Abstract

This paper analyses inference procedures for local polynomial estimators commonly used in regression discontinuity designs. The standard practice in the literature (Porter 2003, Imbens and Lemieux 2008 and Lee and Lemieux 2009) is to rely in asymptotic approximations for the distribution of \( \sqrt{n}h(\hat{\alpha} - \alpha) \) that assume the bandwidth, \( h \), around the discontinuity shrinks towards zero as the sample size increases. However, in practice, the researcher has to use an \( h > 0 \) to implement the estimator. This paper provides more general asymptotic distributions for \( \sqrt{n}h(\hat{\alpha} - \alpha) \) that allow for the bandwidth to be positive providing band-
width robust standard errors that (partially) correct for bandwidth size. It shows that when $h > 0$ the traditional asymptotic approximation is equivalent to assuming that the density of the running variable and the conditional variance of the dependent variable are constant around the cutoff. Simulations are presented that provide evidence of improvement on the test’s size when using the bandwidth robust asymptotic distribution. Estimators for bandwidth robust standard errors are provided and are shown to incorporate the theoretical gains of the improved approximations. Finally, we provide simulation evidence that the use of OLS White standard errors, usually suggested in the literature, remains valid in this context and shows robustness to bandwidth choice. This is due to the fact that the usual White standard errors are consistent estimates of the bandwidth robust standard errors suggested in this paper.

1 Introduction

Regression Discontinuity (RD) designs have been propelled to the spotlight of economic analysis in recent years, especially in the policy and treatment evaluation literatures, as a form of estimating treatment effects in a non-experimental setting. The estimation of the treatment effect exploits the similarities of observations around a known threshold by comparing the outcomes of individuals just above and just below the cutoff. Its appeal comes from the relatively weak assumptions necessary to the identification of treatment effects and inference, which rely on its "quasi-experimental" characteristics.
Hahn, Todd and Van der Klaauw (1999, 2001) presented the conditions for identification of the treatment effect of interest and its estimation exploiting the discontinuities in treatment provision.

Porter (2003) provides more widely used results on the asymptotic properties of the estimators for the treatment effect of interest, obtaining limiting distributions for estimators based on local polynomial regression and partially linear estimation. He also proposes estimators consistent estimators for the asymptotic variance of the estimator that can be used to perform inference for a given dataset.

Imbens and Lemieux (2008) and Lee and Lemieux (2009) offer a broad review of the theoretical and applied literature with emphasis in the identification of the parameter of interest and its potential interpretations as weighted average treatment effects.

These papers evaluate the properties of the proposed estimators under the traditional assumption that the bandwidth used in the estimation procedure shrinks towards zero as the sample size grows. This guarantees identification of the parameter of interest under very mild conditions. The assumption is maintained when deriving the asymptotic distribution of the estimator that is the basis for inference.

In practice, the empiricist is confined to a certain dataset and cannot feasibly choose a bandwidth that is equal to zero or even shrinking towards zero, even if he promises to do so in the event that more data would be available.

This paper analyses the importance of the assumption that the bandwidth
converges to zero to the asymptotic distributions of the local polynomial estimator obtained by Porter (2003) and derives the distribution of the estimator when the bandwidth is allowed to be any positive real number. The results shown provide a different approximation to the estimator's bias and variance that will depend on the bandwidth size chosen by the researcher. These asymptotic distributions are robust to the choice of bandwidth.

It is shown that adopting the assumption that the bandwidth converges to zero is, when the bandwidth is different from zero, equivalent to assuming that the density of the running variable and conditional variance of errors is constant inside the bandwidth around the cutoff, which is most likely not true in empirical work.

We present evidence from Monte Carlo simulations which indicate that the use of these (theoretical) approximations improves the performance of inference about the treatment effect and the ability to correct for bias (potentially) present in the estimates. As one would expect, the benefits of these improved approximations is more relevant for larger bandwidths, for which the usual assumption of a bandwidth converging to zero would be less plausible.

This paper also proposes estimators for the variance of the estimators and provides evidence, through Monte Carlo simulations, that the feasible inference behaves as predicted by the theory, suggesting that the theoretical gains in robustness can be translated to practical benefits in applied work.

Finally, we compare the usual practice in performing inference about the treatment effect in the applied literature and conclude that the inference proce-
dures based on OLS, TSLS and the variance estimators proposed by Imbens and Lemieux (2008) and Lee and Lemieux (2009) are valid since they are consistent estimators for the asymptotic variance developed here, even when the bandwidth does not converge to zero. Hence, we can affirm that, even though the theoretical approximation for the asymptotic distribution of the RD estimates is not accurate for bandwidths away from zero (as expected), the standard errors used for inference by practitioners applying RD is valid and robust to bandwidth choice.

2 Model

We are interested in estimating the average treatment effect, $\alpha$, of a certain treatment or policy that affected part of a population of interest. As discussed in Porter (2003), Lee and Lemieux (2009) etc, RD designs is closely associated with the treatment effect literature. Angrist and Pischke (2009) provide a simple introduction to the intuition of regression discontinuity. There are two "types" of RD designs, sharp and fuzzy and they differ as to how treatment is assigned to a certain observation and the impact of the discontinuity in its assignment. I will focus on the sharp design at this section and emphasize the differences of the fuzzy design when needed.

Suppose that treatment status is a deterministic function of a so called "run-
ning" or "forcing" variable, $x$, such that,

$$d = \begin{cases} 
1 & \text{if } x \geq \bar{x} \\
1 & \text{if } x < \bar{x} 
\end{cases}$$

where $\bar{x}$ is the known cut-off point. Then, let $Y_1$ and $Y_0$ be the potential outcomes corresponding to the two possible treatment assignments. As usual, we cannot observe both potential outcomes, having access only to $Y = dY_1 - (1-d)Y_0$. As described by Hahn, Todd and Van der Klaauw (2001) and Porter (2003), under a smoothness assumption that $E(Y_j | x)$ is continuous at $\bar{x}$ for $j = 0, 1$, the average treatment effect can be estimated by comparing points just above and just below the discontinuity. The discontinuity in treatment assignment at $\bar{x}$ provides the opportunity of identifying $\alpha$ assuming only smoothness of the conditional expectation function around the discontinuity without any parametric functional form restrictions. From Porter (2003), p. 5,

$$\alpha = E(Y_1 - Y_0 | \bar{x})$$

$$= \lim_{x \uparrow \bar{x}} E(y | x) - \lim_{x \downarrow \bar{x}} E(y | x)$$

For an excellent review of RD designs, its applications and interpretation, see Lee and Lemieux (2009).\footnote{Very interestingly, Lee and Lemieux (2009) show that the so called RD gap obtained by the comparison of observations just above and just below the cutoff can be interpreted as a weighted average treatment effect across all individuals, not only the individuals around the cutoff. In this case each individual would have weights directly proportional to the ex ante likelihood that an individual's realization of $X$ will be close to the threshold.}
2.1 Assumptions

In the remainder of the paper, I will adopt the same basic assumptions used by Porter (2003) to derive the asymptotic distribution of the estimator for $\alpha$. The first assumption restricts the choice of kernel used in the estimation.

**Assumption 1 (Porter, 2003 Assumption 1)**  
(a) $k(\cdot)$ is a symmetric, bounded, Lipschitz function, zero outside a bounded set; $\int k(u)du = 1$.

(b) For a positive integer $s$, $\int k(u)u^jdu = 0$, $1 \leq j \leq s - 1$.

As discussed in Porter (2003) assumption 1 is fairly standard and allows for higher order kernels. If $s \geq 3$, the kernel has to be negative for some region of its domain to satisfy part (b) of the assumption. A bounded support set for the kernel avoids the use of a trimming function. Hence, only the local behavior around the discontinuity is the relevant.

Let $f_o$ denote the marginal density of $x$ and $m(x)$ denote the conditional expectation of $y$ given $x$ minus the discontinuity, i.e., $m(x) = E(y \mid x) - \alpha 1[x \geq \pi]$, where $\pi$ is where the discontinuity occurs.

**Assumption 2 (Porter, 2003, Assumption 2)** Suppose the data $(y_i, x_i)_{i=1,2,...,n}$ is i.i.d. and $\alpha$ is defined by

$$\alpha = \lim_{x \uparrow \pi} E(y \mid x) - \lim_{x \downarrow \pi} E(y \mid x)$$

(a) For some compact interval $\mathcal{R}$ of $x$ with $\pi \in \text{int}(\mathcal{R})$, $f_o$ is $l_f$ times continuously differentiable and bounded away from zero; $m(x)$ is $l_m$ times continuously differentiable for $x \in \mathcal{R}\setminus\{\pi\}$, and $m$ is continuous at $\pi$ with finite right and left-hand derivatives to order $l_m$. 


(b) Right and left-hand derivatives of m to order l_m are equal at \( x \).

Assumption 2(a) guarantees smoothness of the density of x and the conditional expectation of y on both sides of the discontinuity while allowing for different right and left-side derivatives of m at \( x \). Also, bounding the density of x on the neighborhood around \( x \) guarantees there is density and, hence "data" around the discontinuity to estimate the jump size.

Assumption 3 describes the behavior of the moments of the outcome variable around the discontinuity. Define, \( \varepsilon = y - E(y \mid x) = y - m(x) - \alpha 1[x \geq x] \).

**Assumption 3 (Porter, 2003, Assumption 3)**

(a) \( \sigma^2(x) = E(\varepsilon^2 \mid x) \) is continuous for \( x \neq \bar{x}, x \in \mathbb{R} \), and right and left-hand limits at \( \bar{x} \) exist.

(b) For some \( \zeta > 0 \), \( E(|\varepsilon|^{2+\zeta} \mid x) \) is uniformly bounded on \( \mathbb{R} \).

Assumption 3(a) allows the conditional variance of the outcome variable to be a function of the running variable and assures it is well behaved around the cutoff. Part (b) bounds the moments so that a central limit theorem can be applied (Porter 2003).

### 2.2 Approximate Asymptotic Distributions

In this section we present some results by Porter (2003) that provide the asymptotic approximations to perform inference regarding the parameter of interest \( \alpha \) and argue that an approximation that is more robust to the choice bandwidth is easily available. Also, note that the traditional assumption that the bandwidth converges to zero produces the same results for applications that would
be obtained by assuming that the density of the running variable \( x \), and the conditional variance of the dependent variable, \( y \), are constant in the bandwidth around the cutoff.

First, the simple case of the Nadaraya-Watson estimator is presented for ease of explanation, followed by the more general case for the local polynomial estimator.

### 2.2.1 Nadaraya-Watson Estimator

The Nadaraya-Watson estimator is the simple kernel estimator of the discontinuity jump which takes a kernel weighted average of observations at each side of the discontinuity and its difference. In the sharp design case, given data \((y_i, x_i)_{i=1,2,...,n}\), let

\[
\hat{\alpha} = \frac{n^{-1} \sum_i h^{-1} k \left( \frac{x - x_i}{h} \right) y_i d_i}{n^{-1} \sum_j h^{-1} k \left( \frac{x - x_j}{h} \right) d_j} - \frac{n^{-1} \sum_i h^{-1} k \left( \frac{x - x_i}{h} \right) y_i (1 - d_i)}{n^{-1} \sum_j h^{-1} k \left( \frac{x - x_j}{h} \right) (1 - d_j)}
\]

where \( d_i = 1[x \geq \bar{x}] \), \( k(\cdot) \) is a kernel function, and \( h \) denotes a bandwidth that controls size of the local neighborhood to be averaged over.

Porter (2003) shows that a asymptotic normality approximation result is available for this estimator if the bandwidth \( h \) is considered to shrink towards zero at certain rates. I reproduce the theorem below for convenient comparison with the results presented later in this paper.

**Theorem 1 (Porter, 2003, Theorem 1)** Suppose Assumptions 1 (a), 2 (a) and 3 hold with \( l_m \) any positive integer and \( l_f \) any nonnegative integer. If \( h \to 0 \),
\( nh \to \infty \) and \( h\sqrt{n} \to C \), where \( 0 \leq C < \infty \), then

\[
\sqrt{nh}(\hat{\alpha} - \alpha) \xrightarrow{d} N \left( 2CK_1(0)(m'^-(\bar{x}) + m'^-(\bar{x})), 4\delta_o \frac{\sigma^2 + \sigma^2}{f_o(\bar{x})} \right)
\]

where \( \delta_o = \int \frac{1}{\omega} k^2(u)du \), \( K_1(\omega) = \int \frac{1}{\omega} k(u)udu \) and \( m'^+(-)(x) \) is the right (left)-hand derivative of \( m(x) \) at point \( x \).

However, for a researcher with access to a given sample of size \( n \) it is known that the size of the chosen bandwidth \( h \) is set as a number and cannot shrink towards zero. Hence, it would be of interest to analyze if a more refined asymptotic approximation for the estimator’s behavior is available that could take in consideration the impact of the choice of a certain value \( h \) for the bandwidth.

It is reasonable to expect that this approximation will not be precise and that it could be improved by approximations of higher order.

In Porter (2003) it is assumed that \( h \to 0 \), while \( nh \to \infty \). This provides the conditions for the theorem reproduced above. However, from the point of view of the empirical researcher who would like to test hypotheses about the treatment effect \( \alpha \), using Porter’s result is equivalent to ignoring the bias and approximating the estimator’s asymptotic variance with a first order approximation.

**Corollary 1** Suppose Assumptions 1 (a), 2 (a) and 3 hold with \( l_m \) any positive integer and \( l_f \) any nonnegative integer. If \( h \to b \), \( nh \to \infty \) then

\[
\sqrt{nh}(\hat{\alpha} - \alpha) \xrightarrow{d} N \left( B_1, \int_0^M k^2(\bar{x}) \frac{\sigma^2 + \sigma^2}{\left( \int_0^M k(u) f_o(\bar{x} + ub)du \right)^2} + \frac{\sigma^2}{\left( \int_0^M k(u) f_o(\bar{x} - ub)du \right)^2} du \right)
\]
where

\[ B_1 = C_1 \int_0^M k(u) u \times \left\{ \frac{m(\pi + ub) - m(\pi)}{ub} \right\} f_o(\pi - ub) \left\{ \frac{m(\pi - ub) - m(\pi)}{ub} \right\} f_o(\pi) du \]

\[ = \frac{\sqrt{nb}}{\int_0^M k(u) f_o(\pi + ub) du} \left( \int_0^M k(u) m(\pi + ub) f_o(\pi + ub) du \right) - \frac{\sqrt{nb}}{\int_0^M k(u) f_o(\pi - ub) du} \left( \int_0^M k(u) m(\pi - ub) f_o(\pi - ub) du \right) \]

and \( C_1 = \lim_{n \to \infty} (nb)^{\frac{1}{2}} b \)

This result is clearly not as "clean" as Porter’s since the asymptotic variance for \( \sqrt{nh}(\hat{\alpha} - \alpha) \) is more complicated and the fact that the bias increases with the number of observations available. Nevertheless, note that Porter (2003) gets around this issue by assuming \( h \to 0 \) and \( (nh)^{\frac{1}{2}} h \to C \). This would imply one unique rate for the convergence of \( h \) and I argue that both approximations have the same issue for the bias, which can vanish in Porter’s approximation only due to the (unfeasible) promise of reducing bandwidths as \( n \) increases. Section 3 presents simulations that indicate that by using the variance formula in Corollary 1 one can obtain test statistics for \( \hat{\alpha} \) that have superior size performance relatively to the ones that would be obtained by using Porter’s approximation. As expected, the performance of the approximations is very similar for small bandwidths for which Porter’s assumptions are more reasonable and diverge in favor of the approximation in Corollary 1 for higher values of \( h \).

To make matters clear, I present a very simple extension of Porter (2003) that includes only a linear term to the approximations for \( f_o(x), m(x) \) and
\( \sigma^2(x) \), used for the asymptotic variance and bias of the estimator \( \hat{\alpha} \) instead of using the true values of these functions as done in Corollary 1.

**Corollary 2** Suppose Assumptions 1 (a), 2 (a) and 3 hold with \( l_m \) any positive integer and \( l_f \) any nonnegative integer, also assume \( \sigma^2(x) \) \( l_\sigma \) times continuously differentiable for \( x \in \mathbb{N} \setminus \{\tau\} \). If \( h \to b, nh \to \infty \) then

\[
\sqrt{nh}(\hat{\alpha} - \alpha) \xrightarrow{d} N(B_{ap}, V_{ap})
\]

where

\[
B_{ap} = C_{ap} \left\{ K_1^1(0)f_o(\tau) \left[ \frac{m^{+}(\tau)}{D_1} + \frac{m^{-}(\tau)}{D_2} \right] + bK_2^1(0) \left[ \frac{m^{+}(\tau)f_o^{+}(\tau)}{D_1} - \frac{m^{-}(\tau)f_o^{-}(\tau)}{D_2} \right] \right\}
\]

\[
V_{ap} = f_o(\tau)K_0^2(0) \left[ \frac{\sigma^{2+}(\tau)}{(D_1)^2} + \frac{\sigma^{2-}(\tau)}{(D_2)^2} \right] +
+ bK_1^2(0) \left[ \frac{\sigma^{2+}(\tau)f_o^{+}(\tau) + \sigma^{2+}(\tau)f_o(\tau)}{(D_1)^2} - \frac{\sigma^{2-}(\tau)f_o^{-}(\tau) + \sigma^{2-}(\tau)f_o(\tau)}{(D_2)^2} \right] +
+ b^2K_2^2(0) \left[ \frac{\sigma^{2+}(\tau)f_o^{+}(\tau)}{(D_1)^2} + \frac{\sigma^{2-}(\tau)f_o^{-}(\tau)}{(D_2)^2} \right]
\]

where \( K_j^i(\omega) = \int_\omega^M k^i(u) w^j du \), \( D_1 = \frac{1}{2}f_o(\tau) + hf_o^{+}(\tau)K_1^1(0) \), \( D_2 = \frac{1}{2}f_o(\tau) - hf_o^{-}(\tau)K_1^1(0) \) and \( C_{ap} = \lim_{n \to \infty} (nb)^{\frac{1}{2}} b \)

Note that this simplifies to the distribution developed by Porter (2003) if the right and left-hand derivatives of \( f_o(x) \) and \( \sigma^2(x) \) evaluated at \( \tau \) are equal to zero. This is the basic idea of the following claim.

**Claim 1** Under the assumptions in corollary 1, if \( h \to b > 0 \), using the asymptotic distribution approximation for the variance of \( \sqrt{nh}(\hat{\alpha} - \alpha) \) in theorem 1 is equivalent to assuming that all the relevant derivatives of \( f_o(x) \) and \( \sigma^2(x) \) are equal to zero in the bandwidth around the cutoff.

12
Even though the assumptions are theoretically very distinct, it is clear that for the practitioner, who has to choose a bandwidth larger than zero to implement the estimator, using Porter’s (2003) results to perform inference is equivalent to assuming that $f_\alpha(x)$ and $\sigma^2(x)$ are constant inside the bandwidth.

In section 3 we present evidence that the use of more precise approximations for both the bias and standard errors improves the size of the (infeasible) empirical tests that test the null that the treatment effect is equal to its true value against the two-sided alternative when the bandwidth $h$ increases.

### 2.2.2 Local Polynomial Estimators

Local polynomial estimators are an attractive option when performing estimation in the regression discontinuity setting given its nice boundary behavior as described by Fan and Gijbels (1996). Porter (2003), assuming that the bandwidth shrinks towards zero, derived the asymptotic properties of the RD estimator based on fitting local polynomials on both sides of the discontinuity and showed that better bias behavior relatively to the Nadaraya-Watson estimator is achieved. The bias reduction is obtained by polynomial "correction" (Porter, 2003).

In applied work, the use of local polynomials is a staple for estimation of treatment effects in RD settings. This is partially due to its easy implementation, nice properties and by the fact that the local linear estimator has been the focus on several papers that helped to disseminate the technique (Hahn, Todd and Van der Klaauw 1999 and 2001, Imbens and Lemieux 2008 and Lee and
Lemieux 2009).

The order $p$ local polynomial is defined as follows. Let $(\hat{\alpha}_{p^+}, \hat{\beta}_{p^+})$ be the solution to the minimization problem.

$$
\min_{a,b_1,\ldots,b_p} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \left( \frac{x_i - \bar{x}}{h} \right) d_i \left[ y_i - a - b_1(x_i - \bar{x}) - \ldots - b_p(x_i - \bar{x})^p \right]^2
$$

and similarly $(\hat{\alpha}_{p^-}, \hat{\beta}_{p^-})$ minimizes the same objective function but with $1 - d_i$ in place of $d_i$. Then, the estimator of the parameter of interest $\alpha$ is given by

$$
\hat{\alpha}_p = \hat{\alpha}_{p^+} - \hat{\alpha}_{p^-}
$$

if $p = 0$ we would be back to the Nadaraya-Watson estimator.

The result presented by Porter (2003) regarding the local polynomial estimator is replicated below for the benefit of the reader. For the proof refer to Porter (2003).

**Theorem 2 (Porter, 2003, Theorem 3)** Suppose Assumptions 1 (a) and 3 hold.

(a) If Assumption 2 (a) holds with $l_m \geq p+1$ and $l_f$ any nonnegative integer. If $h \to 0$, $nh \to \infty$ and $h^{p+1}\sqrt{nh} \to C_a$, where $0 \leq C_a < \infty$, then

$$
\sqrt{nh}(\hat{\alpha}_p - \alpha) \overset{d}{\to} N \left( B_a, \frac{\sigma^2(\bar{x}) + \sigma^2(\bar{x})}{f_o(\bar{x})} e_1' \Gamma^{-1} \Delta \Gamma^{-1} e_1 \right)
$$

(b) If $p$ is odd, Assumption 2 (a) holds with $l_m \geq p+2$ and $l_f$ any nonnegative integer. If $h \to 0$, $nh^3 \to \infty$ and $h^{p+2}\sqrt{nh} \to C_b$, where $0 \leq C_b < \infty$, then

$$
\sqrt{nh}(\hat{\alpha}_p - \alpha) \overset{d}{\to} N \left( B_b, \frac{\sigma^2(\bar{x}) + \sigma^2(\bar{x})}{f_o(\bar{x})} e_1' \Gamma^{-1} \Delta \Gamma^{-1} e_1 \right)
$$
where

\[ B_a = \frac{C_a}{(p+1)!} \left[ m^{(p+1)}(x) - (-1)^{p+1} m^{(p+1)}(x) \right] e'_1 \Gamma^{-1} \begin{bmatrix} \gamma_{p+1} \\ \vdots \\ \gamma_{2p+1} \end{bmatrix} \]

\[ B_b = 2C_b \left( \frac{m^{(p+1)}(x) f'_o(x) + m^{(p+2)}(x)}{(p+1)! \ f_o(x)} + \frac{m^{(p+2)}(x)}{(p+2)!} \right) e'_1 \Gamma^{-1} \begin{bmatrix} \gamma_{p+2} \\ \vdots \\ \gamma_{2p+2} \end{bmatrix} \]

\[ -2C_b \left( \frac{m^{(p+1)}(x) f'_o(x)}{(p+1)! \ f_o(x)} \right) e'_1 \Gamma^{-1} \Gamma_{(+1)} \Gamma^{-1} \begin{bmatrix} \gamma_{p+1} \\ \vdots \\ \gamma_{2p+1} \end{bmatrix} \]

and

\[ \Gamma = \begin{bmatrix} \gamma_0 & \cdots & \gamma_p \\ \vdots & \ddots & \vdots \\ \gamma_p & \cdots & \gamma_{2p} \end{bmatrix}, \quad \Gamma_{(+1)} = \begin{bmatrix} \gamma_1 & \cdots & \gamma_{p+1} \\ \vdots & \ddots & \vdots \\ \gamma_{p+1} & \cdots & \gamma_{2p+1} \end{bmatrix}, \quad \text{and} \ \Delta = \begin{bmatrix} \delta_0 & \cdots & \delta_p \\ \vdots & \ddots & \vdots \\ \delta_p & \cdots & \delta_{2p} \end{bmatrix}, \]

\[ e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \gamma_j = \int_0^\infty k(u)u^j du, \quad \delta_j = \int_0^\infty k^2(u)u^j du \text{ and } m^{(l)+(-)}(x) \]

is the lth right (left)-hand derivative of m(x) at point x.

Similarly to what was done in section 2.2.1, we analyze the asymptotic behavior of the local polynomial estimator under the more general condition that the bandwidth converges to a nonnegative value and suggest that the result be used towards inference about the parameter of interest. By using this more general approximation, we will explicitly take in account the choice of bandwidth in calculating our standard errors, hence, making inference more robust to its choice relative to what would be obtained by using Porter’s theorem 3.
Corollary 3 Suppose Assumptions 1 (a) and 3 hold.

If Assumption 2 (a) holds with $l_m \geq p + 1$ and $l_f$ any nonnegative integer.

If $h \to b$, $nh \to \infty$, then

$$\sqrt{nh}(\hat{\alpha}_p - \alpha) \overset{d}{\to} N \left( B_{ar}, e_1' \left( (\Gamma^+_*)^{-1} \Delta^+_* (\Gamma^+_*)^{-1} + (\Gamma^-_*)^{-1} \Delta^-_* (\Gamma^-_*)^{-1} \right) e_1 \right)$$

where

$$B_{ar} = C_p e_1' \left\{ (\Gamma^+_*)^{-1} \left[ \int_0^M k(u) Z(x + uh) m(x + uh) \, dx \right] - (\Gamma^-_*)^{-1} \left[ \int_0^M k(u) Z(x - uh) m(x - uh) \right] \right\}$$

and

$$\Gamma^+_* = \begin{bmatrix} \gamma^+_0 & \ldots & \gamma^+_p \\ \vdots & \ddots & \vdots \\ \gamma^+_0 & \ldots & \gamma^+_2p \end{bmatrix}, \text{ and } \Delta^+_* = \begin{bmatrix} \delta^+_0 & \ldots & \delta^+_p \\ \vdots & \ddots & \vdots \\ \delta^+_0 & \ldots & \delta^+_2p \end{bmatrix},$$

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \gamma^+_j = \int_0^M k(u) u^j f_o(x + uh) \, dx,$$

$$\gamma^-_j = (-1)^j \int_0^M k(u) u^j f_o(x - uh) \, dx,$$

$$\delta^+_j = \int_0^M k^2(u) u^j \sigma^2(x + uh) f_o(x + uh) \, dx,$$

$$\delta^-_j = (-1)^j \int_0^M k^2(u) u^j \sigma^2(x - uh) f_o(x - uh) \, dx$$

and $C_p = \lim_{n \to \infty} \sqrt{nh}$.

Corollary 3 provides the asymptotic distribution for the RD estimator of the parameter of interest without assuming that the bandwidth converges to zero as sample size increases. Clearly the formulas for asymptotic variance become a lot more cumbersome, nevertheless are still functions of known data and could be calculated if the functions $f_o(x)$ and $\sigma^2(x)$ were known for the values of the support inside the bandwidth centered on the discontinuity point. In fact, that is exactly the approach used in section 3 to obtain the simulated coverages of
the hypotheses tests for \( \alpha \) which show the improvement in the variance and bias
approximation using the results in corollary 3 relatively to the results in Porter’s
theorem 2. As in the Nadaraya-Watson estimator’s case it is straightforward to
see that when \( h \to b > 0 \) using the results in theorem 2 is equivalent to assume
\( f_o(x) \) and \( \sigma^2(x) \) are constant around the cutoff.

**Claim 2** Under the assumptions in corollary 3, if \( h \to b > 0 \), using the asymp-
totic distribution approximation for the variance of \( \sqrt{n}h(\hat{\alpha}_p - \alpha) \) in theorem 2
is equivalent to assuming that all the relevant derivatives of \( f_o(x) \) and \( \sigma^2(x) \) are
equal to zero in the bandwidth around the cutoff.

### 2.3 Variance Estimators

To be able to perform inference about \( \alpha \) using the information in a given sample
we need to obtain appropriate estimates for the unknown terms in the asymp-
totic variance’s formulas developed in Theorem 1 and Corollaries 1 and 2.

Turning to Corollary 1 and noting that the components of the asymptotic
variance of \( \sqrt{n}h(\hat{\alpha} - \alpha) \) can be written as

\[
\begin{align*}
\int_0^M k^2(u) \sigma^2+(\bar{x} + ub)f_o(\bar{x} + ub)du &= E \left[ h^{-1}k^2 \left( \frac{\bar{x} - x}{h} \right) d\varepsilon^2 \right] \\
\int_0^M k^2(u) \sigma^2-(\bar{x} - ub)f_o(\bar{x} - ub)du &= E \left[ h^{-1}k^2 \left( \frac{\bar{x} - x}{h} \right) (1 - d)\varepsilon^2 \right] \\
\int_0^M k(u) f_o(\bar{x} + ub)du &= E \left[ h^{-1}k \left( \frac{\bar{x} - x}{h} \right) d \right] \\
\int_0^M k(u) f_o(\bar{x} - ub)du &= E \left[ h^{-1}k \left( \frac{\bar{x} - x}{h} \right) (1 - d) \right]
\end{align*}
\]

according to equations 5 and 2.
Natural estimators for these terms are given by their sample analogues, under Assumption 2 and a given $h$. Those are, respectively,

\[
(nh)^{-1} \sum_{i=1}^{n} k^2 \left( \frac{\bar{x} - x_i}{h} \right) d_i \hat{\varepsilon}_i^2
\]

\[
(nh)^{-1} \sum_{i=1}^{n} k^2 \left( \frac{\bar{x} - x_i}{h} \right) (1 - d_i) \hat{\varepsilon}_i^2
\]

\[
(nh)^{-1} \sum_{i=1}^{n} k \left( \frac{\bar{x} - x_i}{h} \right) d_i
\]

\[
(nh)^{-1} \sum_{i=1}^{n} k \left( \frac{\bar{x} - x_i}{h} \right) (1 - d_i)
\]

where $\hat{\varepsilon}_i = y_i - d_i \bar{y}^+ - (1 - d_i) \bar{y}^-$, with $\bar{y}^+ = n_1^{-1} \sum_{\bar{x} \leq x_i \leq \bar{x} + M_h} y_i$ and $\bar{y}^- = n_u^{-1} \sum_{\bar{x} - M_h \leq x_i \leq \bar{x}} y_i$ and $n_l$ and $n_u$ are the number of observations used in each sum. Then the estimator for the variance of $\sqrt{nh}(\hat{\alpha} - \alpha)$ is given by

\[
\frac{(nh)^{-1} \sum_{i=1}^{n} k^2 \left( \frac{\bar{x} - x_i}{h} \right) d_i \hat{\varepsilon}_i^2}{\left( (nh)^{-1} \sum_{i=1}^{n} k \left( \frac{\bar{x} - x_i}{h} \right) d_i \right)^2} + \frac{(nh)^{-1} \sum_{i=1}^{n} k^2 \left( \frac{\bar{x} - x_i}{h} \right) (1 - d_i) \hat{\varepsilon}_i^2}{\left( (nh)^{-1} \sum_{i=1}^{n} k \left( \frac{\bar{x} - x_i}{h} \right) (1 - d_i) \right)^2}
\]

\[
= \frac{(nh) \sum_{i=1}^{n} k^2 \left( \frac{\bar{x} - x_i}{h} \right) d_i \hat{\varepsilon}_i^2}{\left( \sum_{i=1}^{n} k \left( \frac{\bar{x} - x_i}{h} \right) d_i \right)^2} + \frac{(nh) \sum_{i=1}^{n} k^2 \left( \frac{\bar{x} - x_i}{h} \right) (1 - d_i) \hat{\varepsilon}_i^2}{\left( \sum_{i=1}^{n} k \left( \frac{\bar{x} - x_i}{h} \right) (1 - d_i) \right)^2}
\]

which is, by an application of the law of large numbers and Slutzky theorem, trivially consistent for the asymptotic variance of $\sqrt{nh}(\hat{\alpha} - \alpha)$ presented in corollaries 1 and 2 for any bandwidth chosen.
If we are using a rectangular kernel, this simplifies to

\[
\frac{(nh)^{-1} \sum_{\tau-Mh \leq x_i \leq \tau + Mh} (\frac{1}{2})^2 d_i \hat{\varepsilon}_i^2}{(nh)^{-1} \sum_{\tau-Mh \leq x_i \leq \tau + Mh} (\frac{1}{2})^2 (1-d_i) \hat{\varepsilon}_i^2} + \frac{(nh)^{-1} \sum_{\tau-Mh \leq x_i \leq \tau + Mh} (\frac{1}{2})^2}{(nh)^{-1} \sum_{\tau-Mh \leq x_i \leq \tau + Mh} (\frac{1}{2})^2 (1-d_i)}
\]

\[
= \frac{(nh)^{-1} (\frac{1}{4}) \left( \sum_{\tau-Mh \leq x_i \leq \tau + Mh} d_i \hat{\varepsilon}_i^2 \right)^2}{(\sum_{\tau-Mh \leq x_i \leq \tau + Mh} d_i)^2} + \frac{(nh) \sum_{\tau-Mh \leq x_i \leq \tau + Mh} (1-d_i) \hat{\varepsilon}_i^2}{(\sum_{\tau-Mh \leq x_i \leq \tau + Mh} (1-d_i))^2}
\]

\[
= \frac{(nh) \sum_{\tau-Mh \leq x_i \leq \tau + Mh} d_i \hat{\varepsilon}_i^2}{n_u^2} + \frac{(nh) \sum_{\tau-Mh \leq x_i \leq \tau} \hat{\varepsilon}_i^2}{n_l^2}
\]

\[
= \frac{nh \hat{\varepsilon}_u^2 + nh \hat{\varepsilon}_l^2}{n_u \hat{\varepsilon}_u^2 + n_l \hat{\varepsilon}_l^2} = \frac{nh}{n_u + n_l} \left( \frac{n_u + n_l \hat{\varepsilon}_u^2}{n_u \hat{\varepsilon}_u^2 + n_l \hat{\varepsilon}_l^2} \right)
\]

\[
= \frac{nh}{n_u + n_l} \left( \frac{\hat{\varepsilon}_u^2 + \hat{\varepsilon}_l^2 + n_l \hat{\varepsilon}_u^2 + n_u \hat{\varepsilon}_l^2}{n_u \hat{\varepsilon}_u^2 + n_l \hat{\varepsilon}_l^2} \right), \text{ if } n_u = n_l
\]

\[
= \frac{2nh}{n_u + n_l} \left( \hat{\varepsilon}_u^2 + \hat{\varepsilon}_l^2 \right)
\]

which is the estimator proposed for the asymptotic variance by Imbens and Lemieux (2008).

We can also see that the White Robust standard errors for OLS would be valid estimators for the variance presented in corollary 1.

It is clear that, even though the asymptotic theory derived in most of the RD literature relies (mainly for identification purposes) on the assumption that \( h \to 0 \) as \( n \to \infty \), the practice suggested for applied work does not depend on that assumption, with the estimators for the standard errors being robust to choice of bandwidth. The main concern, then, should fall on the bias introduced to the estimate of \( \alpha \) by the use of larger bandwidths.
Nevertheless, the issue of bias becomes less important when local polynomials are used to estimate $\alpha$ as seen in section 2.2.2.

3 Simulations

The asymptotic approximations derived in section 2 differ about their assumptions regarding the limit of the bandwidth used, $h$, with Porter (2003) assuming $h \to 0$, while this paper assumes that $h \to b > 0$. The assumption adopted here is more general in the sense that it encompasses Porter’s. As shown by claims 1 and 2, assuming that $h \to 0$ is equivalent to assuming that the probability density function of $X$ and the conditional variance of the errors are constant in the bandwidth around the cutoff. In fact, one would reasonably expect that the approximations should be closely equivalent for bandwidth values close to zero, in which case, the assumptions would be very similar. This section presents evidence from simulations that inference about the treatment effects of interest using the approximation proposed in this paper have better size behavior than Porter’s for larger empirical bandwidths as expected.

Note that the simulations used for comparing the two sets of results are unfeasible since they depend on knowledge about $f_o(x), \sigma^2(x)$ and $m(x)$ around the cutoff. Later we compare the empirical coverages that would be obtained by using estimated standard errors.
3.1 Nadaraya-Watson Estimator

In the simulations presented here we use the Nadaraya-Watson estimator described in section 2.2.1 to estimate \( \alpha \) and perform a test where the null hypotheses is true using the asymptotic standard errors suggested by theorem 1 and corollaries 1 and 2. Note that the standard errors used in these tests are obtained as if the researcher knew the true values for \( f_\alpha(x) \) and \( \sigma^2(x) \) in the range around the cutoff. This exercise aims to evaluate the relative performance of the asymptotic approximations suggested by Porter (2003) \((h \to 0)\) and this paper \((h \to b > 0)\) for different choices of bandwidth.

Even though in all simulations we are using the N-W estimator and the correspondent asymptotic approximations, I present the results under different specifications of the true model so we can have a grasp of the potential differences in the approximate asymptotic biases implied by the theory in each case. To be fair, the comparisons between the bias approximations by Porter’s formula and our formula is not adequate for any true model in which the relationship of \( y \) and \( x \) could be described by a polynomial of order higher than linear (or order higher than \( p + 1 \) in the local polynomial case) since it describes the bias as being a function of the derivative of \( m(x) \) limiting its accuracy to more complex functional forms. Nevertheless, the simulations suggest that even against the simple linear case the bias derived by Porter underestimates the bias greatly relatively to the asymptotic bias presented in corollary 1.

In the simulation, the running variable, \( x \), is an i.i.d. random variable with normal distribution with mean 50 and standard deviation 10. The error term,
is an i.i.d. random variable with normal distribution with mean zero and standard deviation equal to one. The cutoff point is arbitrarily set to 55. In the specifications used, the treatment effect, $\alpha$, is set equal to 10, the intercept of the model is equal to 3 and the linear, quadratic and cubic slopes equal to 0.5, -0.005 and 0.00002, respectively. The sample size used is equal to 1,000 and the we use 2,000 repetitions of the procedure. For each repetition, the estimate of the treatment effect and the theoretical standard errors are obtained for 100 bandwidths ranging from 0.2 through 20 units of $x$ around the cutoff point. The empirical coverages presented are the fraction of rejections in the 2,000 repetitions for a test of size 5% (two-sided).

The graphs in the attachment show the empirical coverage for the test for the (true) null hypotheses that $\alpha = 10$ when the N-W estimator is used to obtain $\hat{\alpha}$ and the (unfeasible) variances for $\sqrt{n h} (\hat{\alpha} - \alpha)$ presented in theorem 1 and corollaries 1 and 2 are used.

The coverages in figure 1 were obtained through the simulations above when the data generating process used for the dependent variable does not depend on $X$ directly, i.e., $y_i = \mu + \alpha \text{Treat}_i + u_i$. In this case, we would not expect bias on the estimates for $\alpha$. As expected, the empirical size for the tests using different standard errors approximations behave very similarly for small bandwidths when Porter’s (2003) assumptions are closer to be true, but the differences increase as the bandwidth size increases, suggesting that the results presented in corollary 1 provide a better approximation for the asymptotic variance of the estimator $\hat{\alpha}$. Note that the linear approximation from corollary 2, even though
improve the performance of the inference slightly does not, in this case, capture the majority of the impact of bandwidth on the standard errors.

Figures 2, 3 and 4 refer to a data generating process in which $Y$ is linearly related to $X$, i.e., $y_i = \mu + \beta_1 x_i + \alpha T reat_i + u_i$. In this case, a large bias on the N-W estimate should be present for any bandwidth away from zero, and this is clearly reflected on figure 2, where the steep decline on the empirical coverage makes clear the deleterious effects of the bias on the estimate and inference. This effect is big enough that overwhelms the gains obtained by the better approximation of the variance of the estimates.

Nevertheless, we can use this example and the knowledge about the DGP to obtain what would be both the approximated variance and bias for this estimator suggested by Porter (2003) and this paper. Hence, figure 3 shows the empirical coverage for the bias corrected test. It is clear that after correcting for bias, inference using the calculations in corollary 1 have better size than using Porter (2003) or corollary 2. To make explicit where the gains are coming from, figure 4 presents the empirical coverage for both Porter’ s (2003) and using this paper’s results and what would be obtained if only we had used only the bandwidth robust standard errors while correcting bias using Porter’s bias formula or using the robust bias formula with Porter’s standard errors. The main part of the improvement is due to better measurement of the bias, but the gains from the more precise calculations for the standard error are non trivial and in line with the results obtained for the simpler DGP in figure 1.

The remainder graphs present equivalent simulations for different DGPs.
Figures 5, 6 and 7 have a quadratic DGP; figures 8, 9 and 10 present a cubic relationship between \( Y \) and \( X \); and finally, in figures 11, 12 and 13 \( y_i = \exp\left(\frac{x_i}{20}\right) + \alpha \text{Treat}_i + u_i \). The "speed" with which the bias becomes a problem varies depending on the DGP, but in general it becomes a nuisance for relatively small bandwidths. In the quadratic case the bias correction suggested by Porter’s approximation or the linear approximation in corollary 2 is actually detrimental to the performance of the test, this is due to the DGP used, in which \( y \) peaks at 50 causing the bias to be more controlled for smaller bandwidths than in the other cases. None of the approximations does a superb job when the DGP follows an exponential pattern.

In any case, for all the simulations we have evidence that the asymptotic approximation for the distribution of \( \sqrt{n}h(\hat{\alpha} - \alpha) \) proposed in corollary 1 that takes (partially) in consideration the effects of the choice of bandwidth provides better approximations to both bias and variance than those derived by Porter (2003) under the assumption that \( h \to 0 \).

These results are unfeasible since they were derived by plugging the true density and variances for \( x \) in the formulas provided in section 2.2.1. To make these tests feasible we need an estimator for the standard errors\(^2\). Figure 14 presents the results for a simulation identical to the one presented in figure 1, when the dependent variable does not depend on \( X \) directly, and provides the empirical coverage for the feasible tests in which the standard errors are

---

\(^2\)In fact we would like an adequate estimator for the bias. Unfortunately, it is very difficult to disentangle the bias from the estimate \( \hat{\alpha} \).
estimated with use of the estimator in equation 1. As can be seen, the empirical coverage of the tests using the estimated standard errors capture, for larger bandwidths, the gains in performance of using std. errors robust to bandwidth choice, outperforming the tests that used Porter’s (2003) approximation for the std. errors taking advantage of the knowledge about the true distribution of $x$. For smaller bandwidths the tests that use estimated std. errors perform slightly worse than both that use the theoretical std. errors. that can be probably attributed to smaller sample sizes available to estimate variances.

3.2 Local Polynomial Estimator

The simulations for the local polynomial estimator are not currently available, and will be included shortly.

4 Work in Progress

This paper is still in early stages of development and in later versions is expected to include similar results for the Fuzzy Regression Discontinuity Design.

5 Conclusion

The use of Regression Discontinuity designs to obtain estimators of treatment effects, $\alpha$, has been widely used in recent years by researchers in economics. The papers that tackled the main theoretical issues in this area, e.g., Hahn, Todd and Van der Klaauw (1999, 2001), Porter (2003), Imbens and Lemieux
(2008) etc assume that the bandwidth around the discontinuity shrinks fast enough towards zero to guarantee identification of the parameter of interest. In practice, however, researchers are usually bounded in their ability to reduce the bandwidth size by data availability constraints.

This paper analyses the asymptotic behavior of these estimators allowing the bandwidth to converge to a positive number. Under this more general conditions, we present asymptotic distributions for local polynomial estimators that are more robust to bandwidth choice by the researcher relatively to the ones presented by Porter (2003). These include more precise formulas for both bias and variances of the estimators of interest. Claims 1 and 2 show that when \( h > 0 \) the traditional asymptotic approximation for the variance of the estimators is equivalent to assume that the density of the running variable and the conditional variance of the dependent variable are constant around the cutoff.

Simulations are presented that provide evidence that the bandwidth robust approximations suggested are more accurate for both bias and variances than the usually used in the literature. This is reflected on improvements on test’s size, specially when the estimator uses larger bandwidths.

Estimators for bandwidth robust standard errors are provided and shown to incorporate the theoretical gains of the improved approximations. More interestingly, the variance estimators (including OLS and TSLS) suggested by Porter (2003), Imbens and Lemieux (2008), Lee and Lemieux (2009) among others are consistent for the bandwidth robust variances developed in sections 2.2.1 and 2.2.2. Hence, we can conclude that, even though the theory developed
previously in the literature does not provide adequate approximations for the asymptotic behavior of the estimators it studies for \( h > 0 \), the empirical methods used by practitioners produces valid tests that can be used for inference in this case.

6 Appendix

Proof of Corollary 1. The Nadaraya-Watson estimator is given by

\[
\hat{\alpha} = \frac{\sum_i h^{-1} k \left( \frac{x_i - \bar{x}}{h} \right) y_i d_i}{\sum_j h^{-1} k \left( \frac{x_j - \bar{x}}{h} \right) d_j} - \frac{\sum_i h^{-1} k \left( \frac{y_i - \bar{y}}{h} \right) y_i (1 - d_i)}{\sum_j h^{-1} k \left( \frac{y_j - \bar{y}}{h} \right) (1 - d_j)}
\]

plugging \( y_i = m(x_i) + \alpha d_i + \varepsilon_i \), and let \( A_1 = n^{-1} \sum_j h^{-1} k \left( \frac{x_i - \bar{x}}{h} \right) d_j \) and \( A_2 = n^{-1} \sum_j h^{-1} k \left( \frac{y_i - \bar{y}}{h} \right) (1 - d_j) \)

\[
\hat{\alpha} = \frac{n^{-1} \sum_i h^{-1} k \left( \frac{x_i - \bar{x}}{h} \right) [m(x_i) + \alpha d_i + \varepsilon_i] d_i}{A_1} - \frac{n^{-1} \sum_i h^{-1} k \left( \frac{y_i - \bar{y}}{h} \right) [m(x_i) + \alpha d_i + \varepsilon_i] (1 - d_i)}{A_2}
\]

\[
= \frac{n^{-1} \sum_i h^{-1} k \left( \frac{x_i - \bar{x}}{h} \right) [m(x_i) - m(\bar{x}) + \varepsilon_i] d_i}{A_1} - \frac{n^{-1} \sum_i h^{-1} k \left( \frac{y_i - \bar{y}}{h} \right) [m(x_i) - m(\bar{x}) + \varepsilon_i] (1 - d_i)}{A_2}
\]

then,

\[
\sqrt{nh} (\hat{\alpha} - \alpha) = \frac{n^{-\frac{1}{2}} \sum_i h^{-\frac{1}{2}} k \left( \frac{x_i - \bar{x}}{h} \right) [m(x_i) - m(\bar{x}) + \varepsilon_i] d_i}{A_1} - \frac{n^{-\frac{1}{2}} \sum_i h^{-\frac{1}{2}} k \left( \frac{y_i - \bar{y}}{h} \right) [m(x_i) - m(\bar{x}) + \varepsilon_i] (1 - d_i)}{A_2}
\]

First, consider the denominator \( A_1 \).

\[
A_1 = n^{-1} \sum_j h^{-1} k \left( \frac{x_j - \bar{x}}{h} \right) d_j
\]
Note that, since the observations are i.i.d.

\[
\text{Var}(A_1) = \text{Var} \left( n^{-1} \sum_j h^{-1} k \left( \frac{x_j - \bar{x}}{h} \right) d_j \right) \\
= \frac{1}{(nh)^2} \sum_j \text{Var} \left( k \left( \frac{x_j - \bar{x}}{h} \right) d_j \right) \\
\leq \frac{1}{(nh)^2} \sum_j E \left[ \left( k \left( \frac{x_j - \bar{x}}{h} \right) d_j \right)^2 \right] \\
= \frac{1}{nh} E \left[ \frac{1}{h^2} \left( \frac{x_j - \bar{x}}{h} \right) d_j \right] \\
= \frac{1}{nh} \int_{-\infty}^{\infty} \frac{1}{h^2} \left( \frac{x - \bar{x}}{h} \right) f_o(x) dx \\
= \frac{1}{nh} \int_{-\infty}^{\infty} \frac{1}{h^2} \left( \frac{x - \bar{x}}{h} \right) f_o(x) dx
\]

let \( u = \frac{x - \bar{x}}{h} \) and hence, \( x = \bar{x} + uh, \) w.l.g. and \( dx = hdu \)

\[
\text{Var}(A_1) = \frac{1}{nh} \int_0^M k^2(u) f_o(\bar{x} + uh) du \\
= \frac{1}{nh} \int_0^M k^2(u) f_o(\bar{x} + uh) du \\
= O \left( \frac{1}{nh} \right) = o(1)
\]

Also, by the L.L.N.

\[
A_1 = n^{-1} \sum_j h^{-1} k \left( \frac{x_j - \bar{x}}{h} \right) d_j \\
= E \left[ h^{-1} k \left( \frac{x - \bar{x}}{h} \right) \right] + o_p(1) \\
= \int_{-\infty}^{\infty} \frac{1}{h} k \left( \frac{x - \bar{x}}{h} \right) f_o(x) dx + o_p(1) \\
= \int_0^M k(u) f_o(\bar{x} + uh) du + o_p(1)
\]

Then, similarly

\[
A_2 = \int_0^M k(u) f_o(\bar{x} - uh) du + o_p(1)
\]
For the numerator:

\[
\begin{align*}
& n^{-\frac{1}{2}} \sum_i h^{-\frac{1}{2}} k \left( \frac{x_i \mp x}{h} \right) [m(x_i) - m(x) + \varepsilon_i] d_i \\
& = n^{-\frac{1}{2}} \sum_i h^{-\frac{1}{2}} k \left( \frac{x_i \mp x}{h} \right) [m(x_i) - m(x)] d_i + n^{-\frac{1}{2}} \sum_i h^{-\frac{1}{2}} k \left( \frac{x_i \mp x}{h} \right) \varepsilon_i d_i
\end{align*}
\]

Note that the second term follows a CLT. For Liapunov’s CLT

\[
\sum_i E \left[ \left( (nh)^{-\frac{1}{2}} k \left( \frac{x_i \mp x}{h} \right) \varepsilon_i d_i \right)^{2+\zeta} \right]
\]

\[
\begin{align*}
& = (nh)^{-\frac{1}{2}} \sum_i E \left[ k \left( \frac{x_i \mp x}{h} \right)^{2+\zeta} \varepsilon_i^{2+\zeta} d_i \right] \\
& \leq (nh)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \frac{1}{h} \left| k \left( \frac{x \mp x}{h} \right) \right|^{2+\zeta} E \left[ \left| \varepsilon \right|^{2+\zeta} \right] f_o(x)dx \\
& = \sup_{x \in \mathbb{R}} E \left[ \left| \varepsilon \right|^{2+\zeta} \right] (nh)^{-\frac{1}{2}} \int_{0}^{M} \frac{1}{h} \left| k \left( \frac{x \mp x}{h} \right) \right|^{2+\zeta} f_o(x)dx \\
& = O(1)o(1)O(1) = o(1)
\end{align*}
\]
For the variance

\[
\sum_i Var \left[ (nh)^{-\frac{1}{2}} k \left( \frac{x_i - \bar{x}}{h} \right) \varepsilon_i d_i \right] = \sum_i E \left[ (nh)^{-1} k^2 \left( \frac{x_i - \bar{x}}{h} \right) \varepsilon_i^2 d_i \right] = E \left[ h^{-1} k^2 \left( \frac{x_i - \bar{x}}{h} \right) \varepsilon_i^2 d_i \right] = E \left[ (h)^{-1} k^2 \left( \frac{x_i - \bar{x}}{h} \right) E (\varepsilon_i^2 | x) d_i \right] = E \left[ h^{-1} k^2 \left( \frac{x_i - \bar{x}}{h} \right) E (\varepsilon_i^2 | x) d_i \right] = \int_{\pi}^{\infty} h^{-1} k^2 \left( \frac{x - \bar{x}}{h} \right) \sigma^2(x) f_o(x) dx = \int_0^M k^2 (u) \sigma^2(\bar{x} + uh) f_o(\bar{x} + uh) du \tag{5}
\]

For the bias term,

\[
n^{-\frac{1}{2}} \sum_i h^{-\frac{1}{2}} k \left( \frac{x_i - \bar{x}}{h} \right) [m(x_i) - m(\bar{x})] d_i
\]

the variance.

\[
Var \left( n^{-\frac{1}{2}} \sum_i h^{-\frac{1}{2}} k \left( \frac{x_i - \bar{x}}{h} \right) [m(x_i) - m(\bar{x})] d_i \right) = (nh)^{-1} \sum_i Var \left[ k \left( \frac{x_i - \bar{x}}{h} \right) [m(x_i) - m(\bar{x})] d_i \right] \leq h^{-1} E \left[ k^2 \left( \frac{x - \bar{x}}{h} \right) [m(x) - m(\bar{x})]^2 d \right] \leq h^{-1} \left\{ \sup_{v \in [0, M]} \left[ |m(\bar{x} + vh) - m(\bar{x})| \right] \right\}^2 E \left[ k^2 \left( \frac{x - \bar{x}}{h} \right) d \right] \leq \left\{ \sup_{v \in [0, M]} \left[ |m(\bar{x} + vh) - m(\bar{x})| \right] \right\}^2 \int_0^M k^2 (u) f_o(\bar{x} + uh) du = \left\{ \sup_{v \in [0, M]} \left[ |m(\bar{x} + vh) - m(\bar{x})| \right] \right\}^2 O(1)
\]
And has expectation given by,

\[
E \left( n^{-\frac{1}{2}} \sum_i h^{-\frac{1}{2}} k \left( \frac{x_i - x}{h} \right) [m(x_i) - m(x)] d_i \right)
\]

\[
= (nh)^{-\frac{1}{2}} \sum_i E \left[ k \left( \frac{x_i - x}{h} \right) [m(x_i) - m(x)] d_i \right]
\]

\[
= \left( \frac{n}{h} \right)^{\frac{1}{2}} E \left[ k \left( \frac{x - x_i}{h} \right) [m(x) - m(x_i)] d_i \right]
\]

\[
= \left( \frac{n}{h} \right)^{\frac{1}{2}} \int_{-\infty}^{\infty} k \left( \frac{x - x_i}{h} \right) [m(x) - m(x_i)] d f_o(x) dx
\]  

(7)

\[
= (nh)^{\frac{1}{2}} \int_0^M k(u) [m(x + uh) - m(x)] f_o(x + uh) du
\]

\[\blacksquare\]

**Proof of Corollary 2.** For the denominator, taking equations 3 and 4, and by using a Taylor series, approximate \( f_o(x + uh) = f_o(x) + f'_o(x)uh + r(u) \)

\[
A_1 = \int_0^M k(u) [f_o(x) + f'_o(x)uh + r(u)] du + o_p(1)
\]

\[
= f_o(x) \int_0^M k(u) du + h f'_o(x) \int_0^M k(u) u du + \int_0^M k(u) r(u, h) du + o_p(1)
\]

\[
= \frac{1}{2} f_o(x) + h f'_o(x) \int_0^M k(u) u du + \int_0^M k(u) r(u, h) du + o_p(1)
\]

then,

\[
A_1 \overset{p}{\to} \frac{1}{2} f_o(x) + h f'_o(x) \int_0^M k(u) u du + \int_0^M k(u) r^+(u, h) du
\]

and similarly

\[
A_2 \overset{p}{\to} \frac{1}{2} f_o(x) - h f'_o(x) \int_0^M k(u) u du + \int_0^M k(u) r^-(u, h) du
\]

If we ignore the higher order terms \( r^+(u, h) \) and \( r^-(u, h) \).
\[ A_1 = \frac{1}{2} f_o(\overline{x}) + h f'_o(\overline{x}) \int_0^M k(u) \, du = D_1 \]
\[ A_2 = \frac{1}{2} f_o(\overline{x}) - h f'_o(\overline{x}) \int_0^M k(u) \, du = D_2 \]

For the numerator, the CLT taking equation 6, if we add differentiability of \( \sigma^2(x) \) w.r.t. \( x \in \mathbb{R} \), we can approximate \( \sigma^2(\overline{x} + uh) \) in the same fashion that \( f_o(\overline{x} + uh) \). Then,

\[
\int_0^M k^2(u) \sigma^2(\overline{x} + uh)f_o(\overline{x} + uh)du \\
= \int_0^M k^2(u) \sigma^2(\overline{x} + uh)[f_o(\overline{x}) + f'_o(\overline{x})uh + r(u)] \, du \\
= \int_0^M k^2(u) [\sigma^{2+}(\overline{x}) + \sigma^{2+}(\overline{x})uh + r_\sigma(u, h)] [f_o(\overline{x}) + f'_o(\overline{x})uh + r^+(u, h)] \, du \\
= \sigma^{2+}(\overline{x}) \left[ f_o(\overline{x}) \int_0^M k^2(u) \, du + hf'_o(\overline{x}) \int_0^M k^2(u) \, du + \int_0^M k^2(u) r^+(u, h) \, du \right] + \\
+ h \sigma^{2+}(\overline{x}) \left[ f_o(\overline{x}) \int_0^M k^2(u) \, du + hf'_o(\overline{x}) \int_0^M k^2(u) \, du + \int_0^M k^2(u) r^+(u, h) \, du \right] + \\
+ \left[ f_o(\overline{x}) \int_0^M k^2(u) r_\sigma(u, h) \, du + hf'_o(\overline{x}) \int_0^M k^2(u) r_\sigma(u, h) \, du + \int_0^M k^2(u) r_\sigma(u, h) r^+(u, h) \, du \right]
\]

Then, if we ignore the remainder terms for both approximations,

\[
\approx \sigma^{2+}(\overline{x}) \left[ f_o(\overline{x}) \int_0^M k^2(u) \, du + hf'_o(\overline{x}) \int_0^M k^2(u) \, du \right] + \\
+ h \sigma^{2+}(\overline{x}) \left[ f_o(\overline{x}) \int_0^M k^2(u) \, du + hf'_o(\overline{x}) \int_0^M k^2(u) \, du \right]
\]
Then the asymptotic variance for \( \sqrt{n} \hat{h}(\hat{\alpha} - \alpha) \) is given by

\[
\begin{align*}
&\frac{\sigma^2(\varpi)}{\left( \frac{1}{2} f_0(\varpi) + hf^{+}_{0}(\varpi) \int_0^M k^2 (u) \, du \right)^2} + \\
&h\sigma^2(\varpi) \left[ f_0(\varpi) \int_0^M k^2 (u) \, du + hf^{+}_{0}(\varpi) \int_0^M k^2 (u) \, du \right] \\
&\left( \frac{1}{2} f_0(\varpi) + hf^{+}_{0}(\varpi) \int_0^M k (u) \, du \right)^2 + \\
&\frac{\sigma^2(\varpi)}{\left( \frac{1}{2} f_0(\varpi) - hf^{-}_{0}(\varpi) \int_0^M k (u) \, du \right)^2} - \\
&h\sigma^2(\varpi) \left[ f_0(\varpi) \int_0^M k^2 (u) \, du - hf^{-}_{0}(\varpi) \int_0^M k^2 (u) \, du \right] \\
&\left( \frac{1}{2} f_0(\varpi) - hf^{-}_{0}(\varpi) \int_0^M k (u) \, du \right)^2 + \\
&\int_0^M k^2 (u) \left\{ \sigma^2(\varpi) \left[ f_0(\varpi) + hu f^{+}_{0}(\varpi) \right] + h\sigma^2(\varpi) \left[ f_0(\varpi) u + hf^{+}_{0}(\varpi) u^2 \right] \right\} du \\
&\left( \frac{1}{2} f_0(\varpi) + hf^{+}_{0}(\varpi) \int_0^M k (u) \, du \right)^2 + \\
&\int_0^M k^2 (u) \left\{ \sigma^2(\varpi) \left[ f_0(\varpi) - hu f^{-}_{0}(\varpi) u \right] - h\sigma^2(\varpi) \left[ f_0(\varpi) u - hf^{-}_{0}(\varpi) u^2 \right] \right\} du \\
&\left( \frac{1}{2} f_0(\varpi) - hf^{-}_{0}(\varpi) \int_0^M k (u) \, du \right)^2 + \\
&\int_0^M k^2 (u) \left[ f_0(\varpi) + uh f^{+}_{0}(\varpi) \right] \left\{ \sigma^2(\varpi) + uh\sigma^2(\varpi) \right\} du \\
&\left( \frac{1}{2} f_0(\varpi) + hf^{+}_{0}(\varpi) \int_0^M k (u) \, du \right)^2 + \\
&\int_0^M k^2 (u) \left[ f_0(\varpi) - uh f^{-}_{0}(\varpi) \right] \left\{ \sigma^2(\varpi) - uh\sigma^2(\varpi) \right\} du \\
&\left( \frac{1}{2} f_0(\varpi) - hf^{-}_{0}(\varpi) \int_0^M k (u) \, du \right)^2
\end{align*}
\]
Or in a more convenient form,

\[
\frac{1}{(\frac{1}{2} f_o(\overline{x}) + hf_o^+(\overline{x})K_1^1(0))^2} \left[ \sigma^{2+}(\overline{x}) f_o(\overline{x}) K_0^2(0) + h\sigma^{2+}(\overline{x}) f_o^+(\overline{x}) K_1^2(0) \right] +
\frac{1}{(\frac{1}{2} f_o(\overline{x}) + hf_o^+(\overline{x})K_1^1(0))^2} \left[ h\sigma^{2+}(\overline{x}) f_o(\overline{x}) K_1^2(0) + h^2 \sigma^{2+}(\overline{x}) f_o^+(\overline{x}) K_2^2(0) \right] +
\frac{1}{(\frac{1}{2} f_o(\overline{x}) - hf_o^-(\overline{x})K_1^1(0))^2} \left[ \sigma^{2-}(\overline{x}) f_o(\overline{x}) K_0^2(0) - h\sigma^{2-}(\overline{x}) f_o^-(\overline{x}) K_1^2(0) \right] +
\frac{1}{(\frac{1}{2} f_o(\overline{x}) - hf_o^-(\overline{x})K_1^1(0))^2} \left[ -h\sigma^{2-}(\overline{x}) f_o(\overline{x}) K_1^2(0) + h^2 \sigma^{2-}(\overline{x}) f_o^-(\overline{x}) K_2^2(0) \right]
\]

\[
= f_o(\overline{x}) K_0^2(0) \left[ \frac{\sigma^{2+}(\overline{x})}{(\frac{1}{2} f_o(\overline{x}) + hf_o^+(\overline{x})K_1^1(0))^2} + \frac{\sigma^{2-}(\overline{x})}{(\frac{1}{2} f_o(\overline{x}) - hf_o^-(\overline{x})K_1^1(0))^2} \right] +
+hK_1^2(0) \left[ \frac{\sigma^{2+}(\overline{x}) f_o^+(\overline{x})}{(\frac{1}{2} f_o(\overline{x}) + hf_o^+(\overline{x})K_1^1(0))^2} + \frac{\sigma^{2-}(\overline{x}) f_o^-(\overline{x})}{(\frac{1}{2} f_o(\overline{x}) - hf_o^-(\overline{x})K_1^1(0))^2} \right] +
+h^2 K_2^2(0) \left[ \frac{\sigma^{2+}(\overline{x}) f_o^+(\overline{x})}{(\frac{1}{2} f_o(\overline{x}) + hf_o^+(\overline{x})K_1^1(0))^2} + \frac{\sigma^{2-}(\overline{x}) f_o^-(\overline{x})}{(\frac{1}{2} f_o(\overline{x}) - hf_o^-(\overline{x})K_1^1(0))^2} \right]
\]

For the bias term, using the same approximations, taking equation 7

\[
(nh)^\frac{1}{2} \int_0^M k(u) \left[ m(\overline{x} + uh) - m(\overline{x}) \right] f_o(\overline{x} + uh) du
\]

\[
= (nh)^\frac{1}{2} \int_0^M k(u) \left[ m(\overline{x}) + uhm^+(\overline{x}) + r_m(u, h) - m(\overline{x}) \right] f_o(\overline{x} + uh) du
\]

\[
= (nh)^\frac{1}{2} \int_0^M k(u) \left[ uhm^+(\overline{x}) + r_m(u, h) \right] f_o(\overline{x} + uh) du
\]

\[
= (nh)^\frac{1}{2} hm^+(\overline{x}) \int_0^M k(u) f_o(\overline{x} + uh) du + (nh)^\frac{1}{2} \int_0^M k(u) r_m(u, h) f_o(\overline{x} + uh) du
\]

\[
= (nh)^\frac{1}{2} f_o(\overline{x}) \left[ m^+(\overline{x}) h \int_0^M k(u) u du + \int_0^M k(u) r_m(u, h) du \right]
\]

\[
+ (nh)^\frac{1}{2} f_o^+(\overline{x}) \left[ m^+(\overline{x}) h^2 \int_0^M k(u) u^2 du + h \int_0^M k(u) ur_m(u, h) du \right]
\]

\[
+ (nh)^\frac{1}{2} f_o^-(\overline{x}) \left[ m^+(\overline{x}) h \int_0^M r(u, h) k(u) u du + \int_0^M k(u) r_m(u, h) du \right]
\]
And by ignoring the remainder terms,

\[ 
\approx \left( nh \right)^{\frac{1}{2}} f_o(x) m^{+}(x) h \int_{0}^{M} k(u) \, du \\
+ \left( nh \right)^{\frac{1}{2}} f_o'(x) m^{+}(x) h^2 \int_{0}^{M} k(u) \, du \\
= \left( nh \right)^{\frac{1}{2}} h m^{+}(x) \left[ f_o(x) \int_{0}^{M} k(u) \, du + h f_o'(x) \int_{0}^{M} k(u) \, du \right] 
\]

The bias will be given approximately by

\[ 
B_{ap} = \left( nh \right)^{\frac{1}{2}} h m^{+}(x) \left[ f_o(x) \int_{0}^{M} k(u) \, du + h f_o'(x) \int_{0}^{M} k(u) \, du \right] \\
+ \left( nh \right)^{\frac{1}{2}} h m^{-}(x) \left[ f_o(x) \int_{0}^{M} k(u) \, du - h f_o'(x) \int_{0}^{M} k(u) \, du \right] \\
= \left( nh \right)^{\frac{1}{2}} h \left\{ \left[ f_o(x) \int_{0}^{M} k(u) \, du \right] \left[ \frac{m^{+}(x)}{2 f_o(x) + h f_o'(x) f_0^{+}(x) K_1^{+}(0)} + \frac{m^{-}(x)}{2 f_o(x) - h f_o'(x) f_0^{-}(x) K_1^{-}(0)} \right] \\
+ h \left[ f_o(x) \int_{0}^{M} k(u) \, du \right] \left[ \frac{m^{+}(x) f_o'(x)}{2 f_o(x) + h f_o'(x) f_0^{+}(x) K_1^{+}(0)} - \frac{m^{-}(x) f_o'(x)}{2 f_o(x) - h f_o'(x) f_0^{-}(x) K_1^{-}(0)} \right] \right\} \\
= \left( nh \right)^{\frac{1}{2}} h \left\{ K_1^{+}(0) f_o(x) \left[ \frac{m^{+}(x)}{2 f_o(x) + h f_o'(x) f_0^{+}(x) K_1^{+}(0)} + \frac{m^{-}(x)}{2 f_o(x) - h f_o'(x) f_0^{-}(x) K_1^{-}(0)} \right] \\
+ h K_2^{+}(0) \left[ \frac{m^{+}(x) f_o'(x)}{2 f_o(x) + h f_o'(x) f_0^{+}(x) K_1^{+}(0)} - \frac{m^{-}(x) f_o'(x)}{2 f_o(x) - h f_o'(x) f_0^{-}(x) K_1^{-}(0)} \right] \right\} 
\]

where \( K_j^{+} = \int_{0}^{M} k^j(u) \, du \).

**Proof of Corollary 3.** The local polynomial estimator is given by

\[ 
\hat{\alpha}_p = \hat{\alpha}_{p+} - \hat{\alpha}_{p-} 
\]

note that,

\[ 
\hat{\alpha}_{p+} = e' \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i Z_i' \right]^{-1} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i y_i \right] \\
= e' D_{n+} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i y_i \right] 
\]

where \( D_{n+} = \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i Z_i' \right]^{-1} \). Similarly, for \( D_{n-} = \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) (1 - d_i) Z_i Z_i' \right]^{-1} \)

\[ 
\hat{\alpha}_{p-} = e' D_{n-} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) (1 - d_i) Z_i y_i \right] 
\]
Then,

\[
\hat{\alpha}_{p+} = e'_1 D_{n+} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i [m(x_i) + \alpha d_i + \varepsilon_i] \right] \\
= e'_1 D_{n+} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i m(x_i) + \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i \alpha + \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i \varepsilon_i \right] \\
= e'_1 D_{n+} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i m(x_i) \right] + \alpha e'_1 D_{n+} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i \right] + \\
+ e'_1 D_{n+} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i \varepsilon_i \right]
\]

note that \( Z_i = Z_i Z'_i \varepsilon_1, e'_1 e_1 = 1 \) then

\[
e'_1 D_{n+} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i \right] = e'_1 D_{n+} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i Z'_i \right] e_1 = 1
\]

and

\[
\hat{\alpha}_{p+} - \alpha = e'_1 D_{n+} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i m(x_i) \right] + e'_1 D_{n+} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i \varepsilon_i \right]
\]

similarly

\[
\hat{\alpha}_{p-} = e'_1 D_{n-} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) (1 - d_i) Z_i m(x_i) \right] + e'_1 D_{n-} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) (1 - d_i) Z_i \varepsilon_i \right]
\]

Then

\[
\sqrt{nh} (\hat{\alpha}_p - \alpha) = \sqrt{nh} (\hat{\alpha}_{p+} - \alpha - \hat{\alpha}_{p-}) \\
= e'_1 D_{n+} \left\{ \sqrt{nh} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i m(x_i) \right] + \left[ \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i \varepsilon_i \right] \right\} - \\
- e'_1 D_{n-} \left\{ \sqrt{nh} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) (1 - d_i) Z_i m(x_i) \right] + \left[ \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) (1 - d_i) Z_i \varepsilon_i \right] \right\}
\]

For the denominator terms \( D_{n+} \) and \( D_{n-} \),

\[
D_{n+}^{-1} = \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i Z'_i
\]

36
and each element of this matrix is given by

\[ [D_{n+1}]_{i,j} = \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i \left( \frac{x_i - \bar{x}}{h} \right)^{j+l-2} \]

which have asymptotic variance converging to zero since

\[
\text{Var} \left( [D_{n+1}]_{j,l} \right) = \frac{1}{(nh)^2} \text{Var} \left( \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i \left( \frac{x_i - \bar{x}}{h} \right)^{j+l-2} \right)
\]

\[\leq \frac{1}{nh} \mathbb{E} \left[ \frac{1}{h} k^2 \left( \frac{x - \bar{x}}{h} \right) d \left( \frac{x - \bar{x}}{h} \right)^{2(j+l-2)} \right]
\]

\[= \frac{1}{nh} \int_{\bar{x}}^{\bar{x}+h} \frac{1}{h} k^2 \left( \frac{x - \bar{x}}{h} \right) \left( \frac{x - \bar{x}}{h} \right)^{2(j+l-2)} f_0(x) dx + o_p(1)
\]

Note that the terms in the integral and the integral itself are \(O(1)\) and \(\frac{1}{nh} = o(1)\). Hence, \(\text{Var} \left( [D_{n+1}]_{j,l} \right) \rightarrow 0\). Now,

\[ [D_{n+1}]_{i,j} = \mathbb{E} \left[ [D_{n+1}]_{i,j} \right] + o_p(1)
\]

\[= \frac{1}{nh} \mathbb{E} \left[ \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i \left( \frac{x_i - \bar{x}}{h} \right)^{j+l-2} \right] + o_p(1)
\]

\[= \mathbb{E} \left[ \frac{1}{h} k \left( \frac{x_i - \bar{x}}{h} \right) d_i \left( \frac{x_i - \bar{x}}{h} \right)^{j+l-2} \right] + o_p(1)
\]

\[= \int_{\bar{x}}^{\bar{x}+h} \frac{1}{h} k \left( \frac{x - \bar{x}}{h} \right) \left( \frac{x - \bar{x}}{h} \right)^{j+l-2} f_0(x) dx + o_p(1)
\]

\[= \int_{0}^{M} k(u) u^{j+l-2} f_0(x+uh) dx + o_p(1)
\]

Let, \(\gamma^+_j = \int_{0}^{M} k(u) u^{j+l-2} f_0(x+uh) dx\) and \(\Gamma^+_p\) is the \((p+1) \times (p+1)\) matrix with

\((j, l)\) element \(\gamma^+_j\) for \(j, l = 1, \ldots, p+1\). Then,

\[D_{n+} \overset{p}{\rightarrow} (\Gamma^+_p)^{-1}
\]

Similarly,

\[D_{n-} \overset{p}{\rightarrow} (\Gamma^+_p)^{-1}
\]
where $\Gamma^*$ is the $(p + 1) \times (p + 1)$ matrix with $(j, l)$ element $\gamma_{j+l}^{-1}$ for $j, l = 1, \ldots, p + 1$, and $\gamma_j^{-1} = (-1)^j \int_0^M k(u) u^j f_o(\bar{x} - uh)\,dx$.

Now we will derive the asymptotic distribution of $\frac{1}{\sqrt{nh}} \sum_{i=1}^n k\left(\frac{x_i - \bar{x}}{h}\right) d_i Z_i \varepsilon_i$.

Following Porter (2003) I use the Cramer-Wold device to derive the asymptotic distribution. Let $\lambda$ be a nonzero, finite vector. Then,

$$
\sum_{i=1}^n E \left[ \frac{1}{\sqrt{nh}} \sum_{i=1}^n k\left(\frac{x_i - \bar{x}}{h}\right) d_i \lambda' Z_i \varepsilon_i \right]^{2+\zeta} \\
= \sum_{i=1}^n \left( \frac{1}{nh} \right)^{\frac{\hat{\zeta}}{2}} \frac{1}{nh} E \left[ k\left(\frac{x_i - \bar{x}}{h}\right)^{2+\zeta} d_i |\lambda' Z_i|^{2+\zeta} \varepsilon_i \right]^{2+\zeta} \\
\leq \left( \frac{1}{nh} \right)^{\frac{\zeta}{2}} h^{\sup_{x \in R} E \left[ |\varepsilon|^{2+\zeta} | x \right] E \left[ k\left(\frac{x - \bar{x}}{h}\right)^{2+\zeta} d \sum_{l=1}^p \lambda_l \left(\frac{x - \bar{x}}{h}\right)^l \right]^{2+\zeta} \\
= \left( \frac{1}{nh} \right)^{\frac{\hat{\zeta}}{2}} O(1) O(1) = O \left( \frac{1}{\sqrt{nh}} \right)^{\frac{\hat{\zeta}}{2}} = o(1)
$$

then, $\frac{1}{\sqrt{nh}} \sum_{i=1}^n k\left(\frac{x_i - \bar{x}}{h}\right) d_i Z_i \varepsilon_i$ follows Liapunov’s CLT. Note that,

$$
E \left[ \frac{1}{\sqrt{nh}} \sum_{i=1}^n k\left(\frac{x_i - \bar{x}}{h}\right) d_i Z_i \varepsilon_i \right] \\
= E \left[ \frac{1}{\sqrt{nh}} k\left(\frac{x_i - \bar{x}}{h}\right) d_i Z_i E [\varepsilon | x] \right] = 0
$$
and

\[ Var \left[ \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i \varepsilon_i \right] = \frac{1}{h} Var \left[ k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i \varepsilon_i \right] \]

\[ = \frac{1}{h} E \left[ k^2 \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i^2 \varepsilon_i^2 \right] \]

\[ = \frac{1}{h} E \left[ k^2 \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i Z_i^t E \left[ \varepsilon_i^2 | x \right] \right] \]

\[ = \int_{\mathcal{R}} \frac{1}{h} k^2 \left( \frac{x - \bar{x}}{h} \right) ZZ' \sigma^2(x) f_o(x) dx \]

It helps to remember that \( ZZ' \) is a function of the \( x \),

\[ ZZ' = \begin{bmatrix} 1 & (\frac{x - \bar{x}}{h}) & \cdots & (\frac{x - \bar{x}}{h})^p \\ (\frac{x - \bar{x}}{h}) & (\frac{x - \bar{x}}{h})^2 & \cdots & (\frac{x - \bar{x}}{h})^{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ (\frac{x - \bar{x}}{h})^p & (\frac{x - \bar{x}}{h})^{p+1} & \cdots & (\frac{x - \bar{x}}{h})^{2p} \end{bmatrix} \]

Let \( \delta^+_j = \int_{\mathcal{R}} \frac{1}{h} k^2 \left( \frac{x - \bar{x}}{h} \right) (\frac{x - \bar{x}}{h})^j \sigma^2(x) f_o(x) dx = \int_0^M k^2 (u) u^j \sigma^2(\bar{x} + uh) f_o(\bar{x} + uh) dx \) and \( \Delta^+ \) is the \((p + 1) \times (p + 1)\) matrix with \((j, l)\) element \( \delta_{j+l-2} \) for \( j, l = 1, \ldots, p + 1 \). Then,

\[ \frac{1}{\sqrt{n}h} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i \varepsilon_i \overset{p}{\rightarrow} N(0, \Delta^+) \]

Similarly we can show that

\[ \frac{1}{\sqrt{n}h} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i \varepsilon_i \overset{p}{\rightarrow} N(0, \Delta^-) \]

where \( \Delta^- \) is the \((p + 1) \times (p + 1)\) matrix with \((j, l)\) element \( \delta_{j+l-2} \) for \( j, l = 1, \ldots, p + 1 \), and \( \delta^-_j = \int_{\mathcal{R}} \frac{1}{h} k^2 \left( \frac{x - \bar{x}}{h} \right) (\frac{x - \bar{x}}{h})^j \sigma^2(x) f_o(x) dx = (-1)^j \int_0^M k^2 (u) u^j \sigma^2(\bar{x} - uh) f_o(\bar{x} - uh) dx \).
The bias term is given by

\[
\sqrt{n} h e_1 \left\{ D_{n+} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i m(x_i) \right] - D_{n-} \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) (1 - d_i) Z_i m(x_i) \right] \right\}
\]

Notice that if the rectangular kernel is used this is nothing else than the difference between the intercepts estimated by the linear projection of \( m(x) \) on \( Z \), above and below the cutoff point using only the data inside the bandwidth.

Note that,

\[
E \left[ \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i m(x_i) \right] = E \left[ \frac{1}{h} k \left( \frac{x_i - \bar{x}}{h} \right) d_i Z_i m(x_i) \right] = \int_{\bar{x}}^{\bar{x} + h} \frac{1}{h} k \left( \frac{x - \bar{x}}{h} \right) Z(x) m(x) dx = \int_0^{M} k(u) Z(\bar{x} + uh) m(\bar{x} + uh) dx
\]

and similarly,

\[
\frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x_i - \bar{x}}{h} \right) (1 - d_i) Z_i m(x_i) \overset{p}{\rightarrow} \int_{\bar{x} - h}^{\bar{x}} \frac{1}{h} k \left( \frac{x - \bar{x}}{h} \right) Z(x) m(x) dx = \int_0^{M} k(u) Z(\bar{x} - uh) m(\bar{x} - uh) dx
\]

Hence, the bias term can be approximated by,

\[
\lim_{n \to \infty} \left( \sqrt{n} h \right) e_1 \left\{ \left( \Gamma_n^+ \right)^{-1} \left[ \int_0^{M} k(u) Z(\bar{x} + uh) m(\bar{x} + uh) dx \right] - \left( \Gamma_n^- \right)^{-1} \left[ \int_0^{M} k(u) Z(\bar{x} - uh) m(\bar{x} - uh) dx \right] \right\}
\]

\[
\Box
\]

References


Figures for: “When the Practitioner Saved the Theorist: Theory and Practice of Inference in Regression Discontinuity”

Figure 1:

Regression Discontinuity Empirical Coverage

![Graph showing empirical coverage for different estimators and models. The x-axis represents the range around the cutoff, and the y-axis represents empirical coverage. The graph includes lines for Porter SE, Linear Approx. SE, True, and a 95% line.]

Figure 2:

Regression Discontinuity Empirical Coverage

![Graph showing empirical coverage for different estimators and models. The x-axis represents the range around the cutoff, and the y-axis represents empirical coverage. The graph includes lines for Porter SE, Linear Approx. SE, True, and a 95% line.]

N-W estimator - True: No X Model

N-W estimator - True: Linear Model
Figure 3:

Regression Discontinuity Empirical Coverage

N-W estimator, Bias Corrected - True: Linear Model

Range around cutoff

Empirical Coverage

Porter SE
Linear Approx. SE
True
95% line

Figure 4:

Regression Discontinuity Empirical Coverage

N-W estimator - True: Linear Model

Range around cutoff

Empirical Coverage

Porter Bias and SE
Porter Bias and true SE
True Bias and SE
True Bias and Porter SE
95% line
Figure 9: Regression Discontinuity Empirical Coverage
N-W estimator, Bias Corrected - True: Cubic Model

Figure 10: Regression Discontinuity Empirical Coverage
N-W estimator - True: Cubic Model
Figure 11: Regression Discontinuity Empirical Coverage
N-W estimator - True: Exponential Model
Range around cutoff
Empirical Coverage
Porter SE
Linear Approx. SE
True
95% line

Figure 12: Regression Discontinuity Empirical Coverage
N-W estimator, Bias Corrected - True: Exponential Model
Range around cutoff
Empirical Coverage
Porter SE
Linear Approx. SE
True
95% line
Figure 13:

Regression Discontinuity Empirical Coverage
N-W estimator - True: Exponential Model

Figure 14:

Regression Discontinuity Empirical Coverage
N-W estimator - True: No X Model