ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Bartalotti, Otávio

Working Paper

Regression Discontinuity and Heteroskedasticity Robust Standard Errors: Evidence from a Fixed-Bandwidth Approximation

IZA Discussion Papers, No. 11560

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Bartalotti, Otávio (2018) : Regression Discontinuity and Heteroskedasticity Robust Standard Errors: Evidence from a Fixed-Bandwidth Approximation, IZA Discussion Papers, No. 11560, Institute of Labor Economics (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/180578

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11560

Regression Discontinuity and Heteroskedasticity Robust Standard Errors: Evidence from a Fixed-Bandwidth Approximation

Otávio Bartalotti

MAY 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11560

Regression Discontinuity and Heteroskedasticity Robust Standard Errors: Evidence from a Fixed-Bandwidth Approximation

Otávio Bartalotti
Iowa State University and IZA

MAY 2018

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Regression Discontinuity and Heteroskedasticity Robust Standard Errors: Evidence from a Fixed-Bandwidth Approximation*

In regression discontinuity design (RD), for a given bandwidth, researchers can estimate standard errors based on different variance formulas obtained under different asymptotic frameworks. In the traditional approach the bandwidth shrinks to zero as sample size increases; alternatively, the bandwidth could be treated as fixed. The main theoretical results for RD rely on the former, while most applications in the literature treat the estimates as parametric, implementing the usual heteroskedasticity-robust standard errors. This paper develops the "fixed-bandwidth" alternative asymptotic theory for RD designs, which sheds light on the connection between both approaches. I provide alternative formulas (approximations) for the bias and variance of common RD estimators, and conditions under which both approximations are equivalent. Simulations document the improvements in test coverage that fixed-bandwidth approximations achieve relative to traditional approximations, especially when there is local heteroskedasticity. Feasible estimators of fixed-bandwidth standard errors are easy to implement and are akin to treating RD estimators as locally parametric, validating the common empirical practice of using heteroskedasticity-robust standard errors in RD settings. Bias mitigation approaches are discussed and a novel, bootstrap higher-order bias correction procedure based on the fixed bandwidth asymptotics is suggested.

JEL Classification: C12, C21

Keywords: average treatment effect, locally parametric inference, local polynomial estimators, fixed bandwidth, heteroskedasticity robust standard errors, bias correction

Corresponding author:
Otávio Bartalotti
479 Heady Hall
Department of Economics
Iowa State University
Ames, IA, 50011
USA

E-mail: bartalot@iastate.edu

* I would like to thank Gray Calhoun, Jeff Wooldridge, Tim Vogelsang, Gary Solon, Matias Cattaneo, Peter Schmidt, Steve Dieterle, Quentin Brummet, Valentin Verdier, Ilya Rahkovsky, Keith Finlay, Michael Darden, Brigham Frandsen, two anonymous referees, and participants at several seminars and conferences where earlier versions of this paper were presented for valuable comments and discussions. I also thank Yang He for his helpful work as research assistant for this paper. Forthcoming in Journal of Econometric Methods.

1 Introduction

Regression discontinuity (RD) designs have been propelled to the spotlight of economic analysis in recent years,[1] especially the policy and treatment evaluation literatures, as a form of estimating treatment effects in a non-experimental setting. In this design, treatment is assigned based on values of an observed characteristic, with the probability of receiving treatment jumping discontinuously at a known threshold. RD's appeal stems from the relatively weak assumptions necessary for the nonparametric identification of treatment effects, specifically smoothness of the conditional expectation of outcome for treated and untreated individuals at the cutoff. This paper examines the effect of the additional assumptions that are explicitly and implicitly made to justify estimation and inference and shows that those assumptions can be weakened by using a heteroskedasticity-robust variance estimator, which is a common approach in empirical research but has not been supported by previous theoretical research.

Local polynomial estimators are the most common choice in empirical and theoretical work due to ease of implementation and flexibility. These kernel-based estimators rely on fitting a polynomial function to a range of the data, the size of which is determined by a bandwidth, h, just around the threshold. To perform inference in this setting, additional assumptions are usually imposed. In particular, asymptotic approximations assume smoothness of the conditional variance of the outcome, $\sigma^2(x)$. That assumption is somewhat restrictive and not necessary as it will be discussed below. Implementation of the local polynomial estimators depends on the choice of the bandwidth, which can be benchmarked against several bandwidth selectors available in the literature.

The traditional nonparametric asymptotic approximations ("small-h") provide formulas for the variance of such estimators as the bandwidth shrinks towards zero asymptotically. Those formulas reflect the fact that the distribution of the kernel based estimator is asymptotically equivalent to the distribution with homoskedastic data when the smoothing parameter shrinks, which may have led researchers to overlook the need to develop robust inference in small samples. A similar point has been recently made by Kim, Sun and Yang (2017) when discussing time series dependence. Moreover, in practice variance estimators that rely on homoskedasticity are discouraged even though they are supported by the theory. Hence, most researchers often use a different "pre-asymptotics" variance formula to compute standard errors, based on the White-Huber-Eicker robust standard errors or some nearest neighbor alternative.[2] This paper provides an analysis of when and why the common practice of using White-Huber-Eicker robust standard errors in the RD context is appropriate.

I focus on presenting an alternative asymptotic approximation for the standard RD treatment effects estimator using a fixed bandwidth that relaxes the smoothness conditions imposed on $\sigma^2(x)$. This "fixed-h"

[1] Lee and Lemieux (2009), in a broad review of the RD literature, presented a list of more than 60 papers that have applied RD design to many different contexts.

[2] The use of such variance estimators in RD designs have been suggested based on intuitive arguments or using "pre-asymptotics" expressions that did not follow directly from the "small-h" theoretical asymptotic approximations—see, for example, Lee and Lemieux (2009) and Calonico et al. (2014).

approximation provides expressions for the estimator's asymptotic variance that incorporate the bandwidth used by the researcher, and lead naturally to the standard error formulas used in the applied literature, clarifying the assumptions behind its use. The results show that the heteroskedasticity robust standard errors continue to provide valid inference under these relaxed assumptions. Furthermore, these standard errors are appropriate for any specific bandwidth chosen.

This paper shows that this form of the variance estimator is valid both under an asymptotic nesting where the bandwidth parameter, h, is fixed, and as under conventional RD asymptotics where $h \to 0$. However, the traditional asymptotic variance estimator is not consistent for fixed bandwidths. Consequently the common empirical strategy of using robust standard errors in RD studies is appropriate especially when heteroskedasticity is a concern.

While I focus on a setting where the bias is "small" and does not affect the asymptotic distribution of the RD estimator, an alternative approximation for the estimator's asymptotic bias that differs from the usual approximations in the literature is presented and provides additional intuition on the robustness-precision trade-off facing the researcher in RD designs. A discussion on how to address bias in practice is presented.

We first consider the procedure proposed by Calonico et al (2014) implement an analytical bias correction reduction based on a first order expansion of the bias, reducing the order of the leading bias term. The framework proposed here is compatible with such bias reduction procedures. Furthermore, I explore a novel approach to bias mitigation based on a higher order correction inspired by the fixed bandwidth asymptotic bias approximations. That strategy is easily implemented by modifying the wild bootstrap RD procedure in Bartalotti, Calhoun and He (2017). Simulation exercises attest that such approach can significantly improve test performance relative to first order based corrections.

This work contributes to the emerging literature on inference for treatment effects in the context of RD designs. Hahn et al. (1999, 2001) and Lee (2008) presented the conditions for identification and estimation in RD designs. Porter (2003) provided 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. Calonico et al. (2014) studied asymptotic approximations for the bias-corrected local polynomial RD estimator as described above. Other studies about estimation, inference, and bandwidth choice in RD designs include Imbens and Kalyanaraman (2012), Cattaneo et al. (2015), Calonico et al. (Forthcoming), Bartalotti and Brummet (2017) among others. McCrary (2008) studied specification testing. Finally, a broad review of the theoretical and applied literature, with emphasis on the identification of the parameter of interest and its potential interpretations, can be found in Imbens and Lemieux (2008) and Lee and Lemieux (2009).

The asymptotic framework using fixed bandwidths follows a growing literature that recognizes the potential for practical improvements in inference procedures in nonparametric methods. Notably, Neave (1970), in the context of spectral density estimation, obtained more accurate approximations to the variance of nonparametric spectral estimates. Neave's work was later extended by Hashimzade and Vogelsang (2008).

Similarly, Fan (1998) provided an alternative approximation for goodness-of-fit tests for density function estimates, obtaining improved approximations to the asymptotic behavior of the test and critical values for inference. More recent studies, on the context of time-series or spatial dependence, analyze and justify the use asymptotic variance formulas based on fixed-bandwidth approximations when pursuing heteroskedasticity and autocorrelation and spatial dependence robust inference (Chen, Liao and Sun, 2014; Sun, 2014; Bester et al. (2016); Kim, Sun and Yang, 2017). This paper contributes to that literature in the context of local polynomial estimators and RD designs.

Section 5 provides simulation evidence that standard errors used by practitioners work well relative to the ones based in formulas suggested by the traditional approach ($h \to 0$). The robust standard errors based on the fixed bandwidth formulas perform well for larger bandwidths, especially when heteroskedasticity is present. The intuition is that, for larger bandwidths, the homoskedasticity condition becomes less likely to hold as it would need to be valid over a larger support of the data. Finally, I provide an empirical application using Lee (2008), exemplifying with actual data the improvements obtained.

2 Model and Estimator

The interest lies in estimating the average treatment effect, τ, of a certain policy affecting part of a population of interest.[3] There are two types of RD designs: sharp and fuzzy. They differ in regard to the assignment of treatment and to the impact of the discontinuity on the assignment process. The main body of the paper focuses on sharp RD designs. The discussion and extensions for the fuzzy design are presented below.

In the sharp design, the treatment status, d, is a deterministic function of a so-called "running" variable, x, such that,

$$d_i = \left\{ \begin{array}{c} 1 \text{ if } x_i \geq \overline{x} \\ 0 \text{ if } x_i < \overline{x}, \end{array} \right\}$$

where $\overline{x}$ 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 et al. (2001) and Porter (2003), under some weak smoothness assumptions, the average treatment effect can be estimated by comparing points just above and just below the discontinuity. If the running variable is continuous, as well as the conditional expectations of Y_0 and Y_1, the discontinuity in treatment assignment at $\overline{x}$ provides the opportunity to identify the average treatment effect at the cutoff without any additional parametric functional form restrictions on the conditional expectations

[3] As discussed in Porter (2003), Imbens and Lemieux (2008), Lee and Lemieux (2009), and Cattaneo et al. (2015), RD designs are closely associated with the treatment effect literature. Angrist and Pischke (2009) have provided a simple introduction to the intuition of regression discontinuity.

of the outcome variable. The average causal effect of the treatment at the discontinuity is

$$\begin{aligned}\tau &\equiv E\left[Y_1 - Y_0 \mid X = \overline{x}\right] \\ &= \lim_{x \downarrow \overline{x}} E\left[Y \mid X = x\right] - \lim_{x \uparrow \overline{x}} E\left[Y \mid X = x\right].\end{aligned}$$

The sharp RD design uses the discontinuity in the conditional expectation of Y given X to uncover the average treatment effect at the cutoff. For a comprehensive review of RD designs and their applications and interpretation, see Lee and Lemieux (2009).

I focus on estimates of τ obtained using local polynomial estimation, which is the most common in applied work. The order p local polynomial estimator is defined as follows. In the sharp design case, given data $(y_i, x_i)_{i=1,2,...,n}$, let $d_i = 1[x_i \geq \overline{x}]$, $k(\cdot)$ be a kernel function, and h denote a bandwidth that controls the size of the local neighborhood to be averaged over. Also, define the $p+1 \times 1$ vector $Z(x) = \left(1, \left(\frac{x-\overline{x}}{h}\right), \left(\frac{x-\overline{x}}{h}\right)^2, ..., \left(\frac{x-\overline{x}}{h}\right)^p\right)'$ and let $\left(\widehat{\alpha}_{p+}, \widehat{\beta}_{p+}\right)'$ be the solution to the minimization problem:

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

while, similarly, $\left(\widehat{\alpha}_{p-}, \widehat{\beta}_{p-}\right)$ minimizes the same objective function, but with $1 - d_i$ replacing d_i. The estimator of the parameter of interest is given by

$$\widehat{\tau} \equiv \widehat{\alpha}_p = \widehat{\alpha}_{p+} - \widehat{\alpha}_{p-}.$$

In the fuzzy design the probability of receiving treatment still changes discontinuously at the threshold, but is not required to go from 0 to 1,

$$\lim_{x \downarrow \overline{x}} \Pr(d \mid X = x) \neq \lim_{x \uparrow \overline{x}} \Pr(d \mid X = x)$$

This framework allows for a greater range of applications since it includes cases in which the incentives to receive treatment change discontinuously at the threshold, but are not strong enough to induce all individuals above it to be treated. The average treatment effect at the cutoff can be identified by the ratio of the change in the conditional expectation for the outcome variable to the change in the conditional probability of receiving treatment (Imbens and Lemieux, 2008):

$$\tau_F \equiv \frac{\lim_{x \downarrow \overline{x}} E\left[Y \mid X = x\right] - \lim_{x \uparrow \overline{x}} E\left[Y \mid X = x\right]}{\lim_{x \downarrow \overline{x}} E\left[d \mid X = x\right] - \lim_{x \uparrow \overline{x}} E\left[d \mid X = x\right]}$$

The estimator of the parameter of interest is given by the ratio

$$\widehat{\tau}_F = \frac{\widehat{\alpha}_p}{\widehat{\theta}_p}$$

where $\widehat{\alpha}_p$ is the estimator described above, and $\widehat{\theta}$ is the estimator for the change in the probability of being in the treated group at the cutoff which is obtained by using the same estimators with the treatment assignment variable, D_i, as the dependent variable.

In practice, the implementation of RD designs is not without challenges, even in simple cases, that require implicit assumptions or choices by the researcher not fully reflected in this basic set up. Importantly for the aims of this paper, is that several of these situations can be better understood by relying on an alternative asymptotic approximation based on a fixed bandwidth. That not only sheds light on the assumptions and issues dealt implicitly by the practitioners, but also naturally allows for the careful use tools in parametric analysis to approach these challenges.

To develop this framework, I consider the following assumptions, which are more flexible than the ones used in the literature on RD designs, e.g., Hahn et al. (2001), Porter (2003), etc. Let $F_o(x)$ denote the density of x, $m(x) = E[y \mid x]$, and $\overline{x}$ is the cutoff that determines treatment. Finally, define $\varepsilon = y - E[y \mid X = x] = y - m(x)$ and $\sigma^2(x) = E\left[\varepsilon^2 \mid X = x\right]$., Importantly, we allow for the conditional variance, $\sigma^2(x)$, to be not continuous and non-smooth in a region around the cutoff.

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

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

$$\alpha = \lim_{x \downarrow \overline{x}} E[y \mid X = x] - \lim_{x \uparrow \overline{x}} E[y \mid X = x].$$

For the compact interval $\dot{\aleph}$, defined as $[\overline{x} - h, \overline{x} + h]$, (a) $F_o(\overline{x}+h) - F_o(h)$ and $F_o(\overline{x}) - F_o(\overline{x}-h)$ are bounded away from zero; (b) $m(x)$ is l_m times continuously differentiable for $x \in \aleph \setminus \{\overline{x}\}$, and m is continuous at $\overline{x}$ with finite right- and left-hand derivatives to order l_m; and (c) the running variable, x, has enough variation within $\dot{\aleph}$ that we can identify $\left(\widehat{\alpha}_{p+}, \widehat{\beta}_{p+}\right)'$ and $\left(\widehat{\alpha}_{p-}, \widehat{\beta}_{p-}\right)'$.

Assumption 3 *(a) $\sigma^2(x) = E\left[\varepsilon^2 \mid X = x\right]$ exists and is well defined on $\dot{\aleph}$.*

(b) For some $\zeta > 0$, $E\left[|\varepsilon|^{2+\zeta} \mid X = x\right]$ is uniformly bounded on $\aleph$.

Assumption 1 is very standard in the RD literature and does not restricts the support of the kernel being used in estimation significantly for most applications. Assumptions 2 and 3 relax some of the usual smoothness conditions present in the literature. A relevant feature of the analysis performed in the next section is that fixed bandwidth asymptotics will allow the researcher to directly incorporate heteroskedasticity into the asymptotic distribution of the estimates of interest, including more complex dependence problems that are generally intractable in a shrinking bandwidth framework, e.g., clustering (Bartalotti and Brummet, 2017). A similar point has been made by Kim, Sun and Yang (2017) in the time series context. Most importantly, the conditional variance is allowed to not be continuous within the bandwidth around the cutoff. Hence, the framework developed here can be used to analyze cases in which the running variable might exhibit discreteness, which could be due to the nature of the running variable, as described by Lee and Card (2008), heaping (Barreca, Lindo and Waddell, 2016), measurement error (Dong, 2015 and Bartalotti, Brummet and Dieterle, 2017) or some other characteristic of the application at hand.

3 Asymptotic Distributions

To derive the asymptotic properties of the estimator for τ, the usual regularity and smoothness conditions in the literature are sufficient for both the traditional and fixed-h approximations, e.g., Hahn et al. (2001), Porter (2003), etc. As discussed in Section 2, we are able to relax some of the smoothness conditions by using a fixed bandwidth approach, which is particularly useful when dealing with scenarios in which the usual nonparametric approach is not feasible. In the following, let $F_o(x)$ denote the density of x and $m(x)$ denote the conditional expectation of y given x, i.e., $m(x) = E[y \mid x]$, and $\overline{x}$ is the value of the running variable in which the discontinuity occurs. Finally, define $\varepsilon = y - E[y \mid X = x] = y - m(x)$ and $\sigma^2(x) = E\left[\varepsilon^2 \mid X = x\right]$.

Using the traditional nonparametric asymptotic approximations, small-h, Hahn et al. (2001), Porter (2003), Imbens and Lemieux (2008) among others obtain the following asymptotic variance formula for the (scaled) $\widehat{\tau}$,

$$V_{small-h} = \frac{\sigma^{2+}(\overline{x}) + \sigma^{2-}(\overline{x})}{f_o(\overline{x})} e_1' \Gamma^{-1} \Delta \Gamma^{-1} e_1 \tag{1}$$

where $e_1' \Gamma^{-1} \Delta \Gamma^{-1} e_1$ is a constant scalar which depends only on the order of the local polynomial, p, and the kernel used.[4] Crucially for practical purposes, it does not depend on the bandwidth (h) by construction. That is natural since the approximation is based on a shrinking bandwidth around the cutoff. However, that does not capture heteroskedasticity or changes in the values for the density of x in the range of data actually used in practice. That issue has been recognized in the literature, e.g. Imbens and Lemieux (2008) and Calonico et al. (2014) and by practitioners who commonly implement the usual heteroskedasticity robust White-Huber-Eicker standard errors (or weighted analogues depending on the choice of kernel) when implementing RD based on intuitive or "pre-asymptotics" arguments.

Interestingly these heteroskedasticity robust standard errors can also be justified by an approximation under the asymptotic nesting where the bandwidth parameter, h, is fixed. The fixed-h asymptotic approximation can be summarized in the following result.

Theorem 1 *Suppose assumptions 1, 3 and 2 hold. If h is fixed, positive, as $n \to \infty$, then*

$$Avar\left[\sqrt{nh}(\widehat{\alpha}_p - \alpha)\right] \xrightarrow{p} V_{fixed-h},$$

where

$$V_{fixed-h} = e_1' \left[\left(\Gamma_+^*\right)^{-1} \Delta_+^* \left(\Gamma_+^*\right)^{-1} + \left(\Gamma_-^*\right)^{-1} \Delta_-^* \left(\Gamma_-^*\right)^{-1} \right] e_1$$

Additionally, if $Bias\left[\sqrt{nh}(\widehat{\alpha}_p - \alpha)\right] = o_p(1)$,*then*

$$\sqrt{nh}(\widehat{\alpha}_p - \alpha) \xrightarrow{d} N\left(0, V_{fixed-h}\right) \tag{2}$$

where $\Gamma_{+(-)}^*$, $\Delta_{+(-)}^*$, e_1, γ_j^+, γ_j^-, δ_j^+, and δ_j^- are defined in the appendix.

[4] See appendix for the definition of $e_1' \Gamma^{-1} \Delta \Gamma^{-1} e_1$.

The fixed-h asymptotic variance formula is given by $V_{fixed-h}$, and depends directly on the behavior of $\sigma^2(x)$, $f_o(x)$ *inside* the bandwidth around the cutoff, p, and the kernel function through $\left(\Gamma_+^*\right)^{-1}\Delta_+^*\left(\Gamma_+^*\right)^{-1}$ and $\left(\Gamma_-^*\right)^{-1}\Delta_-^*\left(\Gamma_-^*\right)^{-1}$. Hence, it captures the impact of h on the asymptotic variance accounting for potential local heteroskedasticity. This asymptotic approximation provides a natural and intuitive way to understand the sources of improvement in inference observed in practice when using these standard errors, as discussed below.

There are two noteworthy cases in which the formulas for the fixed-h asymptotic variance (and bias) simplify to the small-h formulas. First, if in the bandwidth around the cutoff, $f_o(x)$, $\sigma^2(x)$ and $m(x)$ are continuous, then when $h \to 0$, [5]

$$\lim_{h\to 0} V_{fixed-h} = V_{small-h}.$$

Hence, if h is small, fixed-h and small-h provide similar approximations to the asymptotic behavior of $\widehat{\alpha}_p$. This is intuitive since the fixed bandwidth approximation is valid for any value of h, and the difference between the formulas arising from the behavior of $\sigma^2(x)$, $f_o(x)$ *inside* the bandwidth will likely be small as the support of the running variable on which they are being evaluated becomes smaller.

Second, if $f_o(x)$ and $\sigma^2(x)$ are constant around the cutoff, the fixed-h asymptotic variance formula simplifies to the small-h formula, i.e., $V_{fixed-h} = V_{small-h}$.[6] This fact makes clear that the differences between the fixed bandwidth and traditional approximations are due to incorporating the behavior of $f_o(x)$ and $\sigma^2(x)$ in the ranges around the cutoff in the first case, while considering only its values at the cutoff, $f_o(\overline{x})$ and $\sigma^2(\overline{x})$ in the second.

3.1 Addressing the Asymptotic Bias

The asymptotic normality described in Theorem 1 holds even if the bias is non-negligible in the sense that $\sqrt{nh}(\widehat{\alpha}_p - E[\widehat{\alpha}_p]) \xrightarrow{d} N\left(0, V_{fixed-h}\right)$ similarly to the results presented by Pagan and Ullah (1999). Even though not used directly in the asymptotic variance discussion which is the focus of this paper, an ancillary result is the asymptotic bias formula under the fixed bandwidth asymptotic nesting (see appendix),

$$Bias\left[\sqrt{nh}(\widehat{\alpha}_p - \alpha)\right] = e_1' \left\{ \begin{array}{l} \left(\Gamma_+^*\right)^{-1}\left[\int_0^\infty k\left(u\right) Z(\overline{x}+uh) m(\overline{x}+uh) dF_o(\overline{x}+uh)\right] - \\ -\left(\Gamma_-^*\right)^{-1}\left[\int_0^\infty k\left(u\right) Z(\overline{x}-uh) m(\overline{x}-uh) dF_o(\overline{x}-uh)\right] \end{array} \right\}. \tag{3}$$

A more familiar representation of the asymptotic bias term can be obtained if we are willing to assume differentiability of $m(x)$ and $f_o(x)$ up to order $(p+2)$ in the local support $\dot{\aleph}$ and approximate it by polynomial of order $p+2$. Then one can rewrite the asymptotic bias term above using a (cruder) approximation as

[5] Corollary 1, in the appendix, follows the sequential asymptotics literature usually implemented in the context of series estimators (e.g., Stock and Yogo 2005).

[6] See Corollary 2 in the appendix.

$$B_{small-h} = \frac{h^{p+1}}{(p+1)!}\left[m^{(p+1)+}(\overline{x}) - (-1)^{p+1}m^{(p+1)-}(\overline{x})\right] e_1'\Gamma^{-1}\begin{bmatrix}\gamma_{p+1}\\ \vdots \\ \gamma_{2p+1}\end{bmatrix} + \frac{h^{p+2}}{(p+2)!}\left[m^{(p+2)+}(\overline{x}) - (-1)^{p+2}m^{(p+2)-}(\overline{x})\right] e_1'\Gamma^{-1} \tag{4}$$

which are the formulas obtained in the traditional nesting where $h \to 0$. (Porter, 2003; Calonico et al., 2014)[7]

Both equations provide some intuition for conditions under which the condition of negligible asymptotic bias, $Bias\left[\sqrt{nh}(\widehat{\alpha}_p - \alpha)\right] = o_p(1)$, in Theorem 1 holds.

Under the more general conditions discussed in Section 2 with fixed bandwidth, the bias will be asymptotically negligible if the local polynomial of order p correctly captures the relevant features of $m(x)$ in the neighborhood of the cutoff. That could be because the model is correctly specified (which is often implicitly assumed by practitioners when implementing RD) and, hence the bias vanishes asymptotically as $n \to \infty$. More interestingly, that condition will hold *with h fixed* if the researcher treats p as embedded in an increasing sequence as n increases as in a series estimator at the cost of imposing additional smoothness conditions on $m(x)$ in the neighborhood around the cutoff. Details on these conditions could be found on Pagan and Ullah (1999) section 3.9 and Andrews (1991). It is important to notice that in the RD case, the restrictions on $m(x)$ need only to hold in the smaller support $\dot{\aleph}$ since our focus is to improve the polynomial fit only on this neighborhood, hence maintaining a substantial amount of the flexibility afforded by the RDD.

In the case where the running variable is continuous and the usual smoothness conditions hold, the asymptotic bias described above will be negligible if $nh^{2p+3} \to 0$ (Hahn, Todd and van der Klaauw, 2001; Porter, 2003). As Calonico et al. (2014) points out, in that case the usual MSE-optimal bandwidths proposed by Imbens and Kalyanaraman (2012) are too"large" because they do not satisfy this condition. They proposed an analytical bias correction by estimating the first term in Equation 4, reducing the order of the remaining asymptotic bias. Therefore, if such bias correction procedure were to be implemented , the condition $Bias\left[\sqrt{nh}(\widehat{\alpha}_p - \alpha)\right] = o_p(1)$ would hold if $nh^{2p+5} \to 0$. This bias correction procedure has become standard in the literature and is readily available in STATA and R in the package *rdrobust* (Calonico et al., 2017).

As is well known, unless the true specification of the population model is known, the local polynomial estimator is inconsistent. Even though the treatment effect is nonparametrically identified, in practice the local polynomial estimator of the RD design will potentially provide a biased estimate of the average treatment effect, unless the polynomial used correctly specifies the conditional expectation of Y in the bandwidth around the cutoff. Nevertheless, the formulas for $B_{fixed-h}$ provide important additional intuition relative to $B_{small-h}$. While the latter depends on a first-order (or second-order) approximation of the bias

[7]Note that to recover exactly the same notation as used in Porter (2003), one needs to multiply $B_{small-h}$ by the scaling term $\sqrt{nh}$. Then, the first term in $B_{small-h}$ converges to $\frac{C_a}{(p+1)!}$, as in Porter (2003).

term, the former directly captures the effect of the bandwidth used and its impacts on the approximation provided by the a local polynomial of order p in the whole support within the bandwidth.

It is worth noting that both fixed-h and small-h asymptotic approximations are based on the **same estimator** for $\widehat{\alpha}$. For a given bandwidth used in practice, the initial bias present in the estimate is given. Using the small-h approximation for standard errors does nothing to eliminate potential bias in practice and any potential bias should be treated as a separate empirical issue. When implementing RD, the researcher can consider some potential bias mitigation strategies. The analytical bias correction procedures proposed by Calonico et al. (2014) and its bootstrap equivalent in Bartalotti, Calhoun and He (2017) could be used, and rely on first order approximations and have performance dependent on the curvature of $m(x)$ and the bandwidth used. Both adjustments require an additional adjustment to the standard errors due to the estimation of the first-order bias adjustment.

The key insight obtained by the fixed bandwidth asymptotic bias presented above is that using a first-order bias adjustment might not be sufficient to capture the complexity of $m(x)$. Especially for larger bandwidths, one would expect that approach to become less reliable in controlling the asymptotic bias. Hence, the fixed bandwidth asymptotic framework suggests two related empirical strategies of bias mitigation.

As described above, one could implement a series estimator within the fixed bandwidth such that the complexity of the polynomial fitted locally (its order p) increases as more data is accumulated around the cutoff.

Second, the wild bootstrap procedure in Bartalotti, Calhoun and He (2017) can be easily be adjusted to provide higher order bias correction and inference, by estimating a higher order local polynomial, e.g. $(p+2)$, in step 1 of Algorithm 1 in Bartalotti, Calhoun and He (2017, p. 431). Simulations comparing coverage rates obtained by the usual analytical first-order bias mitigation (Calonico et al, 2014) and the higher order bootstrap approach are presented in the Appendix Section 8.1.

Intuitively, the potential bias is due to the potential local model misspecification since the local polynomial used might not be able to correctly capture $m(x)$'s features within the bandwidth. To that point, it is interesting to draw a parallel between $B_{fixed-h}$ with the issue of model misspecification in parametric models (White 1982, 1996). The problem of estimating the average treatment effect at the cutoff can be seen as one of correctly estimating $E[Y \mid X]$ on both sides of the cutoff. By using a relatively small bandwidth, we fit the conditional expectation on a restricted support, and hence expect a polynomial of order p to produce a better fit than if we were trying to fit $E[Y \mid X]$ globally. This better fit is the benefit associated with a local approach, since it allows the conditional expectation to be unrestricted *outside* the bandwidth.

3.2 Fuzzy Regression Discontinuity

The Fuzzy RD case follows the delta method applied to $\widehat{\tau}_F = \frac{\widehat{\alpha}_p}{\widehat{\theta}_p}$ (Porter, 2003),

$$Avar\left[\sqrt{nh}\left(\widehat{\tau}_F - \tau_F\right)\right] \xrightarrow{p} V_{FRD-fixed-h} = \frac{1}{\theta^2}V_\alpha - 2\frac{\alpha}{\theta^3}C_{\alpha\theta} + \frac{\alpha^2}{\theta^4}V_\theta,$$

where V_α and V_θ are the same asymptotic variances presented in Theorem 1 with the outcome and treatment status as the variable of interest. If, additionally $Bias\left[\sqrt{nh}\left(\widehat{\tau}_F - \tau_F\right)\right] = o_p(1)$, and a multivariate CLT holds for $\begin{pmatrix} \sqrt{nh}(\widehat{\alpha} - \alpha) \\ \sqrt{nh}(\widehat{\theta} - \theta) \end{pmatrix}$ (Pagan and Ullah, 1999) then

$$\sqrt{nh}\left(\widehat{\tau}_F - \tau_F\right) \xrightarrow{d} N\left(0, V_{FRD-fixed-h}\right) \tag{5}$$

The proof follows directly from the delta method and is omitted.

All the terms that appear in the asymptotic distribution above, except for $C_{\alpha\theta}$, can be obtained from theorem 1.

Theorem 2 *Let $\sigma_{\varepsilon\eta}(x) = E\left[\varepsilon\eta \mid X = x\right]$. If $\widehat{\alpha}$ and $\widehat{\theta}$ are the local polynomial estimators and the conditions of theorem 1 hold for both estimators, then*

$$C_{\alpha\theta} = e_1' \left[\left(\Gamma_+^*\right)^{-1} \Delta_+^\rho \left(\Gamma_+^*\right)^{-1} + \left(\Gamma_-^*\right)^{-1} \Delta_-^\rho \left(\Gamma_-^*\right)^{-1}\right] e_1$$

where $\Delta_{+(-)}^\rho = \begin{bmatrix} \rho_0^{+(-)} & \cdots & \rho_p^{+(-)} \\ \vdots & \ddots & \vdots \\ \rho_p^{+(-)} & \cdots & \rho_{2p}^{+(-)} \end{bmatrix}$,

$\rho_j^+ = \int_0^\infty k^2(u)\, u^j \sigma_{\varepsilon\eta}(\overline{x} + uh) dF_o(\overline{x} + uh)$,

$\rho_j^- = (-1)^j \int_0^\infty k^2(u)\, u^j \sigma_{\varepsilon\eta}(\overline{x} - uh) dF_o(\overline{x} - uh)$, Γ_+^* *and* Γ_-^* *are defined as in previous results.*

The proof is presented in the Appendix. Similarly to the results for the Sharp RD, the asymptotic covariance formula converges to the small-h asymptotic covariance both as $h \to 0$, or for fixed-h if in the bandwidth around the cutoff, $f_o(x)$ and $\sigma_{\varepsilon\eta}(x)$ are constant.[8]

4 Variance Estimators

To perform inference about α, appropriate estimates for the asymptotic variance formulas from theorem 1 are necessary. The components of $V_{fixed-h}$, Δ_+^* and Γ_+^*, have typical elements given by

$$\gamma_j^+ = \int_0^\infty k(u)\, u^j dF_o(\overline{x} + uh) = E\left[h^{-1} k\left(\frac{\overline{x} - x}{h}\right)\left(\frac{\overline{x} - x}{h}\right)^j d\right]$$

$$\delta_j^+ = \int_0^\infty k^2(u)\, u^j \sigma^2(\overline{x} + uh) dF_o(\overline{x} + uh) = E\left[h^{-1} k\left(\frac{\overline{x} - x}{h}\right)^2\left(\frac{\overline{x} - x}{h}\right)^j d\varepsilon^2\right]$$

[8] See Corollary 4 and 3 in the appendix.

and similarly for γ_j^- and δ_j^-. A natural estimator for the asymptotic variance is given by their sample analogues, using the estimated residuals $\widehat{\varepsilon} = y - \widehat{m}(x)$,

$$\widehat{\gamma}_j^+ = (nh)^{-1} \sum_{i=1}^{n} k\left(\frac{\overline{x} - x_i}{h}\right) \left(\frac{\overline{x} - x_i}{h}\right)^j d_i$$
$$\widehat{\delta}_j^+ = (nh)^{-1} \sum_{i=1}^{n} k\left(\frac{\overline{x} - x_i}{h}\right)^2 \left(\frac{\overline{x} - x_i}{h}\right)^j d_i \widehat{\varepsilon}_i^2,$$

which are consistent by standard arguments in both asymptotic frameworks.

Then, the natural plug-in estimator of the fixed-h variance-covariance matrix is given by

$$\left[\left(\widehat{\Gamma}_+^*\right)^{-1} \widehat{\Delta}_+^* \left(\widehat{\Gamma}_+^*\right)^{-1} + \left(\widehat{\Gamma}_-^*\right)^{-1} \widehat{\Delta}_-^* \left(\widehat{\Gamma}_-^*\right)^{-1}\right]. \tag{6}$$

These are simple averages of the data and kernel weights, and they have the familiar "sandwich form" (Fan and Gijbels 1996). This estimator is analogous to the usual heteroskedasticity robust standard errors in a general weighted least squares framework, and it comes naturally from the fixed-h framework developed above. The fixed-h approach provides a framework that justifies the use of such estimators by practitioners. Intuitively, the variance estimators are "robust to the choice of bandwidths," since they are valid for any finite h, take into consideration the impact of higher order polynomials on the estimator's variance and are flexible regarding the conditional variance and density of X around the cutoff. Note that, $\left[\left(\widehat{\Gamma}_+^*\right)^{-1} \widehat{\Delta}_+^* \left(\widehat{\Gamma}_+^*\right)^{-1} + \left(\widehat{\Gamma}_-^*\right)^{-1} \widehat{\Delta}_-^* \left(\widehat{\Gamma}_-^*\right)^{-1}\right] \xrightarrow{p} \left[\left(\Gamma_+^*\right)^{-1} \Delta_+^* \left(\Gamma_+^*\right)^{-1} + \left(\Gamma_-^*\right)^{-1} \Delta_-^* \left(\Gamma_-^*\right)^{-1}\right] = V_{fixed-h}$.

In the rectangular kernel case, the variance estimator in equation (6) simplifies to the usual heteroskedastic robust variance estimator when using the data just above and below the cutoff.

Variance estimators based on small-h asymptotic variance formulas, as proposed by Porter (2003) and Lee and Lemieux (2009), are not fully robust to local heteroskedasticity. For example, Porter (2003) suggested an estimator for the variance of $\widehat{\alpha}$ based on the small-h approximation that requires only the estimation of the density of x *and* conditional variance of the errors *at* the cutoff. Let

$$\widehat{\sigma}^{2+}(\overline{x}) = \frac{(nh)^{-1} \sum_{i=1}^{n} k\left(\frac{\overline{x} - x_i}{h}\right) d_i \widehat{\varepsilon}_i^2}{\frac{1}{2}\widehat{f}_o(\overline{x})}, \tag{7}$$

$$\widehat{\sigma}^{2-}(\overline{x}) = \frac{(nh)^{-1} \sum_{i=1}^{n} k\left(\frac{\overline{x} - x_i}{h}\right) (1 - d_i) \widehat{\varepsilon}_i^2}{\frac{1}{2}\widehat{f}_o(\overline{x})}, \tag{8}$$

$$\widehat{f}_o(\overline{x}) = (nh)^{-1} \sum_{i=1}^{n} k\left(\frac{\overline{x} - x_i}{h}\right), \tag{9}$$

and

$$\frac{\widehat{\sigma}^{2+}(\overline{x}) + \widehat{\sigma}^{2-}(\overline{x})}{\widehat{f}_o(\overline{x})} e_1' \Gamma^{-1} \Delta \Gamma^{-1} e_1 \tag{10}$$

is the estimator for the traditional asymptotic variance matrix. The matrix $\Gamma^{-1}\Delta\Gamma^{-1}$ can be calculated directly because it is a deterministic function of the kernel and local polynomial order.

An additional drawback of the variance estimator in formula (10) is the need to estimate $f_o(\overline{x})$, which is sidestepped if formula (6) is used. To obtain $\widehat{f}_o(\overline{x})$, we need to choose a kernel and a bandwidth for the density estimator, increasing the number of tuning parameters to be chosen.[9]

5 Simulations

This section presents simulation evidence displaying the empirical coverage of a standard t-statistic used to perform inference about the treatment effect of interest. The objective of the simulations is to determine the extent to which the heteroskedasticity robust variance estimator discussed in Section 4 improves on the small-h, homoskedastic variance estimator when the bias introduced by local misspecification is small. As shown previously, it is expected that both approaches yield similar test performance when the bandwidths are small, and differences in empirical coverage should be of greater importance when local heteroskedasticity is present around the cutoff.

To evaluate the relative performance of tests based on the fixed-h and small-h asymptotic variance approximations and their respective estimators, in this section I present evidence from two data generating processes for which the local polynomial estimator has negligible or mild asymptotic bias (in the sense that the bias does not overwhelm inference completely), in line with the condition imposed on Theorem 1. Obviously, if the bias in the local linear estimator is important, the inference on both approaches would suffer equally, since they use the same estimator.

Evidence from the simulations presented below indicates that, both in the theoretical (unfeasible) and feasible cases, inference using the heteroskedasticity robust standard errors performs well, especially for larger bandwidths and when local heteroskedasticity is present.

The simulations are based on a sharp RD design and consist of 2,000 replications with sample size n equal to 1,000 observations; the effective sample size included depends on the bandwidth used.[10] For the DGPs used the running variable is drawn from $X \sim 2Beta(2,4) - 1$, and $\sigma(x) = 0.1295$ for the homoskedastic case, $\sigma(x) = 0.1295 + (x)^2$ for the first (mild) heteroskedastic case, and $\sigma(x) = 0.1295 + (5x)^2$ for the second (acute) case. DGP 1 is linear in X and will serve as a benchmark of the refinements offered by the robust asymptotic approximation developed in Section 3. DGP 2 follows Calonico et al. (2014) and is based on an empirical RD problem, corresponding to the regression function fitted to Lee's (2008) data both above and below the cutoff. Since this DGP introduces bias for some bandwidths, it will allow us to compare the respective coverage obtained by both approaches in the presence of some bias.

The bandwidths used range from 0.1 to 1. While bandwidth selection algorithms have been proposed in the literature, in practice implementation requires the researcher to use a particular set of bandwidths. For

[9] Imbens and Lemieux (2008) propose a plug-in estimator for $\frac{\sigma^{2+}(\overline{x})+\sigma^{2-}(\overline{x})}{f_o(\overline{x})}$ and obtain their estimate for the asymptotic variance of the local linear estimator by scaling it by $e_1'\Gamma^{-1}\Delta\Gamma^{-1}e_1$. This estimator suffers from the same drawbacks as the one proposed by Porter (2003).

[10] When there are no observations inside the bandwidth for all choices of h considered the sample is dropped.

example, for discrete running variables, a discretely positive bandwidth is necessary. Even with a continuous running variable, sample sizes often are small enough that concerns about precision impel researchers to use a relatively large bandwidth. As pointed out by Calonico et al. (2014), bandwidth selectors in the literature "typically lead to bandwidth choices that are too large for the usual distributional assumptions to be valid." Furthermore, as pointed out in several recent studies, these optimal bandwidth selectors could change significantly from setting to setting in a way not captured by the bandwidth selection algorithm (see Calonico, Cattaneo and Farrell (2017) or Bartalotti and Brummet (2017) for cluster cases). Hence, most researchers use data-driven bandwidth selectors to benchmark their results while reporting results for a set of chosen *ad hoc* bandwidths, which is adequate. The simulations presented here cover a wide range of fixed bandwidth values including those suggested by the selectors, being representative of the practice in the applied literature, highlighting the robustness of the fixed bandwidth approximation regardless of bandwidth choice.

In Appendix 8.1 I present two additional DGPs which have high bias and compare two potential bias mitigation approaches, one based on small-h bias approximation in equation 4 (Calonico et al., 2014) and a novel approach inspired by the fixed-h bias approximation in equation 3, which is based on a modified version of the bias correction bootstrap described in Bartalotti, Calhoun and He (2017).

The empirical coverages presented are the fraction of rejections in the 2,000 repetitions for a two-sided test of nominal size 5%. More specifically, the DGPs are:

- DGP 1: $y_i = 0.48 + 1.27x_i + u_i \qquad \text{if } x < 0,$

  $\qquad\qquad 0.52 + 0.84x_i + u_i \qquad \text{if } x \geq 0.$

- DGP 2: $y_i = 0.48 + 1.27x_i + 7.18x_i^2 + 20.21x_i^3 + 21.54x_i^4 + 7.33x_i^5 + u_i \qquad \text{if } x < 0,$

  $\qquad\qquad 0.52 + 0.84x_i - 3.00x_i^2 + 7.99x_i^3 - 9.01x_i^4 + 3.56x_i^5 + u_i \qquad \text{if } x \geq 0.$

  The simulations use the local linear estimator ($p = 1$), since it is the preferred choice in applied work.[11] The next subsection compares the test coverages obtained by the theoretical fixed-h and small-h asymptotic distributions derived in Section 3. Subsection 5.2 compares the empirical coverages obtained with (feasible) estimated standard errors.

5.1 Simulations for Infeasible Inference

This subsection documents the differences in test coverages based on the theoretical fixed-h and small-h asymptotic variance formulas. The tests are infeasible since they depend on knowledge of $f_o(x)$ and $\sigma^2(x)$ around the cutoff, and are intended to highlight the improvements achieved by the fixed-h asymptotic variance approximation in describing the behavior of the estimators of interest.

[11] Even though the results apply to any choice of kernel, the rectangular kernel is the main focus so that the estimation procedure simplifies to the application of OLS on the data just above and below the cutoff, emphasizing the validity of standard heteroskedastic-robust standard errors.

These comparisons illustrate how conventional small-h approximations, while valid for small bandwidths, becomes unreliable as we move away from the cutoff, even in the absence of bias. That finding is natural, since the small-h asymptotic approximation's derivation is based on the boundary variance and density and should not be expected to adequately describe the estimator's behavior away from the threshold.

The empirical coverages for DGP 1 in the cases studied are presented on Panel A on Table 1. For smaller bandwidths, both asymptotic variances generate similar empirical coverages as expected, but there is a significant decrease in the small-h coverage as the bandwidth increases, while fixed-h inference increasingly outperforms the standard approximation.

DGP 2, based on data in Lee (2008), presents qualitatively similar results, as can be seen on Panel B in Table 1. Note that the coverage varies severely depending on the bandwidth used due to the bias implied by each choice. Nevertheless, the bias is small enough in this case not to overwhelm the tests completely, and it is clear that tests based on fixed-h asymptotic variance produce more adequate coverage even in the presence of bias. Furthermore, the improvements increase with bandwidth sizes, as predicted.

As described in Section 3, the refinements obtained by the fixed-h approach are due to considering the behavior of $f_o(x)$ and $\sigma^2(x)$ inside the bandwidth. Hence, in the presence of heteroskedasticity those should be even more important.

Table 1: Infeasible Inference - Empirical coverage of t-test (%) - nominal test size:5%

Panel A: DGP 1 (Linear)						
	Homoskedastic		Heteroskedastic 1		Heteroskedastic 2	
Bandwidth	Fixed-h	Small-h	Fixed-h	Small-h	Fixed-h	Small-h
0.1	94.4	94.2	94.3	94.0	94.7	84.9
0.2	94.6	94.5	94.5	93.4	95.0	50.4
0.3	95.2	94.8	95.3	92.0	94.3	26.0
0.4	95.5	94.8	95.0	89.2	95.1	15.4
0.5	95.5	94.3	95.3	84.2	94.5	10.3
0.6	95.3	93.3	95.3	77.8	94.5	7.4
0.7	94.6	92.2	95.2	70.0	94.0	5.0
0.8	94.9	91.1	94.8	62.2	95.2	4.3
0.9	95.0	89.7	94.3	57.3	94.0	4.1
1.0	95.2	87.9	94.0	52.5	94.0	3.5
Panel B: DGP 2 (Lee (2008))						
	Homoskedastic		Heteroskedastic 1		Heteroskedastic 2	
Bandwidth	Fixed-h	Small-h	Fixed-h	Small-h	Fixed-h	Small-h
0.1	93.8	93.8	93.8	93.4	93.8	83.5
0.2	83.2	82.8	84.0	81.4	92.1	49.5
0.3	62.3	60.9	67.7	59.7	94.2	27.1
0.4	53.2	51.2	64.9	49.9	94.8	15.8
0.5	66.6	63.0	79.7	59.5	94.9	10.7
0.6	84.1	80.8	89.5	69.3	94.8	7.4
0.7	91.8	88.3	94.1	68.2	95.8	5.0
0.8	87.8	81.7	93.3	59.7	95.0	3.9
0.9	56.1	42.4	86.2	43.5	94.3	2.9
1.0	22.7	13.4	77.8	27.2	95.8	2.8

The coverages for the (mild) heteroskedastic case, and the second (acute) heteroskedastic cases are presented in the last four columns of Table 1. The (infeasible) tests based on fixed-h asymptotics behave well in all cases, highlighting the robustness of the approach. In the first case, the small-h asymptotic approximation presents a more pronounced pattern of decreasing coverage as the bandwidths increase, compared to the homoskedastic case; as expected. In contrast, for the acute heteroskedasticity case, the small-h based

test has a steep decline in coverage as the bandwidth increases, since it is not able to properly capture the effect of the heteroskedasticity for larger bandwidths.

The heteroskedasticity described in this simulation is a "worst case scenario" for small-h asymptotics since the conditional variance of the error at the cutoff, $\sigma^2(\overline{x})$, is at the extreme of the range of values assumed by $\sigma^2(x)$ in any given bandwidth. As can be seen from formula (1), the small-h and fixed-h asymptotic variances will be more similar if $\sigma^2(\overline{x})$ is close to the "weighted average" of $\sigma^2(x)$ inside the bandwidth.

Some points are worth emphasizing. First, the general pattern is that the empirical coverages obtained using fixed-h results from theorem 1 outperform those from the small-h approximations, especially for larger bandwidths. Second, for smaller bandwidths, small-h asymptotics provide similar coverages to the fixed-h approach, making it clear that the core difference is due to the suitability of the restrictions imposed on $f_o(x)$ and $\sigma^2(x)$ as the bandwidth increases (corollary 2). Naturally, those restrictions tend to be less realistic for larger bandwidths. Third, in the presence of heteroskedasticity, the small-h approximation can have poor performance, while the fixed-h asymptotic variance still provides a reliable approximation for the estimator's behavior.

5.2 Simulations for Feasible Inference

This subsection presents simulations for the empirical coverage of the tests using two different estimated standard errors. The first is based on the fixed-h asymptotic variance and is given by formula (6), which is akin to treating the estimates as locally parametric as discussed above. The second is analogous to the ones proposed by Porter (2003) and Imbens and Lemieux (2008), and described by formula (10). Even though it is nowadays common practice to implement variance estimators that are similar to the usual heteroskedasticity robust standard errors in RD applications these exercises are helpful. They provide evidence of the shortcomings in relying on standard error formulas based on small-h asymptotics.

For locally homoskedastic errors, the heteroskedasticity robust standard errors' estimator incorporates the gains of improved inference as described in the theory and infeasible simulations. These results are seen in the first two columns of Table 2. Perhaps surprisingly, tests obtained using small-h standard error estimators behave very similarly to those of fixed-h even at relatively larger bandwidths, for which one would expect a significantly smaller coverage considering the evidence in subsection 5.1. Essentially, the small-h variance estimators benefit from the fact that, by using data on x_i and $\widehat{\varepsilon}_i$ in practice, the estimator for the standard errors partially captures the behavior of $f_o(x)$ and $\sigma^2(x)$ in the range around the cutoff - even though the theoretical small-h asymptotic approximation ignores it.

Table 2: Feasible Inference - Empirical coverage of t-test (%) - nominal test size:5%

Panel A: DGP 1 (Linear)						
	Homoskedastic		Heteroskedastic 1		Heteroskedastic 2	
Bandwidth	Fixed-h	Small-h	Fixed-h	Small-h	Fixed-h	Small-h
0.1	94.0	93.8	94.2	94.0	94.3	98.0
0.2	94.3	94.3	94.3	95.5	94.2	99.4
0.3	94.8	94.2	94.9	97.0	94.2	99.5
0.4	95.1	94.8	95.0	97.8	94.6	99.7
0.5	95.8	94.3	95.1	98.8	94.9	99.7
0.6	94.8	93.8	94.6	98.8	94.5	99.2
0.7	94.5	93.0	94.8	99.0	94.2	99.4
0.8	94.4	92.3	94.2	99.2	94.3	99.5
0.9	94.5	91.6	93.9	99.4	93.5	99.4
1.0	94.3	91.8	93.4	99.2	93.3	99.4
Panel B: DGP 2 (Lee (2008))						
	Homoskedastic		Heteroskedastic 1		Heteroskedastic 2	
Bandwidth	Fixed-h	Small-h	Fixed-h	Small-h	Fixed-h	Small-h
0.1	93.2	93.1	93.5	93.5	93.7	98.0
0.2	82.8	82.5	84.0	85.7	92.0	99.0
0.3	61.7	60.2	67.6	74.5	94.0	99.1
0.4	52.8	50.8	63.9	76.0	94.2	99.5
0.5	66.5	63.7	78.8	90.3	94.7	99.7
0.6	84.6	81.5	89.0	96.5	94.0	99.5
0.7	92.0	89.5	93.5	98.8	94.6	99.6
0.8	87.9	83.9	92.7	98.2	93.9	99.1
0.9	55.9	50.0	85.5	97.0	93.5	99.5
1.0	23.2	20.1	76.5	94.0	94.5	99.7

To see this point, note that the researcher is not able to exactly estimate $f_o(\overline{x})$ and $\sigma^2(\overline{x})$ from a given dataset, as would have been suggested by the theoretical small-h's asymptotic variance formula. By being "forced" to estimate the variance and the density within the bandwidth, the small-h variance estimator is able to *partially* capture the local behavior of those terms.

In the heteroskedastic cases, the fixed-h variance estimator produces tests with coverage very close to the

test's nominal size for DGP 1, while the coverage for small-h rapidly increases towards 1 as the bandwidth increases. Note that the empirical coverage shows under-rejection in this case since, first, $\sigma^2(x)$ increases away from the cutoff and, second, $\widehat{\sigma}^{2+}$ and $\widehat{\sigma}^{2-}$ will significantly overestimate the true weighted average of the conditional variance within the bandwidth.

Hence, there is evidence that heteroskedasticity can be accurately captured by tests based on fixed-h approximations; small-h based standard errors can produce tests with significant size distortion. Furthermore, the distortion on coverage encountered on this particular DGP for heteroskedasticity is not even in the same direction that the unfeasible (theoretical) tests would predict!

Therefore, it seems the heteroskedasticity robust standard errors are a "safer choice" for practitioners since it is based on an asymptotic approximation that is "robust to bandwidth choice" and its computation is very easy once a kernel and bandwidth are chosen. Furthermore, the robust variance estimator has the advantage of not requiring the estimation of $f_o(\overline{x})$. That estimation would entail the choice of a (potentially different) kernel and bandwidth for $\widehat{f}_o(\overline{x})$. These two tuning parameters might significantly alter the empirical size of the tests and depend on the discretion of the researcher.

To the empirical researcher, a useful conclusion can be drawn from these simulations. By performing inference using fixed-h based standard errors, which is akin to treating the estimates as locally parametric and simplifies to the standard heteroskedastic robust standard errors for rectangular kernels, one can feel confident about the standard errors for any bandwidth used.

The researcher can then focus his attention on choosing a bandwidth to deal with the bias at hand.[12] As pointed out in Section 3 and the simulations above, this issue similarly affects the empirical test coverage ragardless of the standard errors used. However, the fixed-h approximation has the benefit of clarifying that, even under the validity of the RD design, local misspecification can be an important factor. Taking advantage of RD to estimate a treatment effect of interest is the exercise of estimating the conditional expectation of the outcome variable *inside* the bandwidth. Naturally, for larger bandwidths, one would expect that the likelihood of misspecification increases, requiring higher-order bias corrections. I discuss potential approaches to address the bias on Section 3.1. Appendix 15 presents simulation evidence that higher-orders bias mitigation procedures motivated by the fixed-h asymptotics provide improved coverage performance relative to first-order bias correction based on the small-h asymptotics, especially for larger bandwidths.

[12]For a detailed discussion on bias correction in RD designs and the inference adjustments needed when performing such correction, see Calonico et al. (2014).

6 Empirical Example

To illustrate the potential differences in the small-h and fixed-h approximations discussed above, this section uses data from Lee's (2008) study of the electoral advantage of incumbency in the United States.[13] Lee (2008) argues that the U.S. Congressional electoral system has a "built-in" RD design. That is, being the incumbent party in a congressional district is a deterministic function of the candidate-party's vote share in the district during the last electoral cycle. This feature can be described in the following model:

$$v_{i2} = \alpha w_{i1} + \beta v_{i1} + \gamma d_{i2} + e_{i2}$$
$$d_{i2} = 1\left[v_{i1} \geq \frac{1}{2}\right],$$

where v_{it} is the democratic candidate's vote share in congressional district i in election year t, w_{it} is a vector of characteristics or agents' choices (potentially unobserved) as of election day on period t, and d_{it} indicates if the Democratic party is the incumbent in district i at election period t. We also assume that $f_{i1}(v \mid w)$, the density of v_{i1} conditional on w_{i1}, is continuous in v. The main issue in the analysis, as discussed in detail by Lee (2008), is that w_{it} is potentially unobserved and would likely be correlated with being incumbent in a certain district. For example, w_{i1} would include party resources, demographic characteristics, and political leaning of districts, all of which could affect both the vote share in periods 1 and 2, thus biasing the estimates for the causal effect of incumbency.

The thought experiment to find the causal effect of incumbency would be to randomly allocate incumbency in districts to Democrats and Republicans while keeping all other characteristics constant. This clearly cannot be done, but by looking at closely contested elections, we can consider that the incumbents in those districts, whether Democrats or Republicans, were decided randomly, given the inherent uncertainty regarding the outcome of such events.[14] This can provide reasonable estimates of the causal effect of incumbency in closely contested elections.

The data includes U.S. congressional election returns from 1946 to 1998, excluding the years ending in '0' and '2' due to the decennial redistricting which characterizes the U.S. congressional electoral system. The running variable is defined as the difference in vote share between the Democratic candidate and the strongest opponent. Hence, the Democrat wins the election when this variable crosses the 0 threshold - i.e., there is a positive difference in vote share, indicating that the Democrat has received more votes.

Table 1 shows the estimated advantage of incumbency and the estimated standard errors given by formulas based on a fixed bandwidth approximation (6), which in the rectangular kernel case is the usual heteroskedasticity robust s.e., and a small-h approximation (10), respectively.

[13] The dataset used in this section have been downloaded from the "Mostly Harmless Econometrics Data Archive" website. (http://economics.mit.edu/faculty/angrist/data1/mhe)

[14] For a discussion on the assumptions necessary for the validity of RD design in this example, see Lee (2008).

Table 3: Incumbency Effects and Estimated Standard Errors - Lee (2008)

Dependent Variable: Democrat Vote Share - Election $t+1$

	All	$\lvert Margin\rvert \leq 0.5$	$\lvert Margin\rvert \leq 0.05$
Panel A: Nadaraya-Watson Estimator ($p=0$)			
Estimated Effect	0.351	0.257	0.096
(Fixed-h Standard Errors)	(0.0041)	(0.0038)	(0.0090)
[Small-h Standard Errors]	[0.0041]	[0.0038]	[0.0090]
Difference (%)	1.6%	0.5%	0.1%
Panel B: Local Linear Estimator ($p=1$)			
Estimated Effect	0.118	0.090	0.048
(Fixed-h Standard Errors)	(0.0056)	(0.0062)	(0.0159)
[Small-h Standard Errors]	[0.0068]	[0.0071]	[0.0180]
Difference (%)	21.4%	14.5%	13.2%
Panel C: Local Polynomial Estimator ($p=4$)			
Estimated Effect	0.077	0.066	0.105
(Fixed-h Standard Errors)	(0.0113)	(0.0144)	(0.0312)
[Small-h Standard Errors]	[0.0167]	[0.0179]	[0.0447]
Difference (%)	46.5%	24.3%	42.4%
Observations	6558	4900	610

Panel A presents estimates of the Nadaraya–Watson estimator ($p=0$) for different bandwidths. Column 1 uses all the data available, column 2 looks only at elections for which the margin of victory in period $t-1$ was within 50% of the total votes, and column 3 uses only elections with margins lower than 5%. Panel B presents similar estimates and standard errors obtained by a local linear estimator ($p=1$), which is the preferred specification in several RD applications in the literature and is expected to significantly reduce bias in the estimates of the incumbency advantage effect.

Panel C presents the results Lee (2008) called the "parametric fit" in the first column, which uses a polynomial of order 4 to fit the whole data. The other two columns use smaller bandwidths to emphasize how the order of the polynomial chosen to fit the data can significantly impact estimates and standard errors, especially in small samples, by over-fitting the data.

The point estimates are exactly the same as presented by Lee (2008) for panel A and the first column in panel C. The remaining estimates are new. The results indicate a significant incumbent advantage in U.S. congressional races, even when comparing districts that had close elections in the previous electoral cycle for which the determination of incumbent status can be considered "as good as randomized."[15] See Table 2 in the online appendix for similar results that include the pre-determined variables that Lee (2008) uses and the

[15]See Caughey and Sekhon (2011) for a discussion on that claim.

comparisons for fixed-h and small-h standard errors. The observed differences in pre-determined variables between incumbents and challengers vanish as we compare districts that previously had competitive races, lending credibility to the RD as an identification strategy for the incumbency effect. More relevant to this paper are the differences between the competing standard error estimates.

The estimated standard errors, shown in Table 1, differ significantly, with the fixed-h standard errors being smaller in most of the cases—as one would expect, given the simulations in Section 5. Also as expected, using smaller bandwidths comes at a large cost in terms of precision of the estimates due to the smaller amount of data available, negatively affecting both standard error estimators. The two last columns in all three panels show an increase of the estimated standard errors when the data is restricted to districts that had victory margins smaller than 5% in the previous election.

Interestingly, it is usual for the relative gap between the two standard errors estimates to decline as the bandwidth shrinks, as predicted in Section 3. For panels A and B and for the first two columns of panel C, this pattern is confirmed as we compare the percent difference between the standard errors within panels.[16] For the third entry in panel C, note that both standard errors become larger than those for the wider bandwidth in column 2, but the relative gap increases. That can be due to the fact that the fixed-h standard error estimate requires the calculation of $4(2p+1)$ terms (see equation 6). Hence, it is more susceptible to the combination of large polynomial orders and small sample sizes induced by the smaller bandwidth. Nevertheless, it still provides tighter confidence intervals than the small-h estimated standard error.

Finally, note that the differences between standard errors increase as the order of the polynomial used to fit the data increases for the same bandwidth. This is due to the fact that the small-h standard error estimator is based on a fixed scaling term for a given kernel and polynomial order choice, $e_1'\Gamma^{-1}\Delta\Gamma^{-1}e_1$. Intuitively, as the polynomial order increases, the greater the distortion is likely to be between using this approximation and the more refined formula implied by the fixed-h approach.

7 Conclusion

This paper focuses on presenting an alternative asymptotic approximation for the standard RD treatment effects estimator using a fixed bandwidth. This approximation provides expressions for the estimator's asymptotic variance that incorporate the bandwidth used by the researcher, and leads naturally to the standard error formulas used in the applied literature, i.e., White-Huber-Eicker robust standard errors when a rectangular kernel is implemented.

The proposed approximation is valid under an asymptotic nesting where the bandwidth parameter, h, is fixed, as well as under conventional RD asymptotics where $h \to 0$, while the traditional asymptotic approximations based on a shrinking bandwidth suggested, for example, in Porter (2003) and Imbens and Lemieux (2008) are not adequate for relatively larger bandwidths used in practice, especially in the presence

[16] A similar pattern emerges from table 2 in the online appendix.

of heteroskedasticity. Furthermore, the approximations provided in this study relax the assumption of continuity of the conditional variance of the outcome at the cutoff, providing a theoretical framework for that case. The common empirical strategy of using robust standard errors in RD studies is appropriate when the misspecification is small and heteroskedasticity is a concern.

Simulations document the theoretical refinements provided by relying on asymptotic variances based on fixed bandwidth asymptotics, and illustrate that tests using heteroskedasticity robust standard errors produces significantly more reliable test coverage than its counterparts using standard errors based on asymptotic approximations where $h \to 0$.

Furthermore, an alternative, fixed-h, approximation for the estimator's potential asymptotic bias is obtained and provides additional intuition on the robustness-precision trade-off facing the researcher in RD designs. I investigate two alternatives to address the bias in practice. The first is the widely used analytical approach based on the traditional asymptotics suggested by Calonico et al. (2014). The second follows more naturally from the fixed bandwidth asymptotics developed here and is based on a iterative bootstrap approach that provides higher order bias mitigation by building upon the wild bootstrap algorithm in Bartalotti, Calhoun and He (2017).

References

[1] Andrews, D. 1991. Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models. *Econometrica*, 59(2), 307-345.

[2] Angrist, J., Pischke, J. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press: Princeton, NJ.

[3] Barreca, A., Lindo, J., and Waddell, G. 2016. Heaping-Induced Bias in Regression Discontinuity Designs. *Economic Inquiry,* 54(1): 268-293.

[4] Bartalotti, O. Brummet, Q. 2017. Regression Discontinuity Designs with Clustered Data. In: *Regression Discontinuity Designs: Theory and Applications. Advances in Econometrics*, Eds. M. Cattaneo and J. Escanciano, Emerald Publishing, V. 38: 383-420.

[5] Bartalotti, O., Brummet, Q., Dieterle, S. 2017. A Correction for Regression Discontinuity Designs with Group-Specific Mismeasurement of the Running Variable *Working Paper*, Department of Economics, Iowa State University.

[6] Bartalotti, O. Calhoun, G., He, Y. 2017. Bootstrap Confidence Intervals for Sharp Regression Discontinuity Designs. In: *Regression Discontinuity Designs: Theory and Applications. Advances in Econometrics*, Eds. M. Cattaneo and J. Escanciano, Emerald Publishing, V. 38: 421-453.

[7] Bester, C. A., Conley, T. G., Hansen, C. B., and Vogelsang, T. J. 2016. Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators. *Econometric Theory*, 32, 154-186.

[8] Calonico, S., Cattaneo, M., and Farrell, M. Forthcoming. On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association*.

[9] Calonico, S., Cattaneo, M., Farrell, M., and Titiunik, R. 2017. rdrobust: Software for regression-discontinuity designs. *Stata Journal* 17(2): 372-404.

[10] Calonico, S., Cattaneo, M., and Titiunik, R. 2014. Robust Nonparametric Confidence Intervals for Regression Discontinuity Designs. *Econometrica* 82(6): 2295-2326.

[11] Cattaneo, M., Frandsen, B., and Titiunik, R. 2015. Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate. *Journal of Causal Inference*, *3(1):1-24*.

[12] Caughey, D., and Sekhon, J. 2011. Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942–2008. *Political Analysis*, 19, 385-408.

[13] Chen, X., Liao, Z. and Sun, Y. 2014 Sieve inference on possibly misspecified semi-nonparametric time series models. *Journal of Econometrics*, 178(3): Pages 639-658.

[14] Fan, J., and Gijbels, I. 1996. *Local Polynomial Modelling and its Applications*, Chapman & Hall: New York, NY.

[15] Dong, Y. 2015. Regression Discontinuity Applications with Rounding Errors in the Running Variable. *Journal of Applied Econometrics* 30(3): 422-446.

[16] Fan, Y. 1998. Goodness-of-Fit Tests Based on Kernel Density Estimators with Fixed Smoothing Parameters. *Econometric Theory*, 14, 604–621.

[17] Hahn, J., Todd, P., and van der Klaauw, W. 1999. Evaluating the Effect of an Antidiscrimination Law Using a Regression-Discontinuity Design. NBER Working Paper Series, n. 7131.

[18] Hahn, J., Todd, P., and van der Klaauw, W. 2001. Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69, 201–209.

[19] Hashimzade, N. and Vogelsang, T. 2008. Fixed-b Asymptotic Approximation of the Sampling Behaviour of Nonparametric Spectral Density Estimators. *Journal of Time Series Analysis*, 29 (1), 142–162.

[20] Imbens, G. and Lemieux, T. 2008. Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142 (2), 615–635.

[21] Imbens, G. and Kalyanaraman, K. 2012. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies*, 79 (3), 933–959.

[22] Kim, M., Sun, Y., and Yang, J. 2017. A Fixed-bandwidth View of the Pre-asymptotic Inference for Kernel Smoothing with Time Series Data. *Journal of Econometrics,* 197(2), 298-322.

[23] Lee, D. 2008. Randomized Experiments from Non-random Selection in U.S. House Elections. *Journal of Econometrics*, 142 (2), 675–697.

[24] Lee, D. and Card, D. 2008. Regression Discontinuity Inference with Specification Error. *Journal of Econometrics*, 142 (2), 655–674.

[25] Lee, D., and Lemieux, T. 2009. Regression Discontinuity Designs in Economics. NBER Working Paper Series, n. 14723.

[26] McCrary, J. 2008. Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*, 142 (2), 698–714.

[27] Neave, H. 1970. An Improved Formula for the Asymptotic Variance of Spectrum Estimates. *The Annals of Mathematical Statistics*, 41, 70–77.

[28] Pagan, A., and Ullah, A. 1999. *Nonparametric Econometrics*, Cambridge University Press.

[29] Porter, J. 2003. Estimation in the Regression Discontinuity Model. Unpublished Working Paper. Department of Economics, University of Wisconsin. *Manuscript date*: May 7, 2003.

[30] Ramanathan, R. 1993. *Statistical Methods in Econometrics*, Academic Press.

[31] Stock, J., and Yogo, M. 2005. Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments. In: Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg (1st ed.), eds. D. Andrews and J. Stock, Cambridge University Press, pp. 109–120.

[32] Sun, Y. 2014 Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. *Journal of Econometrics*, 178(3):659-677.

[33] White, H. 1982. Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50, 1–24.

[34] White, H. 1996. *Estimation Inference and Specification Analysis*, Cambridge University Press.

8 Appendix

8.1 Simulations for Bias Control

Below we present simulation evidence of the small sample performance for the first order bias correction proposed by Calonico et al. (2014) as implemented in the *rdrobust* package in R and the modified Bartalotti, Calhoun and He (2017) wild bootstrap procedure which easily allows for higher order bias mitigation, as discussed on Section 3. To emphasize the bias mitigation aspect the DGPs in this exercise are chosen to exhibit moderate to severe bias. Specifically, we implement the DGPs described in Calonico el al. (2014) and used by Bartalotti, Calhoun and He (2017).

The first DGP is the same as described in Section 5 as DGP 2

- DGP Bias 1: $y_i = 0.48 + 1.27x_i + 7.18x_i^2 + 20.21x_i^3 + 21.54x_i^4 + 7.33x_i^5 + u_i \qquad \text{if } x < 0,$

$$0.52 + 0.84x_i - 3.00x_i^2 + 7.99x_i^3 - 9.01x_i^4 + 3.56x_i^5 + u_i \qquad \text{if } x \geq 0.$$

The population ATE for this DGP is $0.04 (= 0.52 - 0.48)$.

The second DGP is based on Ludwig and Miller's (2007) analysis of the Head Start program. In this case, eligibility to grant-writing assistance was determined at the county level using historical poverty rate, with a sharp threshold that determines the provision of services.

- DGP Bias 2: $y_i = 3.71 + 2.30x_i + 3.28x_i^2 + 1.45x_i^3 + 0.23x_i^4 + 0.03x_i^5 + u_i \qquad \text{if } x < 0,$

$$0.26 + 18.49x_i - 54.81x_i^2 + 74.30x_i^3 - 45.02x_i^4 + 9.83x_i^5 + u_i \qquad \text{if } x \geq 0.$$

and the population ATE is $-3.45 (= 0.26 - 3.71)$.

Finally, for the third DGP, we use CCT's modification of the first DGP, given by

- DGP Bias 3: $y_i = 0.48 + 1.27x_i + 3.59x_i^2 + 14.147x_i^3 + 23.694x_i^4 + x_i^5 + u_i \qquad \text{if } x < 0,$

$$0.52 + 0.84x_i - 0.30x_i^2 + 2.397x_i^3 - 0.901x_i^4 + x_i^5 + u_i \qquad \text{if } x \geq 0.$$

and the population ATE is again 0.04. CCT introduce this DGP because it has high curvature and local linear models are likely to exhibit high bias.

As the focus of this set of simulations is comparing the bias mitigation achieved by the first order bias correction (Calonico et al. 2014) inspired by the usual shrinking bandwidth asymptotics, and the higher order bias correction using the modified bootstrap procedure prompted by the fixed bandwidth approach, all reported results are based on std. errors compatible with the fixed bandwidth asymptotic nesting described on Section 3.

I follow a similar approach to the simulations as described in Bartalotti, Calhoun and He (2017). I simulate 5000 samples from each of the three DGPs and calculate nominal 95% two-sided confidence intervals. We use 999 bootstrap replications to calculate the asymptotic distribution of the bias corrected estimator, and each of those replications uses an additional 500 replications to estimate the bias. The bandwidths are set equal to h and are fixed within each Monte Carlo replication, in line with the fixed-h thought experiment.

As it can be seem in Table 4 below, the bias mitigation obtained by the first order bias correction becomes less capable of capturing the bias distortion as the bandwidth moves away from zero. This is in line with the discussion in Section 3.1, since for larger bandwidths the bias being captured by the expansion in equation 4 becomes a less reliable approximation of the bias in the local support $\dot{\aleph}$. It is important to note that in the exercise described below I used the correct polynomial order ($p = 5$) for the bias estimation in the modified bootstrap, so it reflects a best case scenario in terms of approximating the bias in the local support of the running variable.

Table 4: Bias Corrected Inference - Empirical coverage of t-test (%) - nominal test size:5%

	DGP Bias 1				DGP Bias 2				DGP Bias 3			
	Homosk.		Heterosk.		Homosk.		Heterosk.		Homosk.		Heterosk.	
h	Boot	CCT	Boot	CCT	Boot	CCT	Boot	CCT	Boot	CCT	Boot	CCT
0.1	95.7	92.1	96.1	92.5	95.7	94.0	93.1	94.2	95.7	92.0	95.9	92.8
0.2	95.2	93.5	94.4	92.2	95.2	93.6	91.8	93.5	95.2	93.7	94.6	92.6
0.3	93.8	92.5	94.8	93.0	93.8	81.2	92.1	79.7	93.8	94.3	95.0	94.1
0.4	95.3	87.2	94.4	90.0	95.3	36.2	92.2	25.1	95.3	92.8	94.7	94.0
0.5	93.7	79.2	94.5	81.7	93.7	2.1	93.2	1.2	93.7	93.2	95.2	94.2
0.6	95.2	72.0	95.5	77.3	95.2	0.1	93.2	0.0	95.2	87.3	96.9	93.2
0.7	95.0	73.1	94.9	80.2	95.0	0.0	93.8	0.0	95.0	66.5	96.4	92.8
0.8	94.8	87.2	95.6	88.3	94.8	0.0	94.2	0.0	94.8	27.9	96.7	92.7
0.9	94.3	93.3	95.8	93.5	94.3	0.0	95.0	0.0	94.3	9.6	96.8	92.5
1.0	95.1	90.9	95.8	93.1	95.1	0.0	94.6	0.0	95.1	4.3	96.4	92.0

These simulations reinforce the main point of this paper that the fixed-h asymptotic framework provides a robust and reliable alternative to the usual asymptotic approximations that rely on shrinking the bandwidth when dealing with practical implementation of RD designs. Furthermore, these approximations provide insight on the properties of the estimator that can lead to improved empirical practices, as exemplified by the modified wild bootstrap bias mitigation suggested in this section.

8.2 Additional Results for Empirical Example: Lee (2008)

Table 5: Electoral outcomes and pre-determined election characteristics: 1948-1996 - Lee (2008)

Variable	Nadaraya-Watson Estimator ($p=0$)			Local Linear Estimator ($p=1$)			Local Polynomial Estimator ($p=4$)		
	All	$Margin \leq 0.5$	$Margin \leq 0.05$	All	$Margin \leq 0.5$	$Margin \leq 0.05$	All	$Margin \leq 0.5$	$Margin \leq 0.05$
Democrat vote share $t+1$	0.351	0.257	0.096	0.118	0.090	0.048	0.077	0.066	0.105
(fixed-h std. error)	(0.0041)	(0.0038)	(0.0090)	(0.0056)	(0.0062)	(0.0159)	(0.0114)	(0.0144)	(0.0312)
[small-h std. error]	[0.0041]	[0.0038]	[0.0090]	[0.0068]	[0.0071]	[0.0180]	[0.0167]	[0.0179]	[0.0447]
Democrat vote share $t-1$	0.313	0.215	0.027	0.066	0.032	0.010	-0.004	0.002	0.104
(fixed-h std. error)	(0.0042)	(0.0040)	(0.0106)	(0.0059)	(0.0070)	(0.0214)	(0.0134)	(0.0177)	(0.0431)
[small-h std. error]	[0.0043]	[0.0040]	[0.0106]	[0.0070]	[0.0074]	[0.0211]	[0.0173]	[0.0184]	[0.0525]
Democrat win prob. $t-1$	0.780	0.724	0.137	0.562	0.335	0.160	-0.0116	0.004	0.545
(fixed-h std. error)	(0.0077)	(0.0097)	(0.0392)	(0.0159)	(0.0228)	(0.0802)	(0.0436)	(0.0609)	(0.1809)
[small-h std. error]	[0.0076]	[0.0097]	[0.0391]	[0.0146]	[0.0181]	[0.0782]	[0.0342]	[0.0440]	[0.1944]
Democrat Political Exper.	3.551	3.246	0.673	2.459	1.101	0.584	0.035	0.223	1.627
(fixed-h std. error)	(0.0657)	(0.0796)	(0.2070)	(0.1073)	(0.1412)	(0.4202)	(0.2537)	(0.3491)	(0.8309)
[small-h std. error]	[0.0739]	[0.0820]	[0.2102]	[0.1459]	[0.1576]	[0.4203]	[0.3570]	[0.3924]	[1.0448]
Opposition Political Exper.	-2.631	-2.458	-0.162	-1.676	-0.729	-0.417	0.131	0.145	-0.365
(fixed-h std. error)	(0.0563)	(0.0840)	(0.1648)	(0.0897)	(0.1153)	(0.3643)	(0.2072)	(0.2840)	(0.8260)
[small-h std. error]	[0.0492]	[0.0608]	[0.1651]	[0.0966]	[0.1171]	[0.3291]	[0.2363]	[0.2913]	[0.8189]
Democrat Electoral Exper.	3.480	3.200	0.673	2.405	1.083	0.552	0.0149	0.240	1.557
(fixed-h std. error)	(0.0673)	(0.0813)	(0.2118)	(0.1098)	(0.1446)	(0.4309)	(0.2593)	(0.3566)	(0.8755)
[small-h std. error]	[0.0750]	[0.0835]	[0.214]	[0.1485]	[0.1614]	[0.4289]	[0.3641]	[0.4021]	[1.0656]
Opposition Electoral Exper.	-2.607	-2.415	-0.154	-1.623	-0.713	-0.440	0.122	0.120	-0.054
(fixed-h std. error)	(0.0575)	(0.0638)	(0.1692)	(0.0915)	(0.1179)	(0.3753)	(0.2126)	(0.2917)	(0.8539)
[small-h std. error]	[0.0505]	[0.0622]	[0.1693]	[0.0993]	[0.1201]	[0.3372]	[0.2434]	[0.2989]	[0.8394]
Observations	6558	4900	610	6558	4900	610	6558	4900	610

8.3 Technical Details

8.3.1 Additional Notation

In this appendix, all integrals are defined as Stieltjes integrals to accommodate potentially discrete and hybrid running variables (Ramanathan, 1993). In the results presented in the main text, the following notation is employed.

Let, $\Gamma^*_{+(-)} = \begin{bmatrix} \gamma_0^{+(-)} & \cdots & \gamma_p^{+(-)} \\ \vdots & \ddots & \vdots \\ \gamma_p^{+(-)} & \cdots & \gamma_{2p}^{+(-)} \end{bmatrix}$, $\Delta^*_{+(-)} = \begin{bmatrix} \delta_0^{+(-)} & \cdots & \delta_p^{+(-)} \\ \vdots & \ddots & \vdots \\ \delta_p^{+(-)} & \cdots & \delta_{2p}^{+(-)} \end{bmatrix}$, $e_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}'$; $\gamma_j^+ = \int_0^\infty k(u) u^j dF_o(\overline{x}+uh)$; $\gamma_j^- = (-1)^j \int_0^\infty k(u) u^j dF_o(\overline{x}-uh)$; $\delta_j^+ = \int_0^\infty k^2(u) u^j \sigma^2(\overline{x}+uh) dF_o(\overline{x}+uh)$; $\delta_j^- = (-1)^j \int_0^\infty k^2(u) u^j \sigma^2(\overline{x}-uh) dF_o(\overline{x}-uh)$.

Also, for the results when $nh \to \infty$, $h \to 0$, $\Gamma = \begin{bmatrix} \gamma_0 & \cdots & \gamma_p \\ \vdots & \ddots & \vdots \\ \gamma_p & \cdots & \gamma_{2p} \end{bmatrix}$, $\Delta = \begin{bmatrix} \delta_0 & \cdots & \delta_p \\ \vdots & \ddots & \vdots \\ \delta_p & \cdots & \delta_{2p} \end{bmatrix}$, $\gamma_j = \int_0^\infty k(u) u^j du$, $\delta_j = \int_0^\infty k^2(u) u^j du$.

8.3.2 Main Results

The proofs presented below follow closely from the insights and work in Porter (2003).

Proof of Theorem 1. The local polynomial estimator is given by

$$\widehat{\alpha}_p = \widehat{\alpha}_{p+} - \widehat{\alpha}_{p-}$$

note that,

$$\begin{aligned} \widehat{\alpha}_{p+} &= e_1' \left[\frac{1}{nh} \sum_{i=1}^n k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i Z_i' \right]^{-1} \left[\frac{1}{nh} \sum_{i=1}^n k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i y_i \right] \\ &= e_1' D_{n+} \left[\frac{1}{nh} \sum_{i=1}^n k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i y_i \right] \end{aligned}$$

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

$$\widehat{\alpha}_{p-} = e_1' D_{n-} \left[\frac{1}{nh} \sum_{i=1}^n k\left(\frac{x_i - \overline{x}}{h}\right) (1-d_i) Z_i y_i \right]$$

Then,

$$\begin{aligned} \widehat{\alpha}_{p+} &= e_1' D_{n+} \left[\frac{1}{nh} \sum_{i=1}^n k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i \left[m(x_i) + \alpha d_i + \varepsilon_i \right] \right] \\ &= e_1' D_{n+} \left[\frac{1}{nh} \sum_{i=1}^n k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i m(x_i) + \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i \alpha + \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i \varepsilon_i \right] \end{aligned}$$

note that $Z_i = Z_i Z_i' e_1$, $e_1' e_1 = 1$ then

$$\widehat{\alpha}_{p+} - \alpha = e_1' D_{n+} \left[\frac{1}{nh} \sum_{i=1}^n k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i m(x_i) \right] + e_1' D_{n+} \left[\frac{1}{nh} \sum_{i=1}^n k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i \varepsilon_i \right]$$

and similarly for $\widehat{\alpha}_{p-}$.Then

$$\begin{aligned}\sqrt{nh}(\widehat{\alpha}_p - \alpha) &= \sqrt{nh}(\widehat{\alpha}_{p+} - \alpha - \widehat{\alpha}_{p-}) \\ &= e_1' D_{n+} \left\{ \sqrt{nh} \left[\frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i m(x_i) \right] + \left[\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i \varepsilon_i \right] \right\} - \\ &- e_1' D_{n-} \left\{ \begin{array}{c} \sqrt{nh} \left[\frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x_i - \overline{x}}{h}\right) (1 - d_i) Z_i m(x_i) \right] + \\ + \left[\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} k\left(\frac{x_i - \overline{x}}{h}\right) (1 - d_i) Z_i \varepsilon_i \right] \end{array} \right\}\end{aligned}$$

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

$$D_{n+}^{-1} = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i Z_i'$$

and each element of this matrix is given by

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

which has asymptotic variance converging to zero since

$$\begin{aligned}Var\left(\left[D_{n+}^{-1}\right]_{j,l}\right) &= \frac{1}{(nh)^2} Var\left(\sum_{i=1}^{n} k\left(\frac{x_i - \overline{x}}{h}\right) d_i \left(\frac{x_i - \overline{x}}{h}\right)^{j+l-2} \right) \\ &= \frac{1}{nh^2} Var\left(k\left(\frac{x_i - \overline{x}}{h}\right) d_i \left(\frac{x_i - \overline{x}}{h}\right)^{j+l-2} \right) \\ &\leq \frac{1}{nh} \int_{\overline{x}}^{\overline{x}+h} \frac{1}{h} k^2\left(\frac{x - \overline{x}}{h}\right) \left(\frac{x - \overline{x}}{h}\right)^{2(j+l-2)} dF_o(x)\end{aligned}$$

Note that the terms in the integral and the integral itself are $O(1)$ and $\frac{1}{nh} = o(1)$. Hence, $Var\left(\left[D_{n+}^{-1}\right]_{l,j}\right) \to 0$. Now, using the usual transformation $u = \frac{x - \overline{x}}{h}$

$$\begin{aligned}\left[D_{n+}^{-1}\right]_{l,j} &= E\left\{\left[D_{n+}^{-1}\right]_{l,j}\right\} + o_p(1) \\ &= \frac{1}{nh} E\left[\sum_{i=1}^{n} k\left(\frac{x_i - \overline{x}}{h}\right) d_i \left(\frac{x_i - \overline{x}}{h}\right)^{j+l-2} \right] + o_p(1) \\ &= \int_0^{\infty} k(u)\, u^{j+l-2} dF_o(\overline{x} + uh) + o_p(1)\end{aligned}$$

Let, $\gamma_j^+ = \int_0^{\infty} k(u)\, u^j dF_o(\overline{x} + uh)$ and Γ_+^* is the $(p+1) \times (p+1)$ matrix with (j,l) element γ_{j+l-2}^+ for $j, l = 1, ..., p+1$. Then, $D_{n+} \xrightarrow{p} \left(\Gamma_+^*\right)^{-1}$ and $D_{n-} \xrightarrow{p} \left(\Gamma_-^*\right)^{-1}$, where Γ_-^* is the $(p+1) \times (p+1)$ matrix with (j,l) element γ_{j+l-2}^- for $j, l = 1, ..., p+1$, and $\gamma_j^- = (-1)^j \int_0^{\infty} k(u)\, u^j dF_o(\overline{x} - uh)$.

Now we will derive the asymptotic distribution of $\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} k\left(\frac{x_i - \overline{x}}{h}\right) d_i Z_i \varepsilon_i$. Following Porter (2003) I

use the Cramer-Wold device to derive the asymptotic distribution. Let λ be a nonzero, finite vector. Then,

$$
\begin{aligned}
&E\left[\left|\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)d_i\lambda' Z_i\varepsilon_i\right|^{2+\zeta}\right] \\
&=\left(\frac{1}{nh}\right)^{\frac{\zeta}{2}}\frac{1}{h}E\left[\left|k\left(\frac{x-\overline{x}}{h}\right)\right|^{2+\zeta}d\left|\sum_{l=1}^{p}\lambda_l\left(\frac{x-\overline{x}}{h}\right)^l\right|^{2+\zeta}E\left[|\varepsilon|^{2+\zeta}\mid X=x\right]\right] \\
&\leq\left(\frac{1}{nh}\right)^{\frac{\zeta}{2}}\frac{1}{h}\sup_{x\in\aleph}\left\{E\left[|\varepsilon|^{2+\zeta}\mid X=x\right]\right\}E\left[\left|k\left(\frac{x-\overline{x}}{h}\right)\right|^{2+\zeta}d\sum_{l=1}^{p}\left|\lambda_l\left(\frac{x-\overline{x}}{h}\right)^l\right|^{2+\zeta}\right] \\
&=\left(\frac{1}{nh}\right)^{\frac{\zeta}{2}}\frac{1}{h}\sup_{x\in\aleph}\left\{E\left[|\varepsilon|^{2+\zeta}\mid X=x\right]\right\}\int_{\overline{x}}^{\overline{x}+h}\left|k\left(\frac{x-\overline{x}}{h}\right)\right|^{2+\zeta}\sum_{l=1}^{p}\left|\lambda_l\left(\frac{x-\overline{x}}{h}\right)^l\right|^{2+\zeta}dF_o(x) \\
&=\left(\frac{1}{nh}\right)^{\frac{\zeta}{2}}O(1)O(1)=o(1)
\end{aligned}
$$

then, $\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)d_iZ_i\varepsilon_i$ follows a CLT. Note that,

$$
E\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)d_iZ_i\varepsilon_i\right]=0
$$

and

$$
\begin{aligned}
&Var\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)d_iZ_i\varepsilon_i\right] \\
&=\frac{1}{h}E\left[k^2\left(\frac{x_i-\overline{x}}{h}\right)d_iZ_iZ_i'\varepsilon_i^2\right] \\
&=\int_{\overline{x}}^{\overline{x}+h}\frac{1}{h}k^2\left(\frac{x-\overline{x}}{h}\right)ZZ'\sigma^2(x)dF_o(x)
\end{aligned}
$$

It helps to remember that Z_iZ_i' is a function of the x,

$$
Z_iZ_i'=\begin{bmatrix}
1 & \left(\frac{x_i-\overline{x}}{h}\right) & \cdots & \left(\frac{x_i-\overline{x}}{h}\right)^p \\
\left(\frac{x_i-\overline{x}}{h}\right) & \left(\frac{x_i-\overline{x}}{h}\right)^2 & \cdots & \left(\frac{x_i-\overline{x}}{h}\right)^{p+1} \\
\vdots & \vdots & \ddots & \vdots \\
\left(\frac{x_i-\overline{x}}{h}\right)^p & \left(\frac{x_i-\overline{x}}{h}\right)^{p+1} & \cdots & \left(\frac{x_i-\overline{x}}{h}\right)^{2p}
\end{bmatrix}
$$

Let $\delta_j^+=\int_{\overline{x}}^{\overline{x}+h}\frac{1}{h}k^2\left(\frac{x-\overline{x}}{h}\right)\left(\frac{x-\overline{x}}{h}\right)^j\sigma^2(x)dF_o(x)=\int_0^\infty k^2(u)\,u^j\sigma^2(\overline{x}+uh)dF_o(\overline{x}+uh)du$ and Δ_+^* is the $(p+1)\times(p+1)$ matrix with (j,l) element δ_{j+l-2}^+ for $j,l=1,...,p+1$. Then,

$$
\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)d_iZ_i\varepsilon_i\xrightarrow{p}N(0,\Delta_+^*)
$$

Similarly we can show that

$$
\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)(1-d_i)Z_i\varepsilon_i\xrightarrow{p}N(0,\Delta_-^*)
$$

where Δ_-^* is the $(p+1)\times(p+1)$ matrix with (j,l) element δ_{j+l-2}^- for $j,l=1,...,p+1$, and $\delta_j^- = \int_{\overline{x}-h}^{\overline{x}} \frac{1}{h}k^2\left(\frac{x-\overline{x}}{h}\right)\left(\frac{x-\overline{x}}{h}\right)^j \sigma^2(x)dF_o(x) = (-1)^j\int_0^\infty k^2(u)\,u^j\sigma^2(\overline{x}-uh)dF_o(\overline{x}-uh)$.

The bias term is given by

$$\sqrt{nh}e_1'\left\{D_{n+}\left[\frac{1}{nh}\sum_{i=1}^n k\left(\frac{x_i-\overline{x}}{h}\right)d_iZ_im(x_i)\right]-D_{n-}\left[\frac{1}{nh}\sum_{i=1}^n k\left(\frac{x_i-\overline{x}}{h}\right)(1-d_i)Z_im(x_i)\right]\right\}$$

Note that,

$$\begin{aligned}E\left[\frac{1}{nh}\sum_{i=1}^n k\left(\frac{x_i-\overline{x}}{h}\right)d_iZ_im(x_i)\right] &= E\left[\frac{1}{h}k\left(\frac{x_i-\overline{x}}{h}\right)d_iZ_im(x_i)\right]\\ &= \int_0^\infty k(u)\,Z(\overline{x}+uh)m(\overline{x}+uh)dF_o(\overline{x}+uh)\end{aligned}$$

and similarly, below the cutoff. Hence, the bias term can be approximated by,

$$e_1'\left\{\begin{array}{l}\left(\Gamma_+^*\right)^{-1}\left[\int_0^\infty k(u)\,Z(\overline{x}+uh)m(\overline{x}+uh)dF_o(\overline{x}+uh)\right]-\\ -\left(\Gamma_-^*\right)^{-1}\left[\int_0^\infty k(u)\,Z(\overline{x}-uh)m(\overline{x}-uh)dF_o(\overline{x}-uh)\right]\end{array}\right\}$$

■

There are two noteworthy cases in which the formulas for the fixed-h asymptotic variance and bias simplify to those of the small-h approximation. First, when $h\to 0$, the fixed-h formulas for the asymptotic variance and bias approach the small-h approximation formulas.

Corollary 1 *If, in the bandwidth around the cutoff, $f_o(x)$, $\sigma^2(x)$ and $m(x)$ are continuous, then*

$$\lim_{h\to 0} V_{fixed-h} = V_{small-h}$$
$$\lim_{h\to 0} B_{fixed-h} = B_{small-h}$$

Proof of Corollary 1. First, note that, if $h\to 0$,

$$\gamma_j^+ = \lim_{h\to 0}\int_0^\infty k(u)\,u_o^j f_o(\overline{x}+uh)du = f_o(\overline{x})\int_0^\infty k(u)\,u^j du = f_o(\overline{x})\gamma_j \tag{11}$$

and

$$\delta_j^+ = \lim_{h\to 0}\int_0^\infty k^2(u)\,u^j\sigma^{2+}(\overline{x}+uh)f_o(\overline{x}+uh)du = \sigma^{2+}(\overline{x})f_o(\overline{x})\int_0^\infty k^2(u)\,u^j du = \sigma^{2+}(\overline{x})f_o(\overline{x})\delta_j \tag{12}$$

and similarly for γ_j^- and δ_j^-. Then, for the variance,

$$\begin{aligned}&\lim_{h\to 0}\left(\Gamma_+^*\right)^{-1}\Delta_+^*\left(\Gamma_+^*\right)^{-1}+\left(\Gamma_-^*\right)^{-1}\Delta_-^*\left(\Gamma_-^*\right)^{-1}\\ &= \frac{\sigma^{2+}(\overline{x})+\sigma^{2-}(\overline{x})}{f_o(\overline{x})}e_1'\Gamma^{-1}\Delta\Gamma^{-1}e_1\end{aligned}$$

For the bias, if we approximate $m(\overline{x}+uh)=m(x)$ just above $m(\overline{x})$:

$$m(x) = m(\overline{x})+m'^{+}(\overline{x})(x-\overline{x})+...+\frac{1}{p!}m^{(p)+}(\overline{x})(x-\overline{x})^p+\frac{1}{(p+1)!}m^{(p+1)+}(\overline{x})(x-\overline{x})^{p+1}+o\left(h^{p+1}\right)$$

and similarly for approximating $m(x)$ just below the cutoff. When we evaluate the linear projection of $m(x)$ on $Z(x)$ at $\overline{x}$, we get the intercept $m(\overline{x})$ and the "residual" as described above. A helpful fact is that, by the definition of $Z(x)$,

$$\int_0^\infty k\left(u\right) Z(\overline{x}+uh)u^{p+1}du = \begin{bmatrix} \gamma_{p+1} \\ \vdots \\ \gamma_{2p+1} \end{bmatrix} \tag{13}$$

$$\int_0^\infty k\left(u\right) Z(\overline{x}-uh)u^{p+1}du = \begin{bmatrix} \gamma_{p+1} \\ \vdots \\ (-1)^p\gamma_{2p+1} \end{bmatrix} \tag{14}$$

Note that $\Gamma^{-1}\begin{bmatrix} \gamma_{p+1} \\ \vdots \\ \gamma_{2p+1} \end{bmatrix}$ is equal both above and below the cutoff. The bias formula in theorem 1 is given by

$$e_1'\left\{ \begin{array}{l} \left(\Gamma_+^*\right)^{-1}\left[\int_0^\infty k\left(u\right) Z(\overline{x}+uh)m(\overline{x}+uh)f_o(\overline{x}+uh)du\right] - \\ -\left(\Gamma_-^*\right)^{-1}\left[\int_0^\infty k\left(u\right) Z(\overline{x}-uh)m(\overline{x}-uh)f_o(\overline{x}-uh)du\right] \end{array} \right\} \tag{15}$$

as discussed in Section 3 in the main text the main term is just the difference between the intercepts of the linear projections of $k\left(u\right)m(x)$ on $k\left(u\right)Z(x)$ in the bandwidth above below the cutoff, which is equal to the linear projections evaluated at $\overline{x}$. Hence, plugging the bias formula for the linear projection, formula (15) can be written as

$$\begin{aligned} &e_1'\left\{ \begin{array}{l} \left[\left(\Gamma_+^*\right)^{-1}\int_0^\infty k\left(u\right) Z(\overline{x}+uh)\left(\frac{1}{(p+1)!}m^{(p+1)+}(\overline{x})\left(uh\right)^{p+1}+o\left(h^{p+1}\right)\right)f_o(\overline{x}+uh)du\right] \\ -\left[\left(\Gamma_-^*\right)^{-1}\int_0^\infty k\left(u\right) Z(\overline{x}-uh)\left(\frac{(-1)^{p+1}}{(p+1)!}m^{(p+1)-}(\overline{x})\left(uh\right)^{p+1}+o\left(h^{p+1}\right)\right)f_o(\overline{x}-uh)du\right] \end{array} \right\} \\ &= \frac{h^{p+1}}{(p+1)!}e_1'\left\{ \left\{ \begin{array}{l} m^{(p+1)+}(\overline{x})\left[\left(\Gamma_+^*\right)^{-1}\int_0^\infty k\left(u\right) Z(\overline{x}+uh)u^{p+1}f_o(\overline{x}+uh)du\right] \\ -\left(-1\right)^{p+1}m^{(p+1)-}(\overline{x})\left[\left(\Gamma_-^*\right)^{-1}\int_0^\infty k\left(u\right) Z(\overline{x}-uh)u^{p+1}f_o(\overline{x}-uh)du\right] \end{array} \right\} + o\left(1\right) \right\} \end{aligned}$$

If $h \to 0$, using the equalities in equations (11), (12), (13) and (14),

$$= \frac{\lim_{h\to 0}\left(h^{p+1}\right)}{(p+1)!}\left[m^{(p+1)+}(\overline{x}) - \left(-1\right)^{p+1}m^{(p+1)-}(\overline{x})\right]e_1'\Gamma^{-1}\begin{bmatrix} \gamma_{p+1} \\ \vdots \\ \gamma_{2p+1} \end{bmatrix}$$

which is the same limit of the bias term in the small-h approximation. ■

Hence, if h is small, fixed-h and small-h provide similar approximations to the asymptotic behavior of $\widehat{\alpha}$.

Secondly, if $f_o(x)$ and $\sigma^2(x)$ are constant around the cutoff and $m(x)$ can be exactly approximated by a polynomial of order $p+1$, the fixed-h asymptotic variance and bias approximations simplify to the small-h asymptotic formulas.

Corollary 2 *If, in the bandwidth around the cutoff, $f_o(x)$, $\sigma^2(x)$ are constant and $m(x)$ can be exactly approximated by an expansion of order $p+1$, then the asymptotic variance and bias of $\sqrt{nh}(\widehat{\alpha}_p - \alpha)$ obtained by fixed-h (theorem 1) and small-h (Porter 2003) are the same.*

$$V_{fixed-h} = V_{small-h}$$
$$B_{fixed-h} = B_{small-h}$$

Proof of Corollary 2. First, note that, if $h > 0$ and, in the bandwidth around the cutoff, $f_o(x) = f_o(\overline{x})$, $\sigma^2(x) = \sigma^2(\overline{x})$ and

$$m(x) = m(\overline{x}) + m'^{+}(\overline{x})(x-\overline{x}) + ... + \frac{1}{p!}m^{(p)+}(\overline{x})(x-\overline{x})^p + \frac{1}{(p+1)!}m^{(p+1)+}(\overline{x})(x-\overline{x})^{p+1}$$

then,

$$\gamma_j^+ = \int_0^\infty k(u)u^j f_o(\overline{x}+uh)du = f_o(\overline{x})\int_0^\infty k(u)u^j du = f_o(\overline{x})\gamma_j \tag{16}$$

and

$$\delta_j^+ = \int_0^\infty k^2(u)u^j\sigma^2(\overline{x}+uh)f_o(\overline{x}+uh)du = \sigma^{2+}(\overline{x})f_o(\overline{x})\int_0^\infty k^2(u)u^j du = \sigma^{2+}(\overline{x})f_o(\overline{x})\delta_j \tag{17}$$

and similarly for γ_j^- and δ_j^-. Then, for the variance,

$$\begin{aligned}&\left(\Gamma_+^*\right)^{-1}\Delta_+^*\left(\Gamma_+^*\right)^{-1} + \left(\Gamma_-^*\right)^{-1}\Delta_-^*\left(\Gamma_-^*\right)^{-1}\\ &= \frac{\sigma^{2+}(\overline{x})+\sigma^{2-}(\overline{x})}{f_o(\overline{x})}e_1'\Gamma^{-1}\Delta\Gamma^{-1}e_1\end{aligned}$$

For the bias, the strategy is basically the same as in the proof of corollary 1:

$$m(x) = m(\overline{x}) + m'^{+}(\overline{x})(x-\overline{x}) + ... + \frac{1}{p!}m^{(p)+}(\overline{x})(x-\overline{x})^p + \frac{1}{(p+1)!}m^{(p+1)+}(\overline{x})(x-\overline{x})^{p+1}$$

Once again the bias formula in theorem 1 is given by

$$e_1'\left\{\begin{array}{c}\left(\Gamma_+^*\right)^{-1}\left[\int_0^\infty k(u)Z(\overline{x}+uh)m(\overline{x}+uh)f_o(\overline{x}+uh)du\right] - \\ -\left(\Gamma_-^*\right)^{-1}\left[\int_0^\infty k(u)Z(\overline{x}-uh)m(\overline{x}-uh)f_o(\overline{x}-uh)du\right]\end{array}\right\}$$

Plugging the bias formula for the linear projection:

$$\begin{aligned}&e_1'\left\{\begin{array}{c}\left[\left(\Gamma_+^*\right)^{-1}\int_0^\infty k(u)Z(\overline{x}+uh)\left(\frac{1}{(p+1)!}m^{(p+1)+}(\overline{x})(uh)^{p+1}\right)f_o(\overline{x}+uh)du\right]\\ -\left[\left(\Gamma_-^*\right)^{-1}\int_0^\infty k(u)Z(\overline{x}-uh)\left(\frac{(-1)^{p+1}}{(p+1)!}m^{(p+1)-}(\overline{x})(uh)^{p+1}\right)f_o(\overline{x}-uh)du\right]\end{array}\right\}\\ &= \frac{h^{p+1}}{(p+1)!}e_1'\left\{\begin{array}{c}\left[\left(\Gamma_+^*\right)^{-1}\int_0^\infty k(u)Z(\overline{x}+uh)m^{(p+1)+}(\overline{x})u^{p+1}f_o(\overline{x}+uh)du\right]\\ -\left[\left(\Gamma_-^*\right)^{-1}\int_0^\infty k(u)Z(\overline{x}-uh)\left((-1)^{p+1}m^{(p+1)-}(\overline{x})u^{p+1}\right)f_o(\overline{x}-uh)du\right]\end{array}\right\}\\ &= \frac{h^{p+1}}{(p+1)!}e_1'\left\{\begin{array}{c}m^{(p+1)+}(\overline{x})\left[\left(\Gamma_+^*\right)^{-1}\int_0^\infty k(u)Z(\overline{x}+uh)u^{p+1}f_o(\overline{x}+uh)du\right]\\ -(-1)^{p+1}m^{(p+1)-}(\overline{x})\left[\left(\Gamma_-^*\right)^{-1}\int_0^\infty k(u)Z(\overline{x}-uh)u^{p+1}f_o(\overline{x}-uh)du\right]\end{array}\right\}\end{aligned}$$

Using $f_o(x) = f_o(\overline{x})$ and the equalities in formulas (16), (17), (13) and (14),

$$= \frac{h^{p+1}}{(p+1)!}\left[m^{(p+1)+}(\overline{x}) - (-1)^{p+1}\, m^{(p+1)-}(\overline{x})\right] e_1'\Gamma^{-1}\begin{bmatrix}\gamma_{p+1}\\ \vdots \\ \gamma_{2p+1}\end{bmatrix}$$

■

8.3.3 Fuzzy Regression Discontinuity Design

Proof of Theorem 2. The covariance between the estimators for the outcome of interest and the treatment probability will be given by two independent terms, one for each side of the threshold. The upper side is given by

$$E\left\{e_1'D_{n+}\left[\frac{1}{nh}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)d_iZ_iy_i\right]\left[\frac{1}{nh}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)d_iZ_it_i\right]D_{n+}e_1\right\}$$

Where t_i is the dummy variable indicating that the observation has received treatment. In obtaining the asymptotic covariance, the bias term of the estimator can be ignored, hence

$$\begin{aligned}
&E\left[e_1'D_{n+}\left(\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)d_iZ_i\varepsilon_i\right)\left(\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)d_iZ_i\eta_i\right)'D_{n+}e_1\right]\\
&= E\left[e_1'D_{n+}\left(\frac{1}{nh}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)^2 d_iZ_iZ_i'E\left[\varepsilon_i\eta_i \mid X=x\right]\right)D_{n+}e_1\right]\\
&= e_1'D_{n+}\left[\int_{\overline{x}}^{\overline{x}+h}\frac{1}{h}k\left(\frac{x-\overline{x}}{h}\right)^2 ZZ'\sigma_{\varepsilon\eta}(x)dF_o(x)\right]D_{n+}e_1
\end{aligned}$$

where I used the assumption that $E\left[\varepsilon_i\eta_j \mid X=x\right]=0$ for $j\neq i$.

Similarly for the second term,

$$\begin{aligned}
&E\left[e_1'D_{n-}\left(\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)(1-d_i)Z_i\varepsilon_i\right)\left(\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}k\left(\frac{x_i-\overline{x}}{h}\right)(1-d_i)Z_i\eta_i\right)'D_{n-}e_1\right]\\
&= e_1'D_{n-}\left[\int_{\overline{x}-h}^{\overline{x}}\frac{1}{h}k\left(\frac{x-\overline{x}}{h}\right)^2 ZZ'\sigma_{\varepsilon\eta}(x)dF_o(x)\right]D_{n-}e_1
\end{aligned}$$

Let $\rho_j^+ = \int_{\overline{x}}^{\overline{x}+h}\frac{1}{h}k^2\left(\frac{x-\overline{x}}{h}\right)\left(\frac{x-\overline{x}}{h}\right)^j\sigma_{\varepsilon\eta}(x)dF_o(x) = \int_0^\infty k^2(u)\,u^j\sigma_{\varepsilon\eta}(\overline{x}+uh)dF_o(\overline{x}+uh)$, Δ_+^ρ is the $(p+1)\times(p+1)$ matrix with (j,l) element ρ_{j+l-2}^+ for $j,l=1,...,p+1$, $\rho_j^- = \int_{\overline{x}-h}^{\overline{x}}\frac{1}{h}k^2\left(\frac{x-\overline{x}}{h}\right)\left(\frac{x-\overline{x}}{h}\right)^j\sigma_{\varepsilon\eta}(x)dF_o(x) = (-1)^j\int_0^\infty k^2(u)\,u^j\sigma_{\varepsilon\eta}(\overline{x}-uh)dF_o(\overline{x}-uh)$ and Δ_-^ρ is the $(p+1)\times(p+1)$ matrix with (j,l) element ρ_{j+l-2}^- for $j,l=1,...,p+1$ Then the asymptotic covariance is given by

$$C_{\alpha\theta} = e_1'\left[\left(\Gamma_+^*\right)^{-1}\Delta_+^\rho\left(\Gamma_+^*\right)^{-1} + \left(\Gamma_-^*\right)^{-1}\Delta_-^\rho\left(\Gamma_-^*\right)^{-1}\right]e_1$$

■

Similarly to the result in corollary 1, as $h\to 0$, the fixed-h covariance formulas converges to the small-h asymptotic covariance.

Corollary 3 *If, in the bandwidth around the cutoff, $\sigma_{\varepsilon\eta}(x)$ is continuous and $h \longrightarrow 0$, then the asymptotic covariance, $C_{\alpha\theta}$, obtained by fixed-h (theorem 2) and small-h (Porter 2003) are the same:*

$$\lim_{h \longrightarrow 0} C_{\alpha\theta} = \frac{\sigma^{+}_{\varepsilon\eta}(\overline{x}) + \sigma^{-}_{\varepsilon\eta}(\overline{x})}{f_o(\overline{x})} e_1' \Gamma^{-1} \Delta \Gamma^{-1} e_1$$

Proof of Corollary 3. Using the results in equation 11 and noting that, if $h \to 0$

$$\rho_j^{+} = \lim_{h \to 0} \int_0^{\infty} k^2(u)\, u^j \sigma_{\varepsilon\eta}(\overline{x} + uh) f_o(\overline{x} + uh) du = \sigma^{+}_{\varepsilon\eta}(\overline{x}) f_o(\overline{x}) \delta_j$$

and similarly for ρ_j^{-}. Then,

$$\begin{aligned}
&\lim_{h \to 0} e_1' \left[\left(\Gamma^{*}_{+}\right)^{-1} \Delta^{\rho}_{+} \left(\Gamma^{*}_{+}\right)^{-1} + \left(\Gamma^{*}_{-}\right)^{-1} \Delta^{\rho}_{-} \left(\Gamma^{*}_{-}\right)^{-1} \right] e_1 \\
&= e_1' \left[\left(f_o(\overline{x})\Gamma\right)^{-1} \sigma^{+}_{\varepsilon\eta}(\overline{x}) f_o(\overline{x}) \Delta \left(f_o(\overline{x})\Gamma\right)^{-1} + \left(f_o(\overline{x})\Gamma\right)^{-1} \sigma^{-}_{\varepsilon\eta}(\overline{x}) f_o(\overline{x}) \Delta \left(f_o(\overline{x})\Gamma\right)^{-1} \right] e_1 \\
&= \frac{\sigma^{+}_{\varepsilon\eta}(\overline{x}) + \sigma^{-}_{\varepsilon\eta}(\overline{x})}{f_o(\overline{x})} e_1' \Gamma^{-1} \Delta \Gamma^{-1} e_1
\end{aligned}$$

■

Also, a result similar to the corollary 2 is readily available.

Corollary 4 *If in the bandwidth around the cutoff, $f_o(x)$ and $\sigma_{\varepsilon\eta}(x)$ are constant, then the asymptotic covariance, $C_{\alpha\theta}$, obtained by fixed-h (theorem 2) and small-h (Porter 2003) are the same.*

$$C_{\alpha\theta} = \frac{\sigma^{+}_{\varepsilon\eta}(\overline{x}) + \sigma^{-}_{\varepsilon\eta}(\overline{x})}{f_o(\overline{x})} e_1' \Gamma^{-1} \Delta \Gamma^{-1} e_1$$

Proof of Corollary 4. The proof follows very closely corollary 3. Using the results in equation 11 and noting that, if $h > 0$ and $f_o(x) = f_o(\overline{x})$ and $\sigma_{\varepsilon\eta}(x) = \sigma_{\varepsilon\eta}(\overline{x})$ for any x in the range around the cutoff

$$\rho_j^{+} = \int_0^{\infty} k^2(u)\, u^j \sigma_{\varepsilon\eta}(\overline{x} + uh) f_o(\overline{x} + uh) du = \sigma^{+}_{\varepsilon\eta}(\overline{x}) f_o(\overline{x}) \delta_j$$

and similarly for ρ_j^{-}. Then,

$$\begin{aligned}
&e_1' \left[\left(\Gamma^{*}_{+}\right)^{-1} \Delta^{\rho}_{+} \left(\Gamma^{*}_{+}\right)^{-1} + \left(\Gamma^{*}_{-}\right)^{-1} \Delta^{\rho}_{-} \left(\Gamma^{*}_{-}\right)^{-1} \right] e_1 \\
&= e_1' \left[\left(f_o(\overline{x})\Gamma\right)^{-1} \sigma^{+}_{\varepsilon\eta}(\overline{x}) f_o(\overline{x}) \Delta \left(f_o(\overline{x})\Gamma\right)^{-1} + \left(f_o(\overline{x})\Gamma\right)^{-1} \sigma^{-}_{\varepsilon\eta}(\overline{x}) f_o(\overline{x}) \Delta \left(f_o(\overline{x})\Gamma\right)^{-1} \right] e_1 \\
&= \frac{\sigma^{+}_{\varepsilon\eta}(\overline{x}) + \sigma^{-}_{\varepsilon\eta}(\overline{x})}{f_o(\overline{x})} e_1' \Gamma^{-1} \Delta \Gamma^{-1} e_1
\end{aligned}$$

■