

$$g_s(x) = \begin{cases} \mu(x)f(x) & \text{for } x \geq 0 \\ \sum_{j=1}^{s+1} k_j\mu(-w_jx)f(-w_jx) & \text{for } x < 0 \end{cases} \tag{2.6}$$

is constructed using a Hestenes' extension of μf from from $[0,\infty)$ to $\mathbb{R}$. Then, the bias of $\tilde{\mu}(0)$ has integral representation

$$E(\tilde{\mu}(0)) - \mu(0) = \frac{1}{f(0)}\int_{\mathbb{R}} (g_s(uh) - g_s(0))K(u)du. \tag{2.7}$$

The smoothness properties that g_s inherits from those of f and μ can be explored to establish asymptotic unbiasedness at 0 and the order of bias decay as a function of h. For example, if $\mu f \in \mathcal{C}_b^s([0,\infty))$, K is such that $\int_{\mathbb{R}} u^j K(u)du = 0$ for $j = 1, \cdots, s-1$ and $\int_{\mathbb{R}} |u|^s |g_s^{(s)}(\tau uh)|du < C$ for some $\tau \in (0,1)$ and $C < \infty$ we have

$$|E(\tilde{\mu}(0)) - \mu(0)| \leq \frac{h^s}{s!}\frac{1}{f(0)}\int_{\mathbb{R}} |u|^s |g_s^{(s)}(\tau uh)|du \leq h^s C \to 0 \text{ as } n \to \infty.$$

The asymptotic bias at 0 is eliminated and its order of decay, viz., h^s is easily obtained. We use the insight gained from using Hestenes' extension to define a class of estimators for the jump J_0.

2.2 A class of estimators for J_0

We propose the following Hestenes' based estimator for J_0:

$$\tilde{J}_0 := \tilde{\mu}(0) - \tilde{\mu}(0-) = \frac{1}{\tilde{f}_+(0)}\frac{1}{nh}\sum_{i=1}^{n} K_H\left(\frac{X_i}{h}\right) I_{\{X_i \geq 0\}} Y_i - \frac{1}{\tilde{f}_-(0)}\frac{1}{nh}\sum_{i=1}^{n} K_H\left(\frac{X_i}{h}\right) I_{\{X_i < 0\}} Y_i,$$

[2] Note that since g_s depends on $\{w_j\}_{j=1}^{s+1}$ a more explicit notation would be $g_s\left(x; \{w_j\}_{j=1}^{s+1}\right)$. However, for simplicity we write $g_s(x)$ and leave implicit its dependence on $\{w_j\}_{j=1}^{s+1}$.

where $\tilde{f}_+(0) = \frac{1}{nh}\sum_{i=1}^n K_H\left(\frac{X_i}{h}\right) I_{\{X_i\geq 0\}}$ and $\tilde{f}_-(0) = \frac{1}{nh}\sum_{i=1}^n K_H\left(\frac{X_i}{h}\right) I_{\{X_i<0\}}$. We note that $\tilde{J}_0$ belongs to a *class* of estimators whose elements are indexed by $\{w_j\}_{j=1}^{s+1}$. Hence, for an assumed s, the choice of $\{w_j\}_{j=1}^{s+1}$ indexes an element of the class. The estimators $\tilde{\mu}(0)$ and $\tilde{\mu}(0-)$ retain the simple algebraic structure of NW estimators but with a new "Hestenes kernel" K_H which depends on $\{w_j\}_{j=1}^{s+1}$. Also, note that when $s = 0$, K_H is the reflection kernel used by Schuster (1985) in density estimation in domains with boundaries.

3 Asymptotic properties of $\tilde{J}_0$

We make the following assumptions to establish convergence in distribution of $\tilde{J}_0$. We start with typical restrictions on the kernel.

Assumption 3.1. *The kernel $K : \mathbb{R} \to \mathbb{R}$ is an even function satisfying: 1. $|K(x)| < C$ for any $x \in \mathbb{R}$; 2. $\int_{\mathbb{R}} K(u)du = 1$; 3. $\int_{\mathbb{R}} u^m K(u)du = 0$ for $m = 1, \cdots, s-1$, $\int_{\mathbb{R}} |u|^s|K(u)|du < C$ and $|u|^{s+1}|K(u)| \to 0$ as $u \to \infty$ for $s \in \{2, 4, \cdots\}$; 4. $\int_{\mathbb{R}} |K(u)|^{2+\delta}du < C$ for some $\delta > 0$.*

In the existing literature, it is frequently assumed that $s = 2$. Different form Porter (2003) or Imbens and Kalyanaraman (2012) we do not restrict the kernels to be compactly supported, nor is there any requirement that they satisfy a Lipschitz condition. However, we retain the assumption that the kernels be uniformly bounded and even functions.

Assumption 3.2. *1. The sequence $\{(Y_i, X_i)\}_{i\in\mathbb{N}}$ is independently and identically distributed (IID) as (Y, X).*

2. *The marginal density $f : \mathbb{R} \to [0, \infty)$ of X exists and $f(0) > 0$, $f \in \mathcal{C}_b^s(\mathbb{R})$, $\int_{\mathbb{R}} |f^{(\ell)}(x)|dx < C$ for $\ell = 1, \cdots, s$ and some $s \in \mathbb{N}$.*

3. *$E(Y^2) < C$ and $E(Y|X) = \mu(X)$ has a jump discontinuity J_0 at $x = 0$. We let $\mu(x) = \mu_u(x) + \mu_l(x)$ for $x \in \mathbb{R}$ where $\mu_u(x) = \begin{cases} \mu(x) & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$, $\mu_l(x) = \begin{cases} 0 & \text{if } x \geq 0 \\ \mu(x) & \text{if } x < 0 \end{cases}$ and assume that $\mu_u \in \mathcal{C}_b^s([0, \infty))$ and $\mu_l \in \mathcal{C}_b^s((-\infty, 0))$.*

4. *$E((Y - \mu(X))^2|X) = \mathcal{V}(X) > 0$ with $\mathcal{V} \in \mathcal{C}_b^0(\mathbb{R})$ and $E((Y - \mu(X))^{2+\delta}|X) < C$ for some $\delta > 0$.*

The assumption of integrability $f^{(\ell)}$ for $\ell = 1, \cdots, s$ is new relative to the extant literature. It emerges because of the structure of the Hestenes' based estimator. Although theoretically restrictive, it is in many empirical settings not necessarily binding.

The assumption that $\mathcal{V} \in \mathcal{C}_b^0(\mathbb{R})$ is not necessary to prove the following Theorems 3.1. The conditional variance, as in Porter (2003) or Imbens and Kalyanaraman (2012), could also exhibit a discontinuity at 0 without fundamentally changing our theorem. We assume continuity of $\mathcal{V}$ for simplicity and because discontinuity of the skedastic function does not materially impact the estimation of J_0. The other restrictions on Assumption 3.2 are standard in the RD design literature.

Theorem 3.1. *Suppose $0 < h \to 0$, $nh \to \infty$ and $nh^{2s+1} = O(1)$. Then, under assumptions 3.1 and 3.2,*

$$\sqrt{nh}\left(\tilde{J}_0 - J_0 - \left(\frac{h^s}{s!}B_s\mu_{K,s} + o_p(h^s)\right)\right) \xrightarrow{d} \mathcal{N}\left(0, 2\frac{\mathcal{V}(0)}{f(0)}\int_0^\infty K_H^2(u)du\right),$$

where $B_s = \frac{1}{f(0)}\left(\sum_{q=1}^s \begin{pmatrix} s \\ q \end{pmatrix} [\mu^{(q)}(0+) - \mu^{(q)}(0-)]f^{(s-q)}(0)\right)$ and $\mu_{K,s} = \int_{\mathbb{R}} u^s K(u)du$.

We note that the approximation for the bias of $\tilde{J}_0$, i.e., $\frac{h^s}{s!}B_s\mu_{K,s}$ depends on the density f. Hence, our estimators lack design-adaptability, in contrast with the approximation for the bias of $\check{J}_0$ which depends only on μ (see Porter, 2003). However, it is not possible to directly compare the magnitude of the absolute value of the biases of $\tilde{J}_0$ and $\check{J}_0$ without considering specific μ and f. The use of higher order kernels provides automatic increase in the rate of bias decay for $\tilde{J}_0$.

It is possible to compare the magnitude of the variance of the asymptotic distribution in Theorem 3.1 with that of $\check{J}_0$ in Theorem 3 (a) in Porter (2003) when $p = 1$ (local linear estimator) and $s = 2$. Note that under continuity of $\mathcal{V}$ we need only compare $I_K\left(s, \{w_j\}_{j=1}^{s+1}\right) := \int_0^\infty K_H^2(u)\,du$ when $s = 2$ to Porter's constant $P_K = \begin{pmatrix} 1 & 0 \end{pmatrix}\Gamma^{-1}\Delta\Gamma^{-1}\begin{pmatrix} 1 & 0 \end{pmatrix}^T$, where $\Gamma = \begin{pmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_2 \end{pmatrix}$, $\Delta = \begin{pmatrix} \delta_0 & \delta_1 \\ \delta_1 & \delta_2 \end{pmatrix}$ with $\gamma_j = \int_0^\infty u^j K(u)du$ and $\delta_j = \int_0^\infty u^j K^2(u)du$ for $j = 0, 1, 2$. Table 1 gives the values of I_K when $w_j = j$ and P_K for the Gaussian, Epanechnikov and Triangular/Edge kernels. We note that for the Triangular/Edge kernel, shown by Cheng et al. (1997) to have

Table 1: $I_K\left(2, \{w_j\}_{j=1}^3\right)$ and P_K for Gaussian, Epanechnikov, Triangular/Edge kernels.

	Gaussian	Epanechnikov	Triangular/Edge
$I_K\left(2, \{w_j\}_{j=1}^3\right)$	1.8507	4.9167	4.6667
P_K	1.7860	4.4980	4.8000

optimal properties for boundary estimation problems, $I_K\left(2, \{w_j\}_{j=1}^3\right) < P_K$, leading to a smaller variance for the asymptotic distribution of our proposed $\tilde{J}_0$ relative to the widely used $\check{J}_0$.

It follows directly from the theorem that the mean squared error (MSE) of $\tilde{J}_0$, denoted by $MSE\left(\tilde{J}_0; K, h, \{w_i\}_{i=1}^{s+1}\right)$, is given by

$$\begin{aligned} MSE\left(\tilde{J}_0; K, h, \{w_j\}_{j=1}^{s+1}\right) &= \frac{h^{2s}}{(s!)^2}B_s^2\mu_{K,s}^2 + \frac{1}{nh}2\frac{\mathcal{V}(0)}{f(0)}I_K\left(s, \{w_j\}_{j=1}^{s+1}\right) + o_p\left(h^{2s} + (nh)^{-1}\right) \\ &= \text{AMSE}\left(\tilde{J}_0; K, h, \{w_j\}_{j=1}^{s+1}\right) + o_p\left(h^{2s} + (nh)^{-1}\right). \end{aligned}$$

For given s, K, and $\{w_j\}_{j=1}^{s+1}$ we define

$$h^* := \underset{h}{\operatorname{argmin}}\, \text{AMSE}\left(\tilde{J}_0; K, h, \{w_j\}_{j=1}^{s+1}\right)$$

and by routine optimization we obtain

$$h^* = n^{-\frac{1}{2s+1}}\left(\frac{s!(s-1)!}{2}\right)^{\frac{1}{2s+1}}(B_s\mu_{K,s})^{-\frac{2}{2s+1}}\left(2\frac{\mathcal{V}(0)}{f(0)}I_K\left(s, \{w_j\}_{j=1}^{s+1}\right)\right)^{\frac{1}{2s+1}}. \qquad (3.1)$$

In practice, the use of h^* depends on obtaining estimates of $\mathcal{V}(0)$, $f(0)$, $\mu^{(q)}(0+)$, $\mu^{(q)}(0-)$ and $f^{(s-q)}(0)$ for $q = 1, \cdots, s$ that can be used in equation (3.1) to produce a "plug in" bandwidth. Since $f \in \mathcal{C}_b^s(\mathbb{R})$, several consistent estimators for $f(0)$ and $f^{(s-q)}(0)$ can be obtained. See, e.g., Härdle et al. (1990), Park and Marron (1990), Sheather and Jones (1991) and Wand and Jones (1995). Consistent estimators for $\mathcal{V}(0)$, $\mu^{(q)}(0+)$ and $\mu^{(q)}(0-)$ when $s = 2$ and $q = 1, 2$ can be obtained as in section 4.2 of Imbens and Kalyanaraman (2012).

Lastly, we note that Theorem 3.1 makes clear that the asymptotic distribution of the estimators $\tilde{J}_0$ depends on $\{w_j\}_{j=1}^{s+1}$. Conceptually, a suitably defined loss function can be defined to guide the choice of $\{w_j\}_{j=1}^{s+1}$ and,

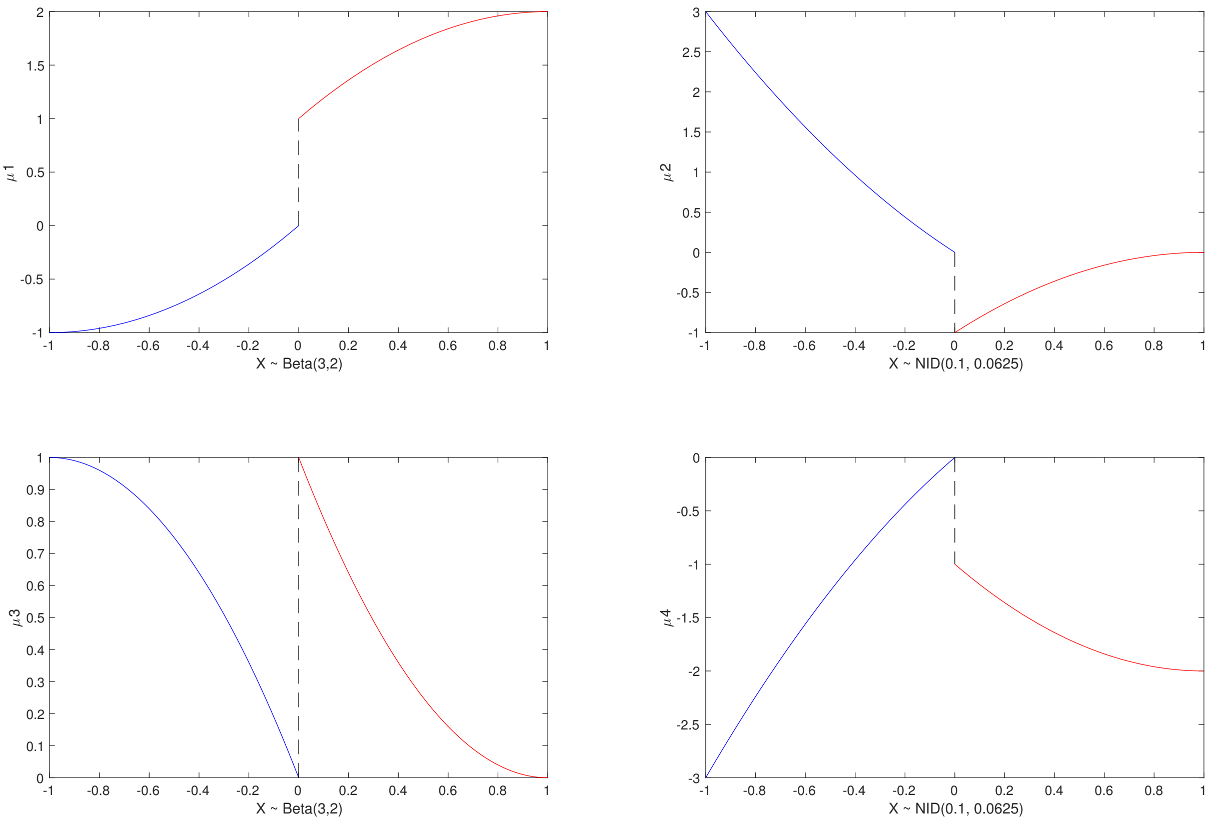

Figure 1: Regressions μ_1, μ_2, μ_3, μ_4 with a discontinuity at $x = 0$

by consequence, $\tilde{J}_0$. We leave this optimal theoretical choice of $\{w_j\}_{j=1}^{s+1}$ for future work. However, as shown by the Monte Carlo simulations in the next section, even a simple choice of $w_j = j$ can lead to an estimator $\tilde{J}_0$ that outperforms the widely used $\check{J}_0$ in terms of root mean squared error.

4 Monte Carlo study

We investigate some of the finite sample properties of our jump estimator $\tilde{J}_0$ and compare them to those of the widely used local linear regression jump estimator $\check{J}_0$. We consider four different regressions $\mu_j(x) = \mu_{ju}(x) + \mu_{jl}(x)$ for $j = 1, \cdots, 4$. The jump discontinuity occurs at $x = 0$ with $\mu_{ju}(x) = 0$ for $x < 0$ and $\mu_{jl}(x) = 0$ for $x \geq 0$. Specifically,

$$\begin{array}{lll} \mu_1(x): & \mu_{1l}(x) = ((x+1)^2 - 1)I_{\{x<0\}}, & \mu_{1u}(x) = (-(x-1)^2 + 2)I_{\{x\geq 0\}} \\ \mu_2(x): & \mu_{2l}(x) = ((x-1)^2 - 1)I_{\{x<0\}}, & \mu_{2u}(x) = (-(x-1)^2)I_{\{x\geq 0\}} \\ \mu_3(x): & \mu_{3l}(x) = (-(x+1)^2 + 1)I_{\{x<0\}}, & \mu_{3u}(x) = ((x-1)^2)I_{\{x\geq 0\}} \\ \mu_4(x): & \mu_{4l}(x) = (-(x-1)^2 + 1)I_{\{x<0\}}, & \mu_{4u}(x) = ((x-1)^2 - 2)I_{\{x\geq 0\}}. \end{array}$$

Figure 1 contains the graphs of each of these regressions.

The data generating processes (DGP) use $Y_i = \mu_j(X_i) + \varepsilon_i$ for $i = 1, \cdots, n$. We set the conditional variance $\mathcal{V}(X) = 4$ and choose $\varepsilon_i \sim NID(0, 4)$. The regressor X_i is generated from two distributions: (a) $X_i \sim NID(0.1, 1/16)$, (b) $X_i = -1 + 2\xi_i$, where $\xi_i \sim BetaID(3, 2)$. The X_i generated using the Beta distribution takes values in $[-1, 1]$ as do most of the X_i generated by our choice of Normal. We note that our choice of distributions for X_i gives density derivatives different from zero at the point of discontinuity, allowing for a meaningful comparison of the estimators' biases.

Table 2: Bias (B), standard error (S), root mean squared error (R) for $\tilde{J}_0$ and $\check{J}_0$ and optimal bandwidth h using the Triangular/Edge kernel

μ	X	n	Estimator	B	S	R	h
μ_1	Beta	1000	$\tilde{J}_0$	0.221	0.235	0.322	0.630
			$\check{J}_0$	0.110	0.267	0.289	0.796
		2000	$\tilde{J}_0$	0.026	0.198	0.200	0.549
			$\check{J}_0$	0.086	0.202	0.220	0.693
μ_2	Normal	1000	$\tilde{J}_0$	-0.050	0.201	0.207	0.402
			$\check{J}_0$	0.051	0.247	0.252	0.696
		2000	$\tilde{J}_0$	-0.053	0.152	0.161	0.350
			$\check{J}_0$	0.053	0.172	0.180	0.606
μ_3	Beta	1000	$\tilde{J}_0$	-0.215	0.232	0.316	0.630
			$\check{J}_0$	-0.104	0.270	0.290	0.796
		2000	$\tilde{J}_0$	-0.034	0.200	0.203	0.549
			$\check{J}_0$	-0.098	0.202	0.225	0.693
μ_4	Normal	1000	$\tilde{J}_0$	0.046	0.192	0.197	0.402
			$\check{J}_0$	-0.055	0.234	0.240	0.696
		2000	$\tilde{J}_0$	0.058	0.149	0.159	0.350
			$\check{J}_0$	-0.049	0.168	0.175	0.606

We implement $\tilde{J}_0$ with $s = 2$ and use the Gaussian, Epanechnikov and Triangular/Edge kernels, the latter being frequently used in nonparametric regression discontinuity estimation (see, Imbens and Lemieux, 2008; Imbens and Kalyanaraman, 2012). We report results from using the the Triangular/Edge kernel. Results using the other kernels are qualitatively similar and are available upon request. We choose $w_j = j$ for $j = 1, 2, 3$ and use the optimal bandwidth h^* given in equation (3.1) in the implementation of $\tilde{J}_0$.

To implement $\check{J}_0$ we use the optimal bandwidth h_{opt} given in equation (7) of Imbens and Kalyanaraman (2012). For both $\tilde{J}_0$ and $\check{J}_0$ the optimal bandwidths are calculated using the true values for $f(0)$, $f^{(1)}(0)$, $\mathcal{V}(0)$, $\mu^{(q)}(0+)$, $\mu^{(q)}(0-)$ for $q = 1, 2$. Our goal is to compare the performance of these estimators without the potential noise that will be introduced by the estimation of the functionals that appear in the expressions for the optimal bandwidths. As mentioned above, an entirely data-driven procedure to calculate $\tilde{J}_0$ and $\check{J}_0$ can be easily implemented.

We generate 2000 samples of size $n = 1000, 2000$ and summarize in Table 2 the performance of the jump estimators at $x = 0$ by calculating their experimental bias (B), standard deviation (S) and root mean squared error (R). We also report the optimal bandwidth h for $\check{J}_0$ and $\tilde{J}_0$.

For both estimators, and all DGPs, as the sample size increases, performance improves in terms of smaller biases (with exceptions), variances, and root mean squared errors. Thus, unsurprisingly, the simulations support the consistency of both jump discontinuity estimators. Except for the design where $n = 1000$ and $X_i \sim BetaID(3, 2)$, $\tilde{J}_0$ has root mean squared error (R) which is smaller than that of $\check{J}_0$. This improved performance derives mostly from a smaller variance, as the absolute value for the magnitude of the bias for both estimators is quite small, relative to the variance. Once again, the simulations confirm the asymptotic results that indicate small leading terms for the bias.

As mentioned in the previous section, a theoretically guided (optimal) choice of $\{w_j\}_{j=1}^{s+1}$ may lead to even stronger improvements on finite sample experimental properties of $\tilde{J}_0$ over $\check{J}_0$, but our simulations already indicate that our estimator can outperform the popular $\check{J}_0$.[3]

5 Conclusion

This paper shows that Hestenes' extensions can be used to define estimators for a jump discontinuity in regression that has desirable asymptotic properties, is easy to calculate and can outperform the jump discontinuity estimators based on local polynomial (linear) estimators that are widely used in RD designs in Economics and other fields. Future work on this new class of estimators should devise a criterion for selecting an optimal element of the class.

References

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., Fan, J., Marron, J. S., 1997. On automatic boundary corrections. Annals of Statistics 25, 1691–1708.

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

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.

Härdle, W., Marron, J. S., Wand, M. P., 1990. Bandwidth choice for density derivatives. Journal of the Royal Statistical Society B 52, 223–232.

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. 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.

Mynbaev, K. T., Martins-Filho, C., 2019. Unified estimation of densities on bounded and unbounded domains. Annals of the Institute of Statistical Mathematics 71, 853–887.

Park, B., Marron, J. S., 1990. Comparison of data-driven bandwidth selectors. Journal of the American Statistical Association 85, 66–72.

Porter, J., 2003. Estimation in the regression discontinuity model. Tech. rep., Department of Economics - University of Wisconsin, Madison.

Schuster, E. F., 1985. Incorporating support constraints into nonparametric estimators of densities. Communications in Statistics - Theory and Methods, 14, 1123–1136.

Sheather, S., Jones, M. C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society B 53, 683–690.

Wand, M. P., Jones, M. C., 1995. Kernel smoothing. Chapman and Hall, Boca Raton, US.

[3]After extensive simulations we found that the sequence $w_j = 1/j$, originally proposed by Hestenes (1941), generally produces large variances and mean squared errors for $\tilde{J}_0$. However, a choice of $w_j = j$ or $w_j = j^2$ provides, for many data generating processes, root mean squared errors that are smaller than those associated with $\check{J}_0$.

A Appendix. Proof of theorem 3.1

Theorem 3.1 *Proof.* Since $\mu_u(0) = \mu(0)$ and $\lim_{x\uparrow 0} \mu_l(x) = \mu_l(0-) = \mu(0-)$, we write

$$\begin{aligned}\tilde{J}_0 - J_0 = & \frac{f(0)}{\tilde{f}_+(0)} \frac{1}{nh} \frac{1}{f(0)} \sum_{i=1}^{n} K_H\left(\frac{X_i}{h}\right) I_{\{X_i\geq 0\}}(Y_i - \mu_u(0)) \\ & - \frac{f(0)}{\tilde{f}_-(0)} \frac{1}{nh} \frac{1}{f(0)} \sum_{i=1}^{n} K_H\left(\frac{X_i}{h}\right) I_{\{X_i< 0\}}(Y_i - \mu_l(0-)).\end{aligned}$$

From Theorems 2 and 3 in Mynbaev and Martins-Filho (2019), $\frac{f(0)}{\tilde{f}_+(0)} = 1 + o_p(1)$ and $\frac{f(0)}{\tilde{f}_-(0)} = 1 + o_p(1)$. Then, letting $\varepsilon_i = Y_i - \mu(X_i)$

$$\begin{aligned}\tilde{J}_0 - J_0 = Z_n \Bigg\{ & \frac{1}{nh}\frac{1}{f(0)} \sum_{i=1}^{n} K_H\left(\frac{X_i}{h}\right) \left[(\mu(X_i) - \mu_u(0)) I_{\{X_i\geq 0\}} - (\mu(X_i) - \mu_l(0-)) I_{\{X_i<0\}}\right] \\ & + \frac{1}{nh}\frac{1}{f(0)} \sum_{i=1}^{n} K_H\left(\frac{X_i}{h}\right) \left(I_{\{X_i\geq 0\}} - I_{\{X_i<0\}}\right)\varepsilon_i \Bigg\} = Z_n \{T_{n,1} + T_{n,2}\} = Z_n T_n \qquad \text{(A.1)}\end{aligned}$$

where $Z_n = 1 + o_p(1)$. Since the sequence $\{(Y_i, X_i)\}_{i\in\mathbb{N}}$ is IID by assumption

$$\begin{aligned}E(T_{n,1}) &= \frac{1}{h}\frac{1}{f(0)} \left(\int_0^\infty K_H\left(\frac{x}{h}\right)(\mu_u(x) - \mu_u(0)) f(x) dx - \int_{-\infty}^0 K_H\left(\frac{x}{h}\right)(\mu_l(x) - \mu_l(0-)) f(x) dx \right) \\ &= \frac{1}{h}\frac{1}{f(0)} (\mathcal{I}_{n,1} - \mathcal{I}_{n,2}).\end{aligned}$$

Given that K is even, by changing variables in both integrals we obtain $\mathcal{I}_{n,1} = h\int_{\mathbb{R}} g_{s,u}(uh)K(u)du$ and $\mathcal{I}_{n,2} = h\int_{\mathbb{R}} g_{s,l}(uh)K(u)du$, where

$$g_{s,u}(x) = \begin{cases} (\mu_u(x) - \mu_u(0)) f(x) & \text{if } x \geq 0 \\ \sum_{j=1}^{s+1} k_j(\mu_u(-w_j x) - \mu_u(0)) f(-w_j x) & \text{if } x < 0 \end{cases}$$

and

$$g_{s,l}(x) = \begin{cases} (\mu_l(x) - \mu_l(0-)) f(x) & \text{if } x < 0 \\ \sum_{j=1}^{s+1} k_j(\mu_l(-w_j x) - \mu_l(0-)) f(-w_j x) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

Consequently, $E(T_{n,1}) = \frac{1}{f(0)}\left(\int_{\mathbb{R}}[g_{s,u}(uh) - g_{s,l}(uh)]K(u)du\right)$. Since $E(\varepsilon|X) = 0$, $E(T_{n,2}) = 0$ and $E(T_n) = E(T_{n,1})$. Given that $g_{s,u}(0) = g_{s,l}(0) = 0$, $\mu_u \in \mathcal{C}_b^s([0,\infty))$ and $\mu_l \in \mathcal{C}_b^s((-\infty,0))$, by Taylor's Theorem, for some $\lambda_u, \lambda_l \in (0,1)$,

$$g_{s,u}(uh) = g_{s,u}^{(1)}(0)uh + \frac{1}{2!} g_{s,u}^{(2)}(0)(uh)^2 + \cdots + \frac{1}{s!} g_{s,u}^{(s)}(\lambda_u hu)(uh)^s$$

and

$$g_{s,l}(uh) = g_{s,l}^{(1)}(0)uh + \frac{1}{2!} g_{s,l}^{(2)}(0)(uh)^2 + \cdots + \frac{1}{s!} g_{s,l}^{(s)}(\lambda_l hu)(uh)^s$$

where

$$g_{s,u}^{(\ell)}(x) = \begin{cases} \sum_{q=0}^{\ell} \begin{pmatrix} \ell \\ q \end{pmatrix} [\mu_u(x) - \mu_u(0)]^{(q)} f^{(\ell-q)}(x) & \text{if } x \geq 0 \\ \sum_{j=1}^{s+1} k_j(-w_j)^{\ell} \sum_{q=0}^{\ell} \begin{pmatrix} \ell \\ q \end{pmatrix} [\mu_u(-w_j x) - \mu_u(0)]^{(q)} f^{(\ell-q)}(-w_j x) & \text{if } x < 0 \end{cases}$$

$$g_{s,l}^{(\ell)}(x) = \begin{cases} \sum_{q=0}^{\ell} \begin{pmatrix} \ell \\ q \end{pmatrix} [\mu_l(x) - \mu_l(0-)]^{(q)} f^{(\ell-q)}(x) & \text{if } x < 0 \\ \sum_{j=1}^{s+1} k_j(-w_j)^{\ell} \sum_{q=0}^{\ell} \begin{pmatrix} \ell \\ q \end{pmatrix} [\mu_l(-w_j x) - \mu_l(0-)]^{(q)} f^{(\ell-q)}(-w_j x) & \text{if } x > 0 \\ \sum_{q=1}^{\ell} \begin{pmatrix} \ell \\ q \end{pmatrix} \mu^{(q)}(0-) f^{(\ell-q)}(0) & \text{if } x = 0 \end{cases}$$

for $\ell = 1, \cdots, s$. Given that $\int_{\mathbb{R}} u^m K(u) du = 0$ for $m = 1, \cdots, s-1$,

$$E(T_n) = \frac{h^s}{f(0)} \frac{1}{s!} \left(\int_{\mathbb{R}} [g_{s,u}^{(s)}(\lambda_u hu) - g_{s,l}^{(s)}(\lambda_l hu)] u^s K(u) du \right).$$

Since $g_{s,u},\ g_{s,l} \in \mathcal{C}_b^s(\mathbb{R})$, $\int_{\mathbb{R}} |u|^s |K(u)| du < C$, $|u|^{s+1}|K(u)| \to 0$ as $u \to \infty$, $\int_{\mathbb{R}} |f_X^{(\ell)}(x)| dx < C$ for $\ell = 1, \cdots, s$ and $h \to 0$ as $n \to \infty$,

$$\int_{\mathbb{R}} [g_{s,u}^{(s)}(\lambda_u hu) - g_{s,l}^{(s)}(\lambda_l hu)] u^s K(u) du \to \left(g_{s,u}^{(s)}(0) - g_{s,l}^{(s)}(0) \right) \int_{\mathbb{R}} u^s K(u) du \text{ as } n \to \infty,$$

where $g_{s,u}^{(s)}(0) = \sum_{q=1}^{s} \begin{pmatrix} s \\ q \end{pmatrix} \mu_u^{(q)}(0+) f^{(s-q)}(0)$ and $g_{s,l}^{(s)}(0) = \sum_{q=1}^{s} \begin{pmatrix} s \\ q \end{pmatrix} \mu_l^{(q)}(0-) f^{(s-q)}(0)$. Hence,

$$\begin{aligned} E(T_n) &= \frac{h^s}{s!} \frac{1}{f(0)} \left(\sum_{q=1}^{s} \begin{pmatrix} s \\ q \end{pmatrix} [\mu^{(q)}(0+) - \mu^{(q)}(0-)] f^{(s-q)}(0) \right) \int_{\mathbb{R}} u^s K(u) du + o(h^s) \\ &= \frac{h^s}{s!} B_s \mu_{K,s} + o(h^s). \end{aligned} \tag{A.2}$$

Now, $nhV(T_n) = nhV(T_{n,1}) + nhE(T_{n,2}^2) + 2nh\text{Cov}(T_{n,1}, T_{n,2})$. By the assumption that $\{(Y_i, X_i)\}_{i \in \mathbb{N}}$ is an IID sequence

$$\begin{aligned} V(T_{n,1}) &= \frac{1}{f(0)^2} \frac{1}{nh^2} V\left(K_H\left(\frac{X}{h}\right) \left[(\mu(X) - \mu_u(0)) I_{\{X \geq 0\}} - (\mu(X) - \mu_l(0-)) I_{\{X<0\}} \right] \right) \\ &\leq \frac{1}{f(0)^2} \frac{1}{nh^2} E\left(K_H^2\left(\frac{X}{h}\right) \left[(\mu(X) - \mu_u(0)) I_{\{X \geq 0\}} - (\mu(X) - \mu_l(0-)) I_{\{X<0\}} \right]^2 \right) \\ &= \frac{1}{f(0)^2} \frac{1}{nh} \int_0^{\infty} \left(K(u) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K\left(\frac{u}{w_j}\right) \right)^2 \Big[(\mu_u(uh) - \mu_u(0))^2 f(uh) \\ &+ (\mu_l(-uh) - \mu_l(0-))^2 f(-uh) \Big] du = \frac{1}{f(0)^2} \frac{1}{nh} \mathcal{T}_{n,1}, \end{aligned}$$

and $nhV(T_{n,1}) \leq \frac{1}{f(0)^2} \mathcal{T}_{n,1}$. Since $\mu_u \in \mathcal{C}_b^s([0,\infty)) \subset \mathcal{C}_b^1([0,\infty))$, by the mean value Theorem there exists $\lambda_u \in (0,1)$ such that $\mu_u(uh) - \mu_u(0) = \mu_u^{(1)}(uh\lambda_u)uh$. Similarly, since $\mu_l \in \mathcal{C}_b^1((-\infty, 0))$, there exists

$\lambda_l \in (0,1)$ such that $\mu_l(0-) - \mu_l(-uh) = \mu_l^{(1)}(-uh\lambda_l)uh$. Hence,

$$|\mathcal{T}_{n,1}| = h^2 \left| \int_0^\infty u^2 K_H^2(u)(\mu_u^{(1)}(uh\lambda_u))^2 f(uh)du + \int_0^\infty u^2 K_H^2(u)(\mu_l^{(1)}(-uh\lambda_l))^2 f(-uh)du \right|$$
$$\leq Ch^2 \left(\int_0^\infty u^2 K_H^2(u) f(uh)du + \int_0^\infty u^2 K_H^2(u) f(-uh)du \right) = Ch^2 \left(\mathcal{T}_{n,11} + \mathcal{T}_{n,12} \right).$$

Since f is continuous at 0, $\int_0^\infty u^2 |K(u)|du < C$ and $|u|^3|K(u)| \to 0$ as $u \to \infty$,

$$\int_0^\infty u^2 K_H^2(u) f(uh)du, \int_0^\infty u^2 K_H^2(u) f(-uh)du \to f(0) \int_0^\infty u^2 K_H^2(u)du.$$

Hence, $\mathcal{T}_{n,1} = O(h^2)$ and $nhV(T_{n,1}) = O(h^2)$.

Now, note that

$$T_{n,2}^2 = \frac{1}{f(0)^2} \frac{1}{n^2h^2} \left(\sum_{i=1}^n K_H^2\left(\frac{X_i}{h}\right) \varepsilon_i^2 (I_{\{X_i \geq 0\}} - I_{\{X_i < 0\}})^2 \right.$$
$$\left. + \sum\sum_{i \neq j} K_H\left(\frac{X_i}{h}\right) K_H\left(\frac{X_j}{h}\right) (I_{\{X_i \geq 0\}} - I_{\{X_i < 0\}})(I_{\{X_j \geq 0\}} - I_{\{X_j < 0\}}) \varepsilon_i \varepsilon_j \right) \text{ and}$$
$$E(T_{n,2}^2) = \frac{1}{f(0)^2} \frac{1}{n^2h^2} \left(E(T_{n,21}) + E(T_{n,22}) \right).$$

Since $\mathcal{V}(X) = E(\varepsilon^2|X)$,

$$E(T_{n,21}) = nE\left(K_H^2\left(\frac{X}{h}\right) \varepsilon^2 (I_{\{X \geq 0\}} - I_{\{X < 0\}})^2 \right) = nE\left(K_H^2\left(\frac{X}{h}\right) (I_{\{X \geq 0\}} - I_{\{X < 0\}})^2 E(\varepsilon^2|X) \right)$$
$$= nE\left(K_H^2\left(\frac{X}{h}\right) (I_{\{X \geq 0\}} - I_{\{X < 0\}})^2 \mathcal{V}(X) \right) = n \int_{\mathbb{R}} K_H^2\left(\frac{x}{h}\right) (I_{\{x \geq 0\}} - I_{\{x < 0\}})^2 \mathcal{V}(x) f(x)dx$$
$$= nh \int_0^\infty K_H^2(u)(\mathcal{V}(hu)f(hu) + \mathcal{V}(-hu)f(-hu))du.$$

By the independence assumption and the fact that $E(\varepsilon|X) = 0$, we have $E(T_{n,22}) = 0$. Hence,

$$E(T_{n,2}^2) = \frac{1}{f(0)^2} \frac{1}{nh} \int_0^\infty K_H^2(u)(\mathcal{V}(hu)f(hu) + \mathcal{V}(-hu)f(-hu))du = \frac{1}{f(0)^2} \frac{1}{nh} \mathcal{T}_{n,2}$$

and $nhE(T_{n,2}^2) = \frac{1}{f(0)^2}\mathcal{T}_{n,2}$. Now, since $0 \leq \mathcal{V}f$ is continuous at 0, $E(Y^2) < 0$, $\int_0^\infty |K(u)|du < C$ and $|u||K(u)| \to 0$ as $u \to \infty$,

$$\mathcal{T}_{n,2} \to 2\mathcal{V}(0)f(0) \int_0^\infty K_H^2(u)du \text{ as } n \to \infty, \tag{A.3}$$

and $nhE(T_{n,2}^2) = O(1)$. Then, by the Cauchy-Schwarz Inequality

$$nh|\text{Cov}(T_{n,1}, T_{n,2})| \leq \left(nhV(T_{n,1})nhE(T_{n,2}^2) \right)^{1/2} \leq \frac{1}{f(0)^2} O(h). \tag{A.4}$$

Hence, given $nhV(T_{n,1}) = O(h^2)$, $nhE(T_{n,2}^2) = O(1)$ and equations (A.2) and (A.4),

$$\left| nhV(T_n) - 2\frac{\mathcal{V}(0)}{f(0)} \int_0^\infty K_H^2(u)du \right| \leq |nhV(T_{n,1})| + \left| nhE(T_{n,2}^2) - 2\frac{\mathcal{V}(0)}{f(0)} \int_0^\infty K_H^2(u)du \right|$$
$$+ 2nh|\text{Cov}(T_{n,1}, T_{n,2})| = o(1).$$

Note that $\sqrt{nh}(T_n - E(T_n)) = \sqrt{nh}(T_{n,1} - E(T_{n,1})) + \sqrt{nh}T_{n,2}$, and since $V(\sqrt{nh}(T_{n,1} - E(T_{n,1}))) = nhV(T_{n,1}) = O(h^2)$ we have $\sqrt{nh}(T_{n,1} - E(T_{n,1})) = O_p(h)$. Hence,

$$\sqrt{nh}(T_n - E(T_n)) = \sqrt{nh}T_{n,2} + o_p(1). \tag{A.5}$$

Now, write $T_{n,2} = \frac{1}{f(0)}\sum_{i=1}^n Z_{n,i}$ where $Z_{n,i} = \frac{1}{nh}K_H\left(\frac{X_i}{h}\right)(I_{X_i\geq 0} - I_{X_i<0})\varepsilon_i$ and note that $E(Z_{n,i}) = 0$ and $V(Z_{n,i}) = \frac{1}{(nh)^2}\int_{\mathbb{R}} K_H^2(x/h)(I_{\{x\geq 0\}} - I_{\{x<0\}})^2\mathcal{V}(x)f(x)dx$. Letting $s_n^2 = \sum_{i=1}^n V(Z_{n,i})$ and defining $W_{n,i} = \frac{Z_{n,i}}{s_n}$, by Liapounov's central limit theorem

$$\sum_{i=1}^n W_{n,i} \stackrel{d}{\to} \mathcal{N}(0,1) \tag{A.6}$$

provided $\lim_{n\to\infty}\sum_{i=1}^n E|W_{n,i}|^{2+\delta} = 0$ for some $\delta > 0$. Now,

$$E|W_{n,i}|^{2+\delta} = (nhc_n)^{-1-\frac{\delta}{2}}E\left(\left|K_H\left(\frac{X}{h}\right)\right|^{2+\delta}|I_{\{X\geq 0\}} - I_{\{X<0\}}|^{2+\delta}|\varepsilon|^{2+\delta}\right)$$

where $c_n = \int_{\mathbb{R}} K_H^2(u)(I_{\{u\geq 0\}} - I_{\{u<0\}})^2\mathcal{V}(hu)f(hu)du$. Given that $E(|\varepsilon|^{2+\delta}|X) < C$,

$$\begin{aligned} E|W_{n,i}|^{2+\delta} &\leq (nhc_n)^{-1-\frac{\delta}{2}}C\int_{\mathbb{R}} |K_H(x/h)|^{2+\delta}\left|I_{\{x\geq 0\}} - I_{\{x<0\}}\right|^{2+\delta} f(x)dx \\ &= (nhc_n)^{-1-\frac{\delta}{2}}C\int_{\mathbb{R}} |K_H(x/h)|^{2+\delta} f(x)dx. \end{aligned}$$

Hence,

$$\sum_{i=1}^n E|W_{n,i}|^{2+\delta} \leq C(nh)^{-\frac{\delta}{2}}c_n^{-1-\frac{\delta}{2}}\int_{\mathbb{R}} |K_H(u)|^{2+\delta} f(hu)du.$$

Since, $nh \to \infty$ and $c_n \to 2\mathcal{V}(0)f(0)\int_0^\infty K_H^2(u)du$, we need only show that $\int_{\mathbb{R}} |K_H(u)|^{2+\delta} f(hu)du = O(1)$ as $n \to \infty$. By defining $w_0 = k_0 = -1$ we have $|K_H(u)|^{2+\delta} \leq \sum_{j=0}^{s+1} 2^{(j+1)(2+\delta)}\left|\frac{k_j}{w_j}\right|^{2+\delta}\left|K\left(\frac{u}{w_j}\right)\right|^{2+\delta}$ and

$$\int_{\mathbb{R}} |K_H(u)|^{2+\delta}f(hu)du \leq \sum_{j=0}^{s+1} 2^{(j+1)(2+\delta)}\left|\frac{k_j}{w_j}\right|^{2+\delta} w_j\int_{\mathbb{R}} |K(u)|^{2+\delta}f(huw_j)du.$$

Hence, if $\int_{\mathbb{R}} |K(u)|^{2+\delta}du < C$

$$\lim_{n\to\infty}\int_{\mathbb{R}} |K_H(u)|^{2+\delta}f(hu)du \leq \sum_{j=0}^{s+1} 2^{(j+1)(2+\delta)}\left|\frac{k_j}{w_j}\right|^{2+\delta} w_jf(0)\int_{\mathbb{R}} |K(u)|^{2+\delta}du < C.$$

Thus, equations (A.5) and(A.6) give

$$\sqrt{nh}(T_n - E(T_n)) \stackrel{d}{\to} \mathcal{N}\left(0, 2\frac{\mathcal{V}(0)}{f(0)}\int_0^\infty K_H^2(u)du\right). \tag{A.7}$$

From equation (A.1) $\sqrt{nh}\left(\tilde{J}_0 - J_0 - Z_nE(T_n)\right) = Z_n\sqrt{nh}\left(T_n - E(T_n)\right)$. Given that $Z_n = 1 + o_p(1)$, equations (A.2), (A.7) and $nh^{2s+1} = O(1)$,

$$\sqrt{nh}\left(\tilde{J}_0 - J_0 - \left(\frac{h^s}{s!}B_s + o_p(h^s)\right)\right) \stackrel{d}{\to} \mathcal{N}\left(0, 2\frac{\mathcal{V}(0)}{f(0)}\int_0^\infty K_H^2(u)du\right).$$

□