3 Special nonparametric models

In this chapter we study the estimation of different nonparametric models whose estimators build on the work we have developed in chapters 1 and 2.

3.1 Estimation of jumps in regression

Regression discontinuity (RD) designs have been widely used in economics, political science and other social and behavioral sciences to empirically evaluate the effects of policy or experimental interventions. See Imbens and Lemieux (2008), Lee and Lemieux (2010) and Cattaneo and Escanciano (2017) for theoretical and empirical overviews of the existing 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, *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 (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 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.

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 local 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 boundary points.

Here, we use the Hestenes' extension to define 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

$$J_0 := \mu(0) - \mu(0-). \tag{1}$$

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]

[1]The assumption that the discontinuity occurs at 0 is made without loss of generality. If it occurs at

Consider a function $g \in \mathcal{C}_b^s([0,\infty))$ 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. Of special interest are the decreasing sequence $w_i = 1/i$ used by Hestenes (1941), and the increasing sequence $w_i = i$. Let $\{k_i\}_{i=1}^{s+1}$ be defined by the system of equations

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

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

$$\phi_s(x) = \sum_{j=1}^{s+1} k_j g(-w_j x) \text{ for } x < 0. \tag{3}$$

Due to equation (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

$$g_s(x) = \begin{cases} g(x) & \text{for } x \geq 0 \\ \phi_s(x) & \text{for } x < 0 \end{cases}, \tag{4}$$

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 $[0,\infty)$ with $f(0) > 0$, i.e., $X \geq 0$. 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

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

$x_d \neq 0$, it suffices to shift μ by defining $\bar{\mu}(z) := \mu(z + x_d)$ where $z = x - x_d$ with discontinuity at $z = 0$.

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

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

$$\tilde{\mu}(0) = \frac{1}{f(0)} \frac{1}{nh} \sum_{i=1}^{n} K_H\left(\frac{X_i}{h}\right) Y_i, \tag{5}$$

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). For K even, we obtain

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

where

$$g_s(x) = \begin{cases} \mu(x) f(x) & \text{for } x \geq 0 \\ \sum_{j=1}^{s+1} k_j \mu(-w_j x) f(-w_j x) & \text{for } x < 0 \end{cases} \tag{6}$$

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

$$E(\tilde{\mu}(0)) - \mu(0) = \frac{1}{f(0)} \int_{\mathbb{R}} (g_s(uh) - g_s(0)) K(u) du. \tag{7}$$

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.

We propose the following Hestenes' based estimator for J_0:

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

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 $\{w_i\}_{i=1}^{s+1}$ 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.

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

Assumption 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 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 1. The conditional variance, as in Porter (2003) or Imbens and Kalyanaraman (2012), could also exhibit a discontinuity at 0 without fundamentally changing our theorems. 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 1. *Suppose* $0 < h \to 0$, $nh \to \infty$ *and* $nh^{2s+1} = O(1)$. *Then, under assumptions 1 and 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* $\mu_{K,s} = \int_{\mathbb{R}} u^s K(u)du$.

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 $\varepsilon_i = Y_i - \mu(X_i)$

$$\begin{aligned}\tilde{J}_0 - J_0 = Z_n\Bigg\{&\frac{1}{nh}\frac{1}{f(0)}\sum_{i=1}^{n} K_H\left(\frac{X_i}{h}\right)\left[(\mu(X_i) - \mu_u(0))I_{\{X_i\geq 0\}} - (\mu(X_i) - \mu_l(0-))I_{\{X_i<0\}}\right] \\ &+ \frac{1}{nh}\frac{1}{f(0)}\sum_{i=1}^{n} K_H\left(\frac{X_i}{h}\right)\left(I_{\{X_i\geq 0\}} - I_{\{X_i<0\}}\right)\varepsilon_i\Bigg\} = Z_n\left\{T_{n,1} + T_{n,2}\right\} = Z_n T_n \qquad (8)\end{aligned}$$

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

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

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,

$$\begin{aligned} E(T_n) &= \frac{h^s}{s!}\frac{1}{f(0)}\left(\sum_{q=1}^{s}\begin{pmatrix} s \\ q \end{pmatrix}[\mu^{(q)}(0+) - \mu^{(q)}(0-)]f^{(s-q)}(0)\right)\int_{\mathbb{R}} u^s K(u)du + o(h^s) \\ &= \frac{h^s}{s!}B_s\mu_{K,s} + o(h^s). \end{aligned} \tag{9}$$

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

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

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,

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

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

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

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

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

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 $u \to \infty$,

$$\mathcal{T}_{n,2} \to 2\mathcal{V}(0)f(0)\int_0^\infty K_H^2(u)du \text{ as } n \to \infty, \tag{10}$$

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

$$nh|\text{Cov}(T_{n,1}, T_{n,2})| \leq \left(nhV(T_{n,1})nhE(T_{n,2}^2)\right)^{1/2} \leq \frac{1}{f(0)^2}O(h). \tag{11}$$

Hence, given $nhV(T_{n,1}) = O(h^2)$, $nhE(T_{n,2}^2) = O(1)$ and equations (9) and (11),

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

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,

$$\sqrt{nh}(T_n - E(T_n)) = \sqrt{nh}T_{n,2} + o_p(1). \tag{12}$$

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

$$\sum_{i=1}^n W_{n,i} \xrightarrow{d} \mathcal{N}(0,1) \tag{13}$$

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

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

Thus, equations (12) and(13) give

$$\sqrt{nh}(T_n - E(T_n)) \xrightarrow{d} \mathcal{N}\left(0, 2\frac{\mathcal{V}(0)}{f(0)}\int_0^\infty K_H^2(u)du\right). \tag{14}$$

From equation (8) $\sqrt{nh}\left(\tilde{J}_0 - J_0 - Z_n E(T_n)\right) = Z_n\sqrt{nh}\left(T_n - E(T_n)\right)$. Given that $Z_n = 1 + o_p(1)$, equations (9), (14) and $nh^{2s+1} = O(1)$,

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

□

It is possible to compare the magnitude of the variance of the asymptotic distribution in Theorem 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,

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

shown by Cheng et al. (1997) to have 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

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

and by routine optimization we obtain

$$h^* = n^{-\frac{1}{2s+1}} \left(\frac{s!(s-1)!}{2}\right)^{\frac{1}{2s+1}} (B_s \mu_{K,s})^{-\frac{2}{2s+1}} \left(2\frac{\mathcal{V}(0)}{f(0)} I_K\left(s, \{w_j\}_{j=1}^{s+1}\right)\right)^{\frac{1}{2s+1}}. \tag{15}$$

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 (15) 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 $\tilde{J}_0$ depend on $\{w_i\}_{i=1}^{s+1}$. Conceptually, a suitably defined loss function can be defined to guide the choice of $\{w_i\}_{i=1}^{s+1}$ and, by consequence, $\tilde{J}_0$. We leave this optimal theoretical choice of $\{w_i\}_{i=1}^{s+1}$ 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.

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.