Adaptive Bandwidth Choice for the Regression Discontinuity Design*

Miikka Rokkanen†

April 14, 2014

Abstract

In this paper I propose an adaptive bandwidth choice algorithm for local polynomial regression-based estimators in the regression discontinuity design. The algorithm allows for different choices for the order of polynomial and kernel function. In addition, the algorithm automatically takes into account the inclusion of additional covariates as well as alternative assumptions on the variance-covariance structure of the error terms. I show that the algorithm produces a consistent estimator of the asymptotically optimal bandwidth and that the resulting regression discontinuity estimator satisfies the asymptotic optimality criterion of Li (1987). Finally, I provide Monte Carlo evidence suggesting that the proposed algorithm also performs well in finite samples.

*I would like to thank Isaiah Andrews, Joshua Angrist, Matias Cattaneo, Victor Chernozhukov, Guido Imbens, Anna Mikusheva, Whitney Newey and Parag Pathak as well as the participants at MIT Econometrics Lunch, MIT Third Year Lunch and HECER Lunch Seminar for their helpful comments and suggestions.

†Department of Economics, Massachusetts Institute of Technology. Email: rokkanen@mit.edu.

1 Introduction

The regression discontinuity (RD) design, originating from Thistlewhite and Campbell (1960), has become a popular approach in economics to identifying causal effects of various treatments. In this design the treatment of interest is either fully or partly determined by whether the value of an observed covariate, often referred to as the running variable, lies below or above a known cutoff. Under relatively weak assumptions, this allows one to identify the causal effect of the treatment for individuals at the cutoff. RD designs have been used to study, for instance, the effect of class size on student achievement (Angrist and Lavy, 1999), parental valuation of school quality (Black, 1999), the effect of financial aid on college enrollment (van der Klaauw, 2002), and the effect of Head Start on child mortality (Ludwig and Miller, 2007). In addition, Hahn, Todd, and van der Klaauw (2001) and Porter (2003), among others, have made important contributions to the literature on identification and estimation of treatment effects in the RD design.[1]

The consistency of the RD estimator relies heavily on the researcher's ability to correctly specify the functional form for the relationship between the running variable and the outcome and the relationship between the running variable and the treatment. This has lead to a widespread interest in nonparametric approaches to estimating these relationships. A common approach in the recent literature has been to use local polynomial regression (LPR), especially local linear regression. As the performance of LPR-based methods depends heavily on the choice of a smoothing parameter, often referred to as the bandwidth, a key question in implementing these methods is how to choose this parameter.

Traditionally, the bandwidth choice in empirical work using LPR-based RD estimators has been based on either ad hoc procedures or on approaches that are not directly suited to the RD design (Ludwig and Miller, 2005; DesJardins and McCall, 2008). However, in a recent influential paper Imbens and Kalyanaraman (2012) studied in depth the problem of optimal bandwidth choice for local linear regression-based RD estimator and proposed an algorithm that can be used to obtain a consistent estimator of the asymptotically optimal RD bandwidth.[2]

This paper contributes to the literature by proposing an adaptive bandwidth choice algorithm for the LPR-based RD estimator by building on previous work by Schucany (1995) and Gerard and

[1]For extensive surveys of the literature, see Cook (2008), Imbens and Lemieux (2008), van der Klaauw (2008), and Lee and Lemieux (2010).

[2]See also Arai and Ichimura (2013) for an alternative approach to optimal bandwidth choice for the local liner regression-based RD estimator. Furthermore, Calonico, Cattaneo, and Titiunik (2014) discuss nonparametric estimation of robust confidence intervals for the local liner regression-based RD estimator.

Schucany (1997).[3] The algorithm is adaptive in the sense that it allows for different choices for the order of polynomial and kernel function. In addition, the algorithm automatically takes into account the inclusion of additional covariates as well as alternative assumptions on the variance-covariance structure of the error terms. Thus, the proposed algorithm provides a convenient approach to bandwidth choice that retains its validity in various settings.

I show that the proposed algorithm produces a consistent estimator of the asymptotically optimal bandwidth. Furthermore, the resulting RD estimator satisfies the asymptotic optimality criterion of Li (1987) and converges to the true parameter value at the optimal nonparametric rate (Stone, 1982; Porter, 2003). I also provide Monte Carlo evidence illustrating that that the propoced algorithm works well in finite sample and compares favorably to the algorithm by Imbens and Kalyanaraman (2012).

The rest of the paper is structured as follows. Section 2 reviews the RD design and the LPR-based RD estimator. Section 3 introduces the proposed bandwidth choice algorithm and discusses its asymptotic properties. Section 4 presents Monte Carlo evidence illustrating the finite-sample performance of the proposed algorithm. Section 5 concludes.

2 Regression Discontinuity Design

2.1 Setting and Parameter of Interest

Suppose one is interestein the causal effect of a binary treatment on some outcome. Let D denote an indicator that equals 1 if an individual receives the treatment and 0 otherwise. Furthermore, let Y_1 and Y_0 denote the potential outcomes when an individual receives and does not receive the treatment. The observed outcome of an individual, denoted by Y, is

$$Y = (1 - D) \times Y_0 + D \times Y_1.$$

In a sharp regression discontinuity (SRD) design D is a deteministic function of a continuous running variable R:[4]

$$D = 1(R \geq c)$$

[3]Similar approaches to optimal bandwidth choice have also been proposed by Ruppert (1997), Doksum, Peterson, and Samarov (2000), and Prewitt (2003).

[4]I focus solely on SRD design in this paper. Fuzzy RD design is a straightforward extension that I leave for future research.

where $1(\cdot)$ is an indicator function equal to 1 if the statement in parentheses is true and 0 otherwise. In words, all individuals with the value of R at or above a cutoff c are assigned to the treatment group while all individuals with the value of R below the cutoff c are assigned to the control group. Furthermore, there is perfect compliance with the treatment assignment: all of the individuals assigned to the treatment group receive the treatment whereas none of the individuals assigned to the control group receive the treatment.

Given the treatment assignment mechanism, a natural parameter of interest in the SRD design is

$$\tau = E[Y_1 - Y_0 \mid R = c],$$

that is, the average effect of the treatment for individuals at the cutoff. Suppose that $E[Y_1 \mid R = r]$ and $E[Y_0 \mid R = r]$ exist and are continuous at $R = c$. Then

$$\tau = \lim_{r \downarrow c} E[Y \mid R = r] - \lim_{r \uparrow c} E[Y \mid R = r]$$

where $m(r) = E[Y \mid R = r]$, $m_+(c) = m(r)$ and $m_-(c) = m(r)$. Thus, under relatively mild assumptions τ is nonparametrically identified and given by the difference in the limits of two conditional expectation functions at the cutoff c.

2.2 Estimation using Local Polynomial Regression

I focus in this paper on the estimation of τ using separate LPRs on both sides of the cutoff.[5] An attractive property of the LPR-based approach is that it allows one to obtain a consistent estimator of τ without reliance on strong functional form assumptions. Moreover, the LPR-based approach reduces (and under some assumptions even eliminates) the bias that afflicts other nonparametric regression function estimates at boundary points.

Suppose we observe a sample (Y_i, R_i), $i = 1, \ldots, n$. The LPR-based estimator of τ using a polynomial of order p , kernel $K(u)$, and bandwidth h is given by

$$\hat{\tau}_p(h) = \hat{\alpha}_p^+(h) - \hat{\alpha}_p^-(h)$$

[5]For a comprehensive treatment of LPR methods, see Fan and Gijbels (1996).

where

$$\begin{bmatrix} \hat{\alpha}_p^+ (h) \\ \hat{\beta}_{1,p}^+ (h) \\ \vdots \\ \hat{\beta}_{p,p}^+ (h) \end{bmatrix} = \underset{\alpha, \{\beta_k\}_{k=1}^p}{\arg\min} \sum_{i=1}^n 1 (R_i \geq c) K \left(\frac{R_i - c}{h} \right) \left(Y_i - \alpha - \sum_{k=1}^p \beta_k (R_i - c)^k \right)^2$$

and

$$\begin{bmatrix} \hat{\alpha}_p^- (h) \\ \hat{\beta}_{1,p}^- (h) \\ \vdots \\ \hat{\beta}_{p,p}^- (h) \end{bmatrix} = \underset{\alpha, \{\beta_k\}_{k=1}^p}{\arg\min} \sum_{i=1}^n 1 (R_i < c) K \left(\frac{R_i - c}{h} \right) \left(Y_i - \alpha - \sum_{k=1}^p \beta_k (R_i - c)^k \right)^2 .$$

I have written the estimator $\hat{\tau}_p (h)$ in a way that makes explicit its dependence on the choice of the order of polynomial p and the bandwidth h. The estimator $\hat{\tau}_p (h)$ also depends on the choice of the kernel $K (u)$, but this does not play a key role in what follows. Covariates could easily be included in the model, but I abstract away from this for notational simplicity.

Let

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, X_p = \begin{bmatrix} 1 & (R_1 - c) & \cdots (R_1 - c)^p \\ \vdots & \vdots & \vdots \\ 1 & (R_n - c) & \cdots (R_n - c)^p \end{bmatrix}$$

and define

$$\begin{aligned} \hat{\beta}_p^+ (h) &= \left(X_p' W_h^+ X_p' \right)^{-1} X_p' W_h^+ Y \\ \hat{\beta}_p^- (h) &= \left(X_p' W_h^- X_p' \right)^{-1} X_p' W_h^- Y \end{aligned}$$

where

$$\begin{aligned} W_h^+ &= diag \left[1 (R_i \geq c) K \left(\frac{R_i - c}{h} \right) \right] \\ W_h^- &= diag \left[1 (R_i < c) K \left(\frac{R_i - c}{h} \right) \right] . \end{aligned}$$

We can now write $\hat{\tau}_p(h)$ equivalently as

$$\hat{\tau}_p(h) = e_1^{'}\left(\hat{\beta}_p^{+}(h) - \hat{\beta}_p^{-}(h)\right)$$

where e_1 is a $(p+1) \times 1$ vector with one as its first element and zeros as the other elements.

Using standard results for Weighted Least Squares (WLS) estimators one can write the heteroskedasticity-robust variance estimators for $\hat{\beta}_p^{+}(h)$ and $\hat{\beta}_p^{-}(h)$ as

$$\begin{aligned}\hat{v}_p^{+}(h) &= \left(X_p^{'}W_h^{+}X_p^{'}\right)^{-1} X_p^{'}W_h^{+}\hat{\Sigma}_{p,h}^{+}W_h^{+}X_p\left(X_p^{'}W_h^{+}X_p^{'}\right)^{-1} \\ \hat{v}_p^{-}(h) &= \left(X_p^{'}W_h^{-}X_p^{'}\right)^{-1} X_p^{'}W_h^{-}\hat{\Sigma}_{p,h}^{-}W_h^{+}X_p\left(X_p^{'}W_h^{-}X_p^{'}\right)^{-1}\end{aligned}$$

where

$$\begin{aligned}\hat{\Sigma}_{p,h}^{+} &= diag\left[\left(Y_i - X_{ip}^{'}\hat{\beta}_p^{+}(h)\right)^2\right] \\ \hat{\Sigma}_{p,h}^{-} &= diag\left[\left(Y_i - X_{ip}^{'}\hat{\beta}_p^{-}(h)\right)^2\right]\end{aligned}$$

and $X_{pi}^{'}$ is the i^{th} row of X_p. Thus,

$$\hat{v}_p(h) = e_1^{'}\left(\hat{v}_p^{+}(h) + \hat{v}_p^{-}(h)\right)e_1$$

provides a heteroskedasticity-robust variance estimator for $\hat{\tau}_p(h)$.[6]

As was mentioned above, there are in general three decisions one has to make when implementing LPR-based estimators: order of polynomial p, kernel $K(u)$ and bandwidth h. I focus in this paper on the choice of h conditional on the choices of p and $K(u)$. This is motivated by the observation that bandwidth choice is commonly viewed as the key decision when implementing LPR-based estimators. As the bandwidth choice algorithm proposed in this paper applies to generic p and $K(u)$, I will make only some remarks regarding these choices.

A common approach in the empirical literature is to use local linear regression-based estimators. These are convenient in practice as the number of parameters needed to be estimated is relativelely small. These estimators have also been shown to have attractive bias properties at boundary points (Fan and Gijbels, 1992) and to obtain the optimal convergence rate (Stone, 1982; Porter, 2003).

[6]Note that in practice one can compute $\hat{\tau}_p(h)$ and $\hat{v}_p(h)$ using WLS with full set of interactions for the running variable controls and the indicator variable for the value of the running variable being above the cutoff. However, for simplicity I use the above notation throughout the paper.

Common choices for the kernel include the uniform kernel $K(u) = \frac{1}{2}1(|u| \leq 1)$, the Epanechnikov kernel $K(u) = \frac{3}{4}\left(1-u^2\right)1(|u| \leq 1)$ and the triangular kernel (some authors refer to this as the edge kernel) $K(u) = (1-|u|)\,1(|u| \leq 1)$. The popularity of the uniform kernel is mainly due to its practical convenience while the Epanechnikov kernel has been shown to be optimal for estimation problems at interior points (Fan, Gasser, Gijbels, Brockmann, and Engel, 1997). The triangular kernel is instead the most appropriate choice for the RD design as it has been shown to be optimal for estimation problems at boundary points (Cheng, Fan, and Marron, 1997). Imbens and Kalyanaraman (2012), for instance, focus on this kernel in their bandwidth choice algorithm.

3 Optimal Bandwidth Choice

3.1 Infeasible Bandwidth Choice

The optimality criteria I use in this paper is the Mean Squared Error (MSE) which can be written as

$$\begin{aligned} MSE\left[\hat{\tau}_p(h)\right] &= E\left[\left(\hat{\tau}_p(h)-\tau\right)^2\right] \\ &= E\left[\hat{\tau}_p(h)-\tau\right]^2 + E\left[\left(\hat{\tau}_p(h)-E\left[\hat{\tau}_p(h)\right]\right)^2\right]. \end{aligned}$$

In words, the MSE equals the sum of the squared bias and the variance of $\hat{\tau}_p(h)$. While the estimation of the variance of $\hat{\tau}_p(h)$ is relatively straightforward, the estimation of the bias of $\hat{\tau}_p(h)$ is problematic. Thus, it is typically difficult to obtain a good estimator of the bandwidth that minimizes the MSE.[7] I follow instead the standard approach in the literature on LPR-based methods and focus on the first-order asymptotic approximation of the MSE refered to as the Asymptotic Mean Squared Error (AMSE). Furthermore, I focus on the case in which the bandwidth is restricted to be the same on both sides of the cutoff as opposed to choosing a different bandwidth to the left and to the right of the cutoff.[8]

I will next state the assumptions used throughout this paper.

Assumption A.

1. The observations (Y_i, R_i), $i = 1, \ldots, n$, are independent and identically distributed.

2. The conditional expectation function $m(r) = E[Y_i \mid R_i = r]$ is at least $p+2$ times continuously

[7] See also the discussion in Imbens and Kalyanaraman (2012).

[8] Arai and Ichimura (2013) propose a bandwidth choice algorithm for the local linear regression-based estimator that uses separate bandwidths to the left and right of the cutoff.

differentiable at $r \neq c$*. Let* $m^{(k)}(r)$ *denote the* k^{th} *derivative of* $m(r)$*.* $\left|m^{(k)}(r)\right|$*,* $k = 0, \ldots, p+2$*, are uniformly bounded on* $(c, c+M]$ *and* $[c-M, c)$ *for some* $M > 0$*.* $\left|m_-^{(k)}(c)\right|$ *and* $\left|m_+^{(k)}(c)\right|$*,* $k = 0, \ldots, p+2$*, exist and are finite, where* $m_-^{(k)}(c)$ *and* $m_+^{(k)}(c)$ *denote the left and right limit of* $m^{(k)}(r)$ *at the cutoff* c*.*

3. The marginal distribution of the running variable R_i*, denoted by* $f(r)$*, is continuous, bounded and bounded away from zero around* c*.*

4. Let ϵ_i *denote the residual* $Y_i - m(R_i)$*. The conditional variance function* $\sigma^2(r) = Var[\epsilon_i \mid R_i = r]$ *is uniformly bounded on* $(c, c+M]$ *and* $[c-M, c)$ *for some* $M > 0$*. The left and right limits of* $\sigma^2(r)$ *at the cutoff, denoted by* $\sigma_+^2(c)$ *and* $\sigma_-^2(c)$*, exist and are finite.*

5. $E\left[\mid \epsilon_i \mid^4 \mid R_i = r\right]$ *is uniformly bounded on* $(c, c+M]$ *and* $[c-M, c)$ *for some* $M > 0$*. The limits* $\lim_{r \downarrow c} E\left[\mid \epsilon_i \mid^4 \mid R_i = r\right]$ *and* $\lim_{r \uparrow c} E\left[\mid \epsilon_i \mid^4 \mid R_i = r\right]$ *exist and are finite.*

6. The kernel $K(u)$ *is non-negative, bounded, different from zero on the compact interval* $[0,1]$ *and continuous on the open interval* $(0,1)$*.*

We can now formally define the AMSE of $\hat{\tau}_p(h)$ as

$$AMSE\left[\hat{\tau}_p(h)\right] = B_p^2 h^{2(p+1)} + \frac{V_p}{nh}$$

where

$$\begin{aligned} B_p &= \frac{m_+^{(p+1)}(c)}{(p+1)!} e_1^{'} \Gamma_+^{-1} \delta_+ - \frac{m_-^{(p+1)}(c)}{(p+1)!} e_1^{'} \Gamma_-^{-1} \delta_- \\ V_p &= \frac{\sigma_+^2(c)}{f(c)} \Gamma_+^{-1} \Lambda_+ \Gamma_+^{-1} + \frac{\sigma_-^2(c)}{f(c)} \Gamma_-^{-1} \Lambda_- \Gamma_-^{-1}. \end{aligned}$$

The vectors/matrices Γ_+, Γ_-, δ_+, δ_-, Λ_+, and Λ_-, defined in the Appendix, depend only on $K(u)$. The AMSE provides an approximation to the MSE for small h and large nh, as shown in Theorem 1. The first term of the AMSE corresponds to the square of the leading term of an asymptotic approximation of the bias of $\hat{\tau}_p(h)$. The second term of the AMSE corresponds to the leading term of an asymptotic approximation of the variance of $\hat{\tau}_p(h)$. The expression illustrates the bias-variance tradeoff inherent in the problem of choosing h: using a larger bandwidth reduces the variance of $\hat{\tau}_p(h)$, but this happens at the cost of larger bias, and vice versa.

Theorem 1 provides an expression for the asymptotically optimal bandwidth h_{opt} that minimizes the AMSE of $\hat{\tau}_p(h)$. We can see that h_{opt} is increasing in the variation of the outcome at the cutoff and decreasing in the squared difference of the curvatures of the two conditional expectation functions at the cutoff. Futhermore, h_{opt} is decreasing in the sample size. The assumption $m_+^{(p+1)}(c) \neq m_-^{(p+1)}(c)$ when p is odd is made to avoid a case in which $B_p = 0$ and consequently $h_{opt} = \infty$. It is possible to derive the optimal bandwidth also for this setting by considering a higher order expansion of the bias of $\hat{\tau}_p(h)$. However, I leave this extension for future work.[9]

Theorem 1. *Suppose that $m_+^{(p+1)}(c) \neq m_-^{(p+1)}(c)$ when p is odd. Then*

$$\begin{aligned} MSE\left[\hat{\tau}_p(h)\right] &= AMSE\left[\hat{\tau}_p(h)\right] + o_p\left(h^{2(p+1)} + \frac{1}{nh}\right) \\ h_{opt} &= \arg\min_h AMSE\left[\hat{\tau}_p(h)\right] \\ &= C_{opt} n^{-\frac{1}{2p+3}} \end{aligned}$$

where

$$C_{opt} = \left(\frac{V_p}{2(p+1)B_p^2}\right)^{\frac{1}{2p+3}}.$$

Thus, the optimal bandwidth takes the form $h_{opt} = C_{opt} n^{-\frac{1}{2p+3}}$ for some constant $C_{opt} > 0$ that depends on the unknown parameters $m_+^{(p+1)}(c)$, $m_-^{(p+1)}(c)$, $\sigma_+^2(c)$, $\sigma_-^2(c)$, and $f(c)$. The problem of optimal bandwidth choice therefore boils down to the optimal choice of C in $h = Cn^{-\frac{1}{2p+3}}$. A common approach in the statistics and econometrics literature on LPR-based estimators is to estimate the unknown parameters that enter C_{opt}. In the RD design literature such plug-in estimator has been proposed for the local linear regression case by Imbens and Kalyanaraman (2012).[10]

3.2 Bandwidth Choice Algorithm

The bandwidth choice algorithm I propose in this paper is based on direct estimation of B_p^2 and V_p without the need to separately estimate the unknown parameters incorporated in these constants. The algorithm is general enough to be directly applicable to settings with arbitrary choices regarding the order of polynomial p and the kernel $K(u)$. The proposed approach also automatically adapts to various departures from the standard setting as discussed below.

The algorithm builds on the work by Schucany (1995) and Gerard and Schucany (1997).[11] The

[9] Arai and Ichimura (2013) propose a bandwidth choice algorithm for local linear regression-based estimator that takes into account the case $m_+^{(2)}(c) = m_-^{(2)}(c)$.

[10] See also the alternative approaches to optimal bandwidth choice by Ludwig and Miller (2005) and DesJardins and McCall (2008) as well as the discussion regarding these approaches in Imbens and Kalyanaraman (2012).

[11] Similar approaches have also been proposed by Ruppert (1997), Doksum, Peterson, and Samarov (2000) and Prewitt (2003).

approach I take to estimate B_p^2 is motivated by the observation that

$$\begin{aligned}\hat{\tau}_p(h) - \tau &= B_p h^{p+1} + o_p\left(h^{p+1}\right) + O_p\left((nh)^{-\frac{1}{2}}\right) \\ \hat{\tau}_{p+1}(h) - \tau &= O\left(h^{p+2}\right) + o_p\left(h^{p+2}\right) + O_p\left((nh)^{-\frac{1}{2}}\right).\end{aligned}$$

That is, the leading term of the bias of $\hat{\tau}_{p+1}(h)$ is of higher order than that of $\hat{\tau}_p(h)$. Thus, letting $\hat{b}_p^2(h)$ denote the squared difference between these two estimators we get that

$$\hat{b}_p^2(h) = B_p^2 h^{2(p+1)} + o_p\left(h^{2(p+1)}\right) + O_p\left((nh)^{-1}\right).$$

The approach I take to estimate V_p is motivated by a similar observation as one can write the heteroskedastivity-robust variance estimator for $\hat{\tau}_p(h)$ as

$$\hat{v}_p(h) = \frac{V_p}{nh} + o_p\left(\frac{1}{nh}\right) + O_p\left((nh)^{-\frac{3}{2}}\right).$$

Taken together, these observations imply that one can estimate B_p^2 and V_p consistently by regressing $\hat{b}_p^2(h_k)$ on $h_k^{2(p+1)}$ and $\hat{v}_p(h)$ on $(nh_k)^{-1}$ using a collection of initial bandwidths h_k, $k = 1, \ldots, K$. The resulting Ordinary Least Squares (OLS) estimators of B_p^2 and V_p are

$$\begin{aligned}\hat{B}_p^2 &= \frac{\sum_{k=1}^K \hat{b}_p^2(h_k) h_k^{2(p+1)}}{\sum_{k=1}^K h_k^{4(p+1)}} \\ \hat{V}_p &= \frac{\sum_{k=1}^K \hat{v}_p(h_k)(nh_k)^{-1}}{\sum_{k=1}^K (nh_k)^{-2}}\end{aligned}$$

where the constant term in both regresions is restricted to zero.

By plugging in $\hat{B}_p^2$ and $\hat{V}_p$ to the expression for C_{opt} the estimator of the asymptotically optimal bandwidth h_{opt} becomes

$$\hat{h}_{opt} = \left(\frac{\hat{V}_p}{2(p+1)\hat{B}_p^2}\right)^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}}.$$

The asymptotic properties of the bandwidth estimator $\hat{h}_{opt}$ and the resulting RD estimator $\hat{\tau}_p\left(\hat{h}_{opt}\right)$ are stated in Theorem 2. First, $\hat{h}_{opt}$ is a consistent estimator of the asymptotically optimal bandwidth h_{opt}. Second, the RD estimator $\hat{\tau}_p\left(\hat{h}_{opt}\right)$ satisfies the asymptotic optimality criterion of Li (1987). What this means is that, in terms of the MSE, the performance of the RD estimator using the estimated bandwidth is asymptotically as good as the performance of the RD estimator using the true optimal bandwidth. Third, the RD estimator $\hat{\tau}_p\left(\hat{h}_{opt}\right)$ converges to τ at the optimal nonparametric rate (Stone, 1982; Porter, 2003).

Theorem 2. *Suppose* $h_k = c_k h^{-\gamma}$, $k = 1, \ldots, K$, *for some positive, finite constants* c_k *and* $\frac{1}{2p+5} \leq \gamma < \frac{1}{2p+3}$. *Then*

$$\begin{aligned}
\frac{\hat{h}_{opt}}{h_{opt}} - 1 &= o_p(1) \\
\frac{MSE\left[\hat{\tau}_p\left(\hat{h}_{opt}\right)\right]}{MSE\left[\hat{\tau}_p\left(h_{opt}\right)\right]} - 1 &= o_p(1) \\
\hat{\tau}_p\left(\hat{h}_{opt}\right) - \tau &= O_p\left(n^{-\frac{p+1}{2p+3}}\right).
\end{aligned}$$

Note that the consistency of the bandwidth estimator $\hat{h}_{opt}$, and consequently the optimality properties of the resulting RD estimator $\hat{\tau}_p\left(\hat{h}_{opt}\right)$, require that the initial bandwidths used to estimate B_p^2 and V_p converge to zero at a slower rate than the asymptotically optimal bandwidth h_{opt}. While this is not necessary for the consistency of $\hat{V}_p$, it is needed to ensure the consistency of $\hat{B}_p$.

A remaining question is how one should choose the parameters c_k, γ, and K that define the collection of intial bandwidths h_k, $k = 1, \ldots, K$, in Theorem 1. Unfortunately, the asymptotic theory presented above has very little to say regarding these parameters. I propose to use the rate $\gamma = \frac{1}{2p+5}$ and the quantiles $0.50, 0.51, \ldots, 0.99$ of the distribution of $|R_i - c|$ as the constants c_k, $k = 1, \ldots, K$. It should be emphasized, however, that these choices are not motivated by any theoretical considerations. One could potentially improve the performance of the algorithm by using more appropriate parameter values, but I leave this question for future research. Ideally, the resulting RD estimator $\hat{\tau}_p\left(\hat{h}_{opt}\right)$ is reasonably insensitive to these choices which is an important specification check when applying the algorithm.

4 Monte Carlo Experiments

In this section I compare the performance of the proposed bandwidth choice algorithm to the performance of the algorithm by Imbens and Kalyanaraman (2012). I follow Imbens and Kalyanaraman (2012) and explore finite sample behavior in Monte Carlo expriments that are based on the data from Lee (2008) who studies the effect of incumbency on the probability of re-election. As is common in empirical practice, I focus on the local linear regression-based estimator. Furthermore, I focus on the triangular kernel due to its optimality property mentioned above.

I consider the following functional forms for $m(r)$:[12]

$$m_1(r) = \begin{cases} 0.48 + 1.27r + 7.18r^2 + 20.21r^3 + 21.54r^4 + 7.33r^5, r < 0 \\ 0.52 + 0.84r - 3.00r^2 + 7.99r^3 - 9.01r^4 + 3.56r^5, r \geq 0 \end{cases}$$

$$m_2(r) = \begin{cases} 3r^2, r < 0 \\ 4r^2, r \geq 0 \end{cases}$$

$$m_3(r) = \begin{cases} 0.42 + 0.84r - 3.00r^2 + 7.99r^3 - 9.01r^4 + 3.56r^5, r < 0 \\ 0.52 + 0.84r - 3.00r^2 + 7.99r^3 - 9.01r^4 + 3.56r^5, r \geq 0 \end{cases}$$

$$m_4(r) = \begin{cases} 0.42 + 0.84r + 7.99r^3 - 9.01r^4 + 3.56r^5, r < 0 \\ 0.52 + 0.84r + 7.99r^3 - 9.01r^4 + 3.56r^5, r \geq 0 \end{cases}$$

In all of the designs the running variable R_i and the residual ϵ_i are generated as $R_i \sim 2Beta(2,4) - 1$ and $\epsilon_i \sim N(0, 0.1295^2)$. I compare the behavior of the bandwidth choice algorithms in samples of size 100, 500, 1,000, 5,000, 10,000, and 50,000 using 1,000 replications.

The results from the Monte Carlo experiments are reported in Tables 1-4. The relative behavior of the two algorithms is similar across the different Monte Carlo designs and sample sizes. There are a few observations one can make based on these results. First, the adaptive algorithm tends to produce smaller bandwidths that vary somewhat more from one sample to another. Second, the bias of the RD estimator produced by the adative algorithm tends to be smaller. For the variance the situation is less clear: the adaptive algorithm tends to produce a less precise RD estimator in designs 1 and 3 while the opposite is true for designs 3 and 4. Finally, in terms of the MSE the adaptive algorithm performs better than the algorithm by Imbens and Kalyanaraman (2012) in designs 1, 2 and 4 once the sample size is at least 500 or 1,000 depending on the design. In design 2 the proposed algorithm performs instead worse than he algorithm by Imbens and Kalyanaraman (2012) across all of the sample sizes.

Taken together, the results from the Monte Carlo experiments suggest that the adaptive algorithm has good finite-sample properties. This seems to be especially true for moderate sample sizes

[12]See the discussion in Imbens and Kalyanaraman (2012) regarding the choice of these functional forms.

typically encountered in empirical applications. The proposed algorithm also compares well to the algorithm by Imbens and Kalyanaraman (2012).

5 Conclusions

This paper introduces an adaptive bandwidth choice algorithm for local polynomial regression-based estimators in the RD design. The algorithm is adaptive in the sense that it allows for different choices for the order of polynomial and kernel function. In addition, the algorithm automatically takes into account the inclusion of additional covariates as well as alternative assumptions on the variance-covariance structure of the error terms. I show that the algorithm produces a consistent estimator of the asymptotically optimal bandwidth that minimizes the AMSE as well as that the resulting RD estimator satisfies the asymptotic optimality criterion of Li (1987) and converges to the true parameter value at the optimal nonparametric convergence rate (Stone, 1982; Porter, 2003). Furthermore, Monte Carlo experiments suggest that the proposed algorithm has satisfactory finite-sample behavior and performs well in comparison to the algorithm by Imbens and Kalyanaraman (2012) for a local linear regression-based estimator.

I focus in the paper on sharp RD designs in which treatment is fully determined by the running variable. However, the proposed algorithm can be straightforwardly extended to fuzzy RD designs in which there is imperfect compliance with the treatment assignment. Another setting the approach can be applied to is the regression kink design (Card, Lee, Pei, and Weber, 2012) in which a continuous treatment variable has a kink instead of a discontinuity at a known cutoff. I leave these extensions for future research.

References

Angrist, J. D., and V. Lavy (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quaterly Journal of Economics*, 114(2), 533–575.

Arai, Y., and H. Ichimura (2013): "Bandwidth Selection for Differences of Nonparametric Estimators with an Application to the Regression Discontinuity Design," Unpublished manuscript.

Black, S. E. (1999): "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quaterly Journal of Economics*, 114(2), 577–599.

Calonico, S., M. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," Unpublished manuscript.

Card, D., D. S. Lee, Z. Pei, and A. Weber (2012): "Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kink Design," Unpublished manuscript.

Cheng, M.-Y., J. Fan, and J. S. Marron (1997): "On Automatic Boundary Corrections," *The Annals of Statistics*, 25(4), 1691–1708.

Cook, T. D. (2008): ""Waiting for Life to Arrive": A History of the Regression-Disconinuity Design in Psychology, Statistics and Economics," *Journal of Econometrics*, 142(2), 636–654.

DesJardins, S. L., and B. P. McCall (2008): "The Impact of the Gates Millennium Scholars Program on the Retention, College Finance- and Work-Related Choices, and Future Educational Aspirations of Low-Income Minority Students," Unpublished manuscript.

Doksum, K., D. Peterson, and A. Samarov (2000): "On Variable Bandwidth Selection in Local Polynomial Regression," *Journal of the Royal Statistical Society: Series B*, 62(3), 431–448.

Fan, J., T. Gasser, I. Gijbels, M. Brockmann, and J. Engel (1997): "Local Polynomial Regression: Optimal Kernels and Asymptotic Minimax Efficiency," *Annals of the Institute of Statistical Mathematics*, 49(1), 79–99.

Fan, J., and I. Gijbels (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *The Annals of Statistics*, 20(4), 2008–2036.

——— (1996): *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.

Gerard, P. D., and W. R. Schucany (1997): "Methodology for Nonparametric Regression from Independent Sources," *Computational Statistics & Data Analysis*, 25, 287–304.

Hahn, J., P. Todd, and W. van der Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69(1), 201–209.

Imbens, G., and K. Kalyanaraman (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Ecomic Studies*, 79(3), 933–959.

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

Lee, D. S. (2008): "Randomized Experiments from Non-Random Selection in U.S. House Elections," *Journal of Econometrics*, 142(2), 675–697.

Lee, D. S., and T. Lemieux (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355.

Li, K.-C. (1987): "Asymptotic Optimality for Cp, CL, Cross-Validation and Generalized Cross-Validation: Discrete Index Set," *The Annals of Statistics*, 15(3), 958–975.

Ludwig, J., and D. L. Miller (2005): "Does Head Start Improve Children's Life Changes? Evidence from a Regression Discontinuity Design," Working Paper 11702, National Bureau of Economic Research (NBER).

——— (2007): "Does Head Start Improve Children's Life Changes? Evidence from a Regression Discontinuity Design," *Quaterly Journal of Statistics*, 122(1), 159–208.

Porter, J. (2003): "Estimation in the Regression Discontinuity Model," Unpublished manuscript.

Prewitt, K. A. (2003): "Efficient Bandwidth Selection in Non-Parametric Regression," *Scandinavian Journal of Statistics*, 30(1), 75–92.

Ruppert, D. (1997): "Empirical-Bias Bandwidths for Local Polynomial Nonparametric Regression and Density Estimation," *Journal of American Statistical Association*, 92(439), 1049–1062.

Schucany, W. R. (1995): "Adaptive Bandwidth Choice for Kernel Regression," *Journal of the American Statistical Association*, 90(430), 535–540.

Stone, C. J. (1982): "Optimal Global Rates of Convergence for Nonparametric Regression," *The Annals of Statistics*, 10(4), 1040–1053.

Thistlewhite, D. L., and D. T. Campbell (1960): "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," *Journal of Educational Psychology*, 51(6), 309–317.

van der Klaauw, W. (2002): "Estimating the Effect of Financial Aid Offers on College Enrollment - A Regression-Discontinuity Approach," *International Economic Review*, 43(4), 1249–1287.

——— (2008): "Regression-Discontinuity Analysis: A Survey of Recent Developments in Economics," *Labour*, 22(2), 219–245.

Appendix: Proofs

Preliminaries

The estimator of $m_+(c)$ using a p^{th} order polynomial and a bandwidth h,$\hat{\alpha}_p^+(h)$, is obtained by solving

$$\min_{\alpha,\{\beta_k\}_{k=1}^p} \sum_{i=1}^N K_+\left(\frac{R_i - c}{h}\right)\left(Y_i - \alpha - \sum_{k=1}^p \beta_k (R_i - c)^k\right)^2$$

where $K_+(u) = 1(u \geq 0) K(u)$. Let $\alpha^+ = m_+(c)$ and $\beta_k^+ = m_+^{(k)}(c)$, $k = 1, \ldots, p$, and define $U_i = Y_i - \alpha^+ - \sum_{k=1}^p \beta_k^+ (R_i - c)^k$. Using this notation the minimization problem can be rewritten as

$$\min_{(\alpha-\alpha^+),\left\{h^k\left(\beta_k-\beta_k^+\right)\right\}_{k=1}^p} \sum_{i=1}^N 1(R_i \geq c) K_+\left(\frac{R_i - c}{h}\right)\left(U_i - (\alpha - \alpha^+) - \sum_{k=1}^p h^k\left(\beta_k - \beta_k^+\right)\left(\frac{R_i - c}{h}\right)^k\right)^2.$$

From the first order conditions we get

$$\begin{bmatrix} \hat{\alpha}_p^+(h) - \alpha^+ \\ h\left(\hat{\beta}_{1,p}^+(h) - \beta_1^+\right) \\ \vdots \\ h^p\left(\hat{\beta}_{p,p}^+(h) - \beta_p^+\right) \end{bmatrix} = \left[\frac{1}{nh_n}\sum_{i=1}^N K_+\left(\frac{R_i - c}{h}\right) X_{pi} X_{pi}'\right]^{-1}\left[\frac{1}{nh}\sum_{i=1}^N K_+\left(\frac{R_i - c}{h}\right) X_{pi} U_i\right]$$

where $X_{pi} = \begin{bmatrix} 1 & \left(\frac{R_i - c}{h}\right) & \cdots & \left(\frac{R_i - c}{h}\right)^p \end{bmatrix}'$.

Finally, define

$$\xi(r) = m(r) - \alpha^+ - \sum_{k=1}^p \beta_k^+ (r - c)^p - \frac{m_+^{(p+1)}(c)}{(p+1)!}(r - c)^{p+1}$$

and note that

$$\sup_{r \in (c, c+Mh]} |\xi(r)| = O\left(h^{p+2}\right).$$

Lemma 1

$$\frac{1}{nh}\sum_{i=1}^N K_+\left(\frac{R_i - c}{h}\right) X_{pi} X_{pi}' = f(c)\Gamma_+ + o(1) + o_p(1)$$

where

$$\Gamma_+ = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_p \\ \gamma_1 & \gamma_2 & \cdots & \gamma_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_p & \gamma_{p+1} & \cdots & \gamma_{2p} \end{bmatrix}$$

and

$$\gamma_k = \int_0^\infty u^k K(u)\, du.$$

Proof: Let us write

$$\frac{1}{nh}\sum_{i=1}^{N} K_+\left(\frac{R_i - c}{h}\right) X_{pi}X_{pi}' = \begin{bmatrix} A_{0,n} & A_{1,n} & \cdots & A_{p,n} \\ A_{1,n} & A_{2,n} & \cdots & A_{p+1,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{p,n} & A_{p+1,n} & \cdots & A_{2p,n} \end{bmatrix}$$

where

$$A_{k,n} = \frac{1}{nh}\sum_{i=1}^{N} K_+\left(\frac{R_i - c}{h}\right)\left(\frac{R_i - c}{h}\right)^k.$$

For the mean of $A_{k,n}$ we get

$$\begin{aligned} E\left[A_{k,n}\right] &= \frac{1}{h}E\left[K_+\left(\frac{R_i - c}{h}\right)\left(\frac{R_i - c}{h}\right)^k\right] \\ &= \frac{1}{h}\int_c^\infty K\left(\frac{r-c}{h}\right)\left(\frac{r-c}{h}\right)^k f(r)\, dr \\ &= \int_0^\infty K(u)\, u^k f(c+hu)\, du \\ &= f(c)\int_0^\infty u^k K(u)\, du + o(1) \end{aligned}$$

where the third equality follows from a change of variables $u = \frac{r-c}{h}$ and the fourth equality from the dominated convergence theorem.

For the variance of $A_{k,n}$ we get

$$\begin{aligned} Var\left[A_{k,n}\right] &\leq \frac{1}{nh^2} E\left[K_+\left(\frac{R_i-c}{h}\right)^2\left(\frac{R_i-c}{h}\right)^{2k}\right] \\ &= \frac{1}{nh^2}\int_c^{\infty} K\left(\frac{r-c}{h}\right)^2\left(\frac{r-c}{h}\right)^{2k} f(r)\,dr \\ &= \frac{1}{nh}\int_0^{\infty} K(u)^2 u^{2k} f(c+hu)\,du \\ &= o(1) \end{aligned}$$

where the second equality follows from a change of variables $u = \frac{r-c}{h}$ and the third equality from the dominated convergence theorem. □

Lemma 2

$$E\left[\frac{1}{nh}\sum_{i=1}^{N} K_+\left(\frac{R_i-c}{h}\right) X_{pi} U_i\right] = \frac{m_+^{(p+1)}(c)}{(p+1)!} f(c)\,\delta_+ h^{p+1} + O\left(h^{p+2}\right)$$

where

$$\delta_+ = \begin{bmatrix} \delta_0 \\ \vdots \\ \delta_p \end{bmatrix}$$

and

$$\delta_k = \int_0^{\infty} u^{k+p+1} K(u)\,du$$

Proof: Let

$$\frac{1}{nh}\sum_{i=1}^{N} K_{+}\left(\frac{R_i-c}{h}\right) X_{pi}U_i \quad = \quad \begin{bmatrix} A_{0,n} \\ \vdots \\ A_{p,n} \end{bmatrix}$$

where

$$\begin{aligned} A_{k,n} &= \frac{1}{nh}\sum_{i=1}^{n} K_{+}\left(\frac{R_i-c}{h}\right)\left(\frac{R_i-c}{h}\right)^{k} U_i \\ &= \frac{1}{nh}\sum_{i=1}^{n} K_{+}\left(\frac{R_i-c}{h}\right)\left(\frac{R_i-c}{h}\right)^{k} \left(\frac{m_{+}^{(p+1)}(c)}{(p+1)!}(R_i-c)^{p+1}+\xi(R_i)+\epsilon_i\right). \end{aligned}$$

We can now write

$$\begin{aligned} E\left[A_{k,n}\right] &= \frac{1}{h}E\left[K_{+}\left(\frac{R_i-c}{h}\right)\left(\frac{R_i-c}{h}\right)^{k}\left(\frac{m_{+}^{(p+1)}(c)}{(p+1)!}(R_i-c)^{p+1}+\xi(R_i)\right)\right] \\ &= \frac{m_{+}^{(p+1)}(c)}{(p+1)!}h^{p}E\left[K_{+}\left(\frac{R_i-c}{h}\right)\left(\frac{R_i-c}{h}\right)^{k+p+1}\right] \\ &\quad +\frac{1}{h}E\left[K_{+}\left(\frac{R_i-c}{h}\right)\left(\frac{R_i-c}{h_n}\right)^{k}\xi(R_i)\right] \\ &= \frac{m_{+}^{(p+1)}(c)}{(p+1)!}h^{p}\int_{c}^{\infty} K\left(\frac{r-c}{h}\right)\left(\frac{r-c}{h}\right)^{k+p+1} f(r)\,dr \\ &\quad +O\left(h^{p+1+\eta}\right)\frac{1}{h_n}\int_{c}^{\infty} K\left(\frac{r-c}{h_n}\right)\left(\frac{r-c}{h_n}\right)^{k} f(r)\,dr \\ &= \frac{m_{+}^{(p+1)}(c)}{(p+1)!}h^{p+1}\int_{0}^{\infty} K(u)\,u^{k+p+1}f(c+hu)\,du \\ &\quad +O\left(h^{p+2}\right)\int_{0}^{\infty} K(u)\,u^{k}f(c+hu)\,du \\ &= \frac{m_{+}^{(p+1)}(c)}{(p+1)!}f(c)\,h^{p+1}\left(\int_{0}^{\infty} u^{k+p+1}K(u)\,du+o(1)\right)+O\left(h^{p+2}\right). \end{aligned}$$

where the fourth equality from a change of variables $u=\frac{r-c}{h_n}$ and the fifth equality from the dominated convergence theorem. $\square$

Lemma 3

$$Var\left[\frac{1}{nh}\sum_{i=1}^{n}K_{+}\left(\frac{R_i-c}{h}\right)X_{pi}U_i\right] = \frac{1}{nh}\left(\sigma_{+}^{2}(c)\,f(c)\,\Lambda_{+}+o(1)+O\left(h^{2(p+1)}\right)\right)$$

where

$$\Lambda_{+} = \begin{bmatrix} \lambda_0 & \lambda_1 & \cdots & \lambda_p \\ \lambda_1 & \lambda_2 & \cdots & \lambda_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_p & \lambda_{p+1} & \cdots & \lambda_{2p} \end{bmatrix}$$

and

$$\lambda_k = \int_0^{\infty} u^k K(u)^2\,du.$$

Proof: Note that

$$\begin{aligned} Var\left[\frac{1}{nh}\sum_{i=1}^{n}K_{+}\left(\frac{R_i-c}{h}\right)X_{pi}U_i\right] &= E\left[Var\left[\frac{1}{nh}\sum_{i=1}^{n}K_{+}\left(\frac{R_i-c}{h}\right)X_{pi}U_i \mid X_{pi}\right]\right] \\ &\quad + Var\left[E\left[\frac{1}{nh}\sum_{i=1}^{n}K_{+}\left(\frac{R_i-c}{h}\right)X_{pi}U_i \mid X_{pi}\right]\right]. \end{aligned}$$

Let us first look at

$$E\left[Var\left[\frac{1}{nh}\sum_{i=1}^{n}K_{+}\left(\frac{R_i-c}{h}\right)X_{pi}U_i \mid X_{pi}\right]\right] = E\left[Var\left[\frac{1}{nh}\sum_{i=1}^{n}K_{+}\left(\frac{R_i-c}{h}\right)X_{pi}\epsilon_i \mid X_{pi}\right]\right].$$

I will only consider the variance of

$$A_{k,n} = \frac{1}{nh}\sum_{i=1}^{n}K_{+}\left(\frac{R_i-c}{h}\right)\left(\frac{R_i-c}{h}\right)^{k}\epsilon_i$$

as the covariance terms can be handled in a similar fashion. For this we get

$$\begin{aligned} E\left[Var\left[A_{k,n} \mid X_{pi}\right]\right] &= \frac{1}{nh^2} E\left[K_+ \left(\frac{R_i - c}{h}\right)^2 \left(\frac{R_i - c}{h}\right)^{2k} \sigma^2\left(R_i\right)\right] \\ &= \frac{1}{nh^2} \int_c^\infty K\left(\frac{r-c}{h}\right)^2 \left(\frac{r-c}{h}\right)^{2k} \sigma^2\left(r\right) f\left(r\right) dr \\ &= \frac{1}{nh} \int_0^\infty K\left(u\right)^2 u^{2k} \sigma^2\left(c+hu\right) f\left(c+hu\right) du \\ &= \frac{1}{nh} \sigma_+^2\left(c\right) f\left(c\right) \left(\int_0^\infty u^{2k} K\left(u\right)^2 du + o\left(1\right)\right) \end{aligned}$$

where the fourth equality follows from a change of variables $u = \frac{r-c}{h}$ and the fifth equality from the dominated convergence theorem.

Let us now turn to the second term and note that

$$E\left[\frac{1}{nh}\sum_{i=1}^{n} K_+\left(\frac{R_i - c}{h}\right) X_{pi} U_i \mid X_{pi}\right] = \frac{1}{nh}\sum_{i=1}^{n} K_+\left(\frac{R_i - c}{h}\right) X_{pi} \left(\frac{m_+^{(p+1)}\left(c\right)}{(p+1)!}\left(R_i - c\right)^{p+1} + \xi\left(R_i\right)\right).$$

I will only consider the variance of

$$A_{k,n} = \frac{1}{nh}\sum_{i=1}^{n} K_+\left(\frac{R_i - c}{h}\right)\left(\frac{R_i - c}{h}\right)^k \left(\frac{m_+^{(p+1)}\left(c\right)}{(p+1)!}\left(R_i - c\right)^{p+1} + \xi\left(R_i\right)\right)$$

as the covariance terms can be handled in a similar fashion. For this we get

$$
\begin{aligned}
Var\left[A_{k,n}\right] &\leq \frac{1}{nh^2}E\left[K_+\left(\frac{R_i-c}{h}\right)^2\left(\frac{R_i-c}{h}\right)^{2k}\left(\frac{m_+^{(p+1)}(c)}{(p+1)!}(R_i-c)^{p+1}+\xi(R_i)\right)^2\right] \\
&\leq \frac{O(1)}{nh^2}E\left[K_+\left(\frac{R_i-c}{h}\right)^2\left(\frac{R_i-c}{h}\right)^{2k}\left(\left(\frac{m_+^{(p+1)}(c)}{(p+1)!}\right)^2(R_i-c)^{2(p+1)}+\xi(R_i)^2\right)\right] \\
&= \frac{O(1)}{nh^2}h^{2(p+1)}\int_c^\infty K\left(\frac{r-c}{h}\right)^2\left(\frac{r-c}{h}\right)^{2(k+p+1)}f(r)\,dr \\
&\quad+\frac{O(1)}{nh^2}O\left(h^{2(p+2)}\right)\int_c^\infty K\left(\frac{r-c}{h}\right)^2\left(\frac{r-c}{h}\right)^{2k}f(r)\,dr \\
&= \frac{O(1)}{nh}h^{2(p+1)}\int_0^\infty K(u)^2u^{2(k+p+1)}f(c+hu)\,du \\
&\quad+\frac{O(1)}{nh}O\left(h^{2(p+2)}\right)\int_0^\infty K(u)^2u^{2k}f(c+hu)\,du \\
&= \frac{1}{nh}O\left(h^{2(p+1)}\right)
\end{aligned}
$$

where the fourth equality follows from a change of variables $u=\frac{r-c}{h}$ and the fifth equality from the dominated convergence theorem. □

Lemma 4

$$
E\begin{bmatrix}\hat{\alpha}_p^+(h)-\alpha^+ \\ h\left(\hat{\beta}_{1,p}(h)-\beta_1^+\right) \\ \vdots \\ h^p\left(\hat{\beta}_{p,p}(h)-\beta_p^+\right)\end{bmatrix} = \frac{m_+^{(p+1)}(c)}{(p+1)!}\Gamma_+^{-1}\delta_+h^{p+1}+o_p\left(h^{p+1}\right)
$$

$$
Var\begin{bmatrix}\hat{\alpha}_p^+(h)-\alpha^+ \\ h\left(\hat{\beta}_{1,p}(h)-\beta_1^+\right) \\ \vdots \\ h^p\left(\hat{\beta}_{p,p}(h)-\beta_p^+\right)\end{bmatrix} = \frac{1}{nh}\frac{\sigma_+^2(c)}{f(c)}\Gamma_+^{-1}\Lambda_+\Gamma_+^{-1}+o_p\left(\frac{1}{nh}\right)
$$

Proof: The result follows from Lemma 1, Lemma 2 and Lemma 3 using the continuous mapping theorem. □

Lemma 5

$$\frac{1}{nh_n}\sum_{i=1}^{n} K_{+}\left(\frac{R_i - c}{h_n}\right)^2 U_i^2 X_{pi} X_{pi}^{'} \quad = \quad \sigma_{+}^2(c) f(c)\Lambda_{+} + O\left(h^{2(p+1)}\right) + o_p(1)$$

where

$$\Lambda_{+} \quad = \quad \begin{bmatrix} \lambda_0 & \lambda_1 & \cdots & \lambda_p \\ \lambda_1 & \lambda_2 & \cdots & \lambda_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_p & \lambda_{p+1} & \cdots & \lambda_{2p} \end{bmatrix}$$

and

$$\lambda_k \quad = \quad \int_0^{\infty} u^k K(u)^2 \, du.$$

Proof: Let us start by writing

$$\frac{1}{nh_n}\sum_{i=1}^{n} K_{+}\left(\frac{R_i - c}{h_n}\right)^2 U_i^2 X_{pi} X_{pi}^{'} \quad = \quad \begin{bmatrix} A_{0,n} & \cdots & A_{p,n} \\ \vdots & \ddots & \vdots \\ A_{p,n} & \cdots & A_{2p,n} \end{bmatrix}$$

where

$$\begin{aligned} A_{k,n} \quad &= \quad \frac{1}{nh_n}\sum_{i=1}^{n} K_{+}\left(\frac{R_i - c}{h_n}\right)^2 \left(\frac{R_i - c}{h_n}\right)^k U_i^2 \\ &= \quad \frac{1}{nh}\sum_{i=1}^{n} K_{+}\left(\frac{R_i - c}{h}\right)^2 \left(\frac{R_i - c}{h}\right)^k \left(\frac{m_{+}^{(p+1)}(c)}{(p+1)!}(R_i - c)^{p+1} + \xi(R_i) + \epsilon_i\right)^2 . \end{aligned}$$

The expectation of $A_{k,n}$ can be written as

$$
\begin{aligned}
E\left[A_{k,n}\right] &= \frac{1}{h}E\left[K_{+}\left(\frac{R_i-c}{h}\right)^2\left(\frac{R_i-c}{h}\right)^k\left(\frac{m_{+}^{(p+1)}(c)}{(p+1)!}\left(R_i-c\right)^{p+1}+\xi\left(R_i\right)+\epsilon_i\right)^2\right] \\
&= \frac{1}{h}E\left[K_{+}\left(\frac{R_i-c}{h_n}\right)^2\left(\frac{R_i-c}{h_n}\right)^k\sigma^2\left(R_i\right)\right]+R_n \\
&= \frac{1}{h}\int_c^{\infty}K\left(\frac{r-c}{h}\right)^2\left(\frac{r-c}{h}\right)^k\sigma^2(r)\,f(r)\,dr+R_n \\
&= \int_0^{\infty}K(u)^2u^k\sigma^2(c+hu)\,f(c+hu)\,du+R_n \\
&= \sigma_{+}^2(c)\,f(c)\int_0^{\infty}u^kK(u)^2\,du+o(1)+R_n
\end{aligned}
$$

where the fourth equality from a change of variables $u=\frac{r-c}{h}$ and the fifth equality from the dominated convergence theorem. Furthermore, notice that

$$
\begin{aligned}
R_n &= \frac{1}{h}E\left[K_{+}\left(\frac{R_i-c}{h}\right)^2\left(\frac{R_i-c}{h_n}\right)^k\left(\frac{m_{+}^{(p+1)}(c)}{(p+1)!}\left(R_i-c\right)^{p+1}+\xi\left(R_i\right)\right)^2\right] \\
&\leq O(1)\,\frac{1}{h}E\left[K_{+}\left(\frac{R_i-c}{h}\right)^2\left(\frac{R_i-c}{h}\right)^k\left(\left(\frac{m_{+}^{(p+1)}(c)}{(p+1)!}\right)^2\left(R_i-c\right)^{2(p+1)}+\xi\left(R_i\right)^2\right)\right] \\
&= O(1)\,\frac{1}{h}\int_c^{\infty}K\left(\frac{r-c}{h}\right)^2\left(\frac{r-c}{h}\right)^k\left(R_i-c\right)^{2(p+1)}f(r)\,dr \\
&\quad +O(1)\,\frac{1}{h}\int_0^{\infty}K\left(\frac{r-c}{h}\right)^2\left(\frac{r-c}{h}\right)^k\xi\left(R_i\right)^2f(r)\,dr \\
&= O(1)\,h^{2(p+1)}\int_0^{\infty}K(u)^2u^{k+2(p+1)}f(c+hu)\,du \\
&\quad +O(1)\,O\left(h^{2(p+2)}\right)\int_0^{\infty}K(u)^2u^kf(c+hu)\,du \\
&= O\left(h^{2(p+1)}\right)
\end{aligned}
$$

where the first inequality follows from the c_r inequality and the third equality from a change of variables $u=\frac{r-c}{h}$.

For the variance of $A_{k,n}$ we get

$$\begin{aligned}
Var\left[A_{k,n}\right] &\leq \frac{1}{nh^2}E\left[K_+\left(\frac{R_i-c}{h}\right)^4\left(\frac{R_i-c}{h}\right)^{2k}\left(\frac{m_+^{(p+1)}(c)}{(p+1)!}(R_i-c)^{p+1}+\xi(R_i)+\epsilon_i\right)^4\right] \\
&\leq \frac{1}{nh^2}O(1)\,E\left[K_+\left(\frac{R_i-c}{h}\right)^4\left(\frac{R_i-c}{h}\right)^{2k}\left(\left(\frac{m_+^{(p+1)}(c)}{(p+1)!}\right)^4(R_i-c)^{4(p+1)}+\xi(R_i)^4+\epsilon_i^4\right)\right] \\
&= \frac{1}{nh^2}O(1)\int_c^\infty K\left(\frac{r-c}{h}\right)^4\left(\frac{r-c}{h}\right)^{2k}\left(\frac{m_+^{(p+1)}(c)}{(p+1)!}\right)^4(r-c)^{4(p+1)}f(r)\,dr \\
&\quad +\frac{1}{nh^2}O(1)\int_c^\infty K\left(\frac{r-c}{h}\right)^4\left(\frac{r-c}{h}\right)^{2k}\xi(r)^4 f(r)\,dr \\
&\quad +\frac{1}{nh^2}O(1)\int_c^\infty K\left(\frac{r-c}{h}\right)^4\left(\frac{r-c}{h}\right)^{2k}E\left[\epsilon_i^4\mid R_i=r\right]f(r)\,dr \\
&= \frac{1}{nh}O(1)\,h^{4(p+1)}\int_0^\infty K(u)^4u^{2k+4(p+1)}f(c+hu)\,du \\
&\quad +\frac{1}{nh}O(1)\,o\left(h^{4(p+1)}\right)\int_0^\infty K(u)^4(u)^{2k}f(c+hu)\,du \\
&\quad +\frac{1}{nh}O(1)\int_0^\infty K(u)^4u^{2k}E\left[\epsilon_i^4\mid R_i=c+hu\right]f(c+hu)\,du \\
&= O\left(\frac{1}{nh}\right)
\end{aligned}$$

where the second inequality follows from the c_r inequality and the second equality from a change of variables $u=\frac{r-c}{h}$. □

Lemma 6

$$\frac{1}{nh_n}\sum_{i=1}^n K_+\left(\frac{R_i-c}{h_n}\right)^2\hat{U}_i^2X_{pi}X_{pi}' \quad = \tfrac{1}{nh_n}\textstyle\sum_{i=1}^n K_+\left(\frac{R_i-c}{h_n}\right)^2U_i^2X_{pi}X_{pi}' + \quad o_p(1)$$

where

$$\hat{U}_i \quad = \quad Y_i-\hat{\alpha}_p^+(h)-\sum_{k=1}^p\hat{\beta}_{p,k}^+(h)(R_i-c)^k$$

Proof: Let us start by writing

$$\begin{aligned}\hat{U}_i^2 &= \left(U_i + \hat{U}_i - U_i\right)^2 \\ &= U_i^2 + \left(\hat{U}_i - U_i\right)^2 + 2U_i\left(\hat{U}_i - U_i\right) \\ &= U_i^2 + \left(\tilde{\beta}_p^+(h)' X_{pi}\right)^2 + 2U_i\left(\tilde{\beta}_p^+(h)' X_{pi}\right)\end{aligned}$$

where $\tilde{\beta}_p^+(h) = \left[\hat{\alpha}_p^+(h) - \alpha^+, h\left(\hat{\beta}_{1,p}(h) - \beta_1^+\right), \ldots, h^p\left(\hat{\beta}_{p,p}(h) - \beta_p^+\right)\right]'$. Note that $\tilde{\beta}_p^+(h) = o_p(1)$ by Lemma 4. We can now write

$$\frac{1}{nh}\sum_{i=1}^{n} K_+\left(\frac{R_i - c}{h}\right)^2 \hat{U}_i^2 X_{pi} X_{pi}' = \begin{bmatrix} A_{0,n} & \cdots & A_{p,n} \\ \vdots & \ddots & \vdots \\ A_{p,n} & \cdots & A_{2p,n}\end{bmatrix}$$

where

$$\begin{aligned}A_{k,n} &= \frac{1}{nh}\sum_{i=1}^{n} K_+\left(\frac{R_i - c}{h}\right)^2 \left(\frac{R_i - c}{h}\right)^k \hat{U}_i^2 \\ &= \frac{1}{nh}\sum_{i=1}^{n} K_+\left(\frac{R_i - c}{h}\right)^2 \left(\frac{R_i - c}{h}\right)^k U_i^2 + R_n\end{aligned}$$

and

$$\begin{aligned}R_n &= \sum_{l=0}^{p}\frac{1}{nh}\sum_{i=1}^{n} K_+\left(\frac{R_i - c}{h}\right)^2 \left(\frac{R_i - c}{h}\right)^k \left(\tilde{\beta}_p^+(h)' X_{pi}\right)^2 \\ &\quad + O(1)\sum_{l=0}^{p}\frac{1}{nh}\sum_{i=1}^{n} K_+\left(\frac{R_i - c}{h}\right)^2 \left(\frac{R_i - c}{h}\right)^{k+2l} U_i\left(\tilde{\beta}_p^+(h)' X_{pi}\right) \\ &\leq o_p(1)\frac{1}{nh}\sum_{i=1}^{n} K_+\left(\frac{R_i - c}{h}\right)^2 \sum_{l=0}^{p}\left(\frac{R_i - c}{h}\right)^{k+2l} \\ &\quad + o_p(1)\frac{1}{nh}\sum_{i=1}^{n} K_+\left(\frac{R_i - c}{h}\right)^2 \sum_{l=0}^{p}\left(\frac{R_i - c}{h}\right)^{k+l} U_i\end{aligned}$$

where the inequality follows from the c_r inequality. The result follows from observing that

$$\begin{aligned}\frac{1}{nh}\sum_{i=1}^{n}K_{+}\left(\frac{R_i-c}{h}\right)^2\sum_{l=0}^{p}\left(\frac{R_i-c}{h}\right)^{k+2l} &= O(1)+o_p(1)\\ \frac{1}{nh}\sum_{i=1}^{n}K_{+}\left(\frac{R_i-c}{h}\right)^2\sum_{l=0}^{p}\left(\frac{R_i-c}{h}\right)^{k+l}U_i &= O(1)+o_p(1)\end{aligned}$$

which can be shown using similar steps as above. □

Lemma 7

$$E\left[v_p^{+}(h)\right] = \frac{1}{nh}\frac{\sigma_{+}^2(c)}{f(c)}\Gamma_{+}^{-1}\Lambda_{+}\Gamma_{+}^{-1}+o_p\left(\frac{1}{nh}\right)$$

Proof: The result follows from Lemma 1, Lemma 5 and Lemma 6 using the continuous mapping theorem. □

Proof of Theorem 1

From Lemma 4 and similar result for the LPR estimator using observations to the left of the cutoff we get that

$$\begin{aligned}Bias\left[\hat{\tau}_p(h)\right] &= B_p h^{p+1}+o_p\left(h^{p+1}\right)\\ Var\left[\hat{\tau}_p(h)\right] &= \frac{V_p}{nh}+o_p\left(\frac{1}{nh}\right)\end{aligned}$$

where

$$\begin{aligned}B_p &= \frac{m_{+}^{(p+1)}(c)}{(p+1)!}\Gamma_{+}^{-1}\delta_{+}-\frac{m_{-}^{(p+1)}(c)}{(p+1)!}\Gamma_{-}^{-1}\delta_{-}\\ V_p &= \frac{\sigma_{+}^2(c)}{f(c)}\Gamma_{+}^{-1}\Lambda_{+}\Gamma_{+}^{-1}-\frac{\sigma_{-}^2(c)}{f(c)}\Gamma_{-}^{-1}\Lambda_{-}\Gamma_{-}^{-1}.\end{aligned}$$

Thus, we have that

$$\begin{aligned}MSE\left[\hat{\tau}_p(h)\right] &= Bias\left[\hat{\tau}_p(h)\right]^2+Var\left[\hat{\tau}_p(h)\right]\\ &= AMSE\left[\hat{\tau}_p(h)\right]+o_p\left(h^{2(p+1)}+\frac{1}{nh}\right)\end{aligned}$$

where

$$AMSE\left[\hat{\tau}_p\left(h\right)\right] = B_p^2 h^{2(p+1)} + \frac{V_p}{nh}.$$

Finally, note that $AMSE\left[\hat{\tau}_p\left(h\right)\right]$ is globally convex in h. Thus differentiating this with respect to h and taking the first order condition gives us

$$\begin{aligned} h_{opt} &= \arg\min_h AMSE\left[\hat{\tau}_p\left(h\right)\right] \\ &= \left(\frac{V_p}{2\left(p+1\right)B_p^2}\right)^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}}. \end{aligned}$$

□

Proof of Theorem 2

From Lemma 4 and similar result for the LPR estimator using observations to the left of the cutoff as well corresponding results for a LPR-based estimator using a $(p+1)^{th}$ order polynomial we get that

$$\begin{aligned} \hat{\tau}_p\left(h\right) &= \tau + B_p h^{p+1} + o_p\left(h^{p+1}\right) + O_p\left(\left(nh\right)^{-\frac{1}{2}}\right) \\ \hat{\tau}_{p+1}\left(h\right) &= \tau + o_p\left(h^{p+1}\right) + O_p\left(\left(nh\right)^{-\frac{1}{2}}\right). \end{aligned}$$

This implies that

$$\hat{\tau}_p\left(h\right) - \hat{\tau}_{p+1}\left(h\right) = B_p h^{p+1} + o_p\left(h^{p+1}\right) + O_p\left(\left(nh\right)^{-\frac{1}{2}}\right)$$

and consequently that

$$\begin{aligned} \hat{b}_p^2\left(h\right) &= \left(\hat{\tau}_p\left(h\right) - \hat{\tau}_{p+1}\left(h\right)\right) \\ &= B_p^2 h^{2(p+1)} + o_p\left(h^{2(p+1)}\right) + O_p\left(\left(nh\right)^{-1}\right). \end{aligned}$$

In addition, from Lemma 6 and similar result for the variance estimator to the left of the cutoff we get that

$$\hat{v}_p\left(h\right) = \frac{V_p}{nh} + o_p\left(\frac{1}{nh}\right) + O_p\left(\left(nh\right)^{-\frac{3}{2}}\right).$$

Thus, by plugging in h_k we get

$$\begin{aligned}
\hat{b}_p^2\left(h_k\right) h_k^{-2(p+1)} &= B_p^2 + o_p\left(1\right) + O_p\left(n^{-(1-2p+3)\gamma}\right) \\
\hat{v}_p\left(h_k\right) n h_k &= V_p + o_p\left(1\right) + O_p\left(n^{-\frac{1}{2}(1-\gamma)}\right).
\end{aligned}$$

This implies that

$$\begin{aligned}
\hat{B}_p^2 &= \frac{\sum_{k=1}^K \hat{b}_p^2\left(h_k\right) h_k^{2(p+1)}}{\sum_{k=1}^K h_k^{4(p+1)}} \\
&= \frac{\sum_{k=1}^K \hat{b}_p^2\left(h_k\right) h_k^{-2(p+1)} h_k^{4(p+1)}}{\sum_{k=1}^K h_k^{4(p+1)}} \\
&= B_p^2 + o_p\left(1\right) + O_p\left(n^{-(1-2p+3)\gamma}\right)
\end{aligned}$$

and that

$$\begin{aligned}
\hat{V}_p &= \frac{\sum_{k=1}^K \hat{v}_p\left(h_k\right)\left(n h_k\right)^{-1}}{\sum_{k=1}^K \left(n h_k\right)^{-2}} \\
&= \frac{\sum_{k=1}^K \hat{v}_p\left(h_k\right) n h_k \left(n h_k\right)^{-2}}{\sum_{k=1}^K \left(n h_k\right)^{-2}} \\
&= V_p + o_p\left(1\right) + O_p\left(n^{-\frac{1}{2}(1-\gamma)}\right).
\end{aligned}$$

Furthermore, plugging in the estimators $\hat{B}_p$ and $\hat{V}_p$ to the expression for C_{opt}, we get

$$\begin{aligned}
\hat{C}_{opt} &= \left(\frac{\hat{V}_p}{2\left(p+1\right)\hat{B}_p^2}\right)^{\frac{1}{2p+3}} \\
&= \left(\frac{V_p}{2\left(p+1\right)B_p^2}\right)^{\frac{1}{2p+3}} + o_p\left(1\right) \\
&= C_{opt} + o_p\left(1\right)
\end{aligned}$$

and consequently that

$$\begin{aligned}
\frac{\hat{h}_{opt}}{h_{opt}} &= \frac{\hat{C}_{opt} n^{-\frac{1}{2p+3}}}{C_{opt} n^{-\frac{1}{2p+3}}} \\
&= \frac{\hat{C}_{opt}}{C_{opt}} \\
&= 1 + o_p\left(1\right).
\end{aligned}$$

Finally, note that

$$\begin{aligned} MSE\left[\hat{\tau}_p\left(h_{opt}\right)\right] &= AMSE\left[\hat{\tau}_p\left(h_{opt}\right)\right] + o_p\left(h_{opt}^{2(p+1)} + \frac{1}{nh_{opt}}\right) \\ &= n^{-\frac{2(p+1)}{2[p+3}}\left(B_p C_{opt}^{2(p+1)} + V_p C_{opt}^{-1} + o_p\left(1\right)\right) \end{aligned}$$

and that

$$\begin{aligned} MSE\left[\hat{\tau}_p\left(\hat{h}_{opt}\right)\right] &= AMSE\left[\hat{\tau}_p\left(\hat{h}_{opt}\right)\right] + o_p\left(\hat{h}_{opt}^{2(p+1)} + \frac{1}{n\hat{h}_{opt}}\right) \\ &= n^{-\frac{2(p+1)}{2[p+3}}\left(B_p C_{opt}^{2(p+1)} + V_p C_{opt}^{-1} + o_p\left(1\right)\right). \end{aligned}$$

Thus, we get that

$$\frac{MSE\left[\hat{\tau}_p\left(\hat{h}_{opt}\right)\right]}{MSE\left[\hat{\tau}_p\left(h_{opt}\right)\right]} = 1 + o_p\left(1\right)$$

which implies that

$$\hat{\tau}_p\left(\hat{h}_{opt}\right) = \tau + O_p\left(n^{-\frac{p+1}{2[p+3}}\right).$$

□

Table 1: Monte Carlo Simulations for Design 1

		$\hat{h}$		$\hat{\tau}$		
		Mean	SE	Bias	SE	RMSE
N = 100	IK	0.5637	0.1318	0.0335	0.0816	0.0882
	Adaptive	0.3302	0.0926	0.0294	0.1617	0.1643
N = 500	IK	0.4739	0.0585	0.0432	0.0359	0.0561
	Adaptive	0.2814	0.0751	0.0274	0.0525	0.0592
N = 1,000	IK	0.4161	0.0468	0.0423	0.0245	0.0489
	Adaptive	0.2477	0.0704	0.0217	0.0397	0.0453
N = 5,000	IK	0.3399	0.0337	0.0385	0.0110	0.0400
	Adaptive	0.1671	0.0443	0.0136	0.0210	0.0250
N = 10,000	IK	0.3311	0.0266	0.0380	0.0086	0.0390
	Adaptive	0.1379	0.0337	0.0107	0.0166	0.0197
N = 50,000	IK	0.1988	0.0184	0.0223	0.0070	0.0234
	Adaptive	0.0883	0.0118	0.0054	0.0083	0.0099

Table 2: Monte Carlo Simulations for Design 2

		$\hat{h}$		$\hat{\tau}$		
		Mean	SE	Bias	SE	RMSE
N = 100	IK	0.5581	0.1535	0.0287	0.0926	0.0969
	Adaptive	0.3317	0.0970	0.0045	0.1635	0.1636
N = 500	IK	0.4189	0.0712	0.0087	0.0369	0.0379
	Adaptive	0.3012	0.0802	0.0015	0.0506	0.0506
N = 1,000	IK	0.3643	0.0472	0.0025	0.0259	0.0261
	Adaptive	0.2699	0.0701	0.0004	0.0371	0.0371
N = 5,000	IK	0.2624	0.0203	0.0016	0.0137	0.0138
	Adaptive	0.2191	0.0566	0.0012	0.0176	0.0176
N = 10,000	IK	0.2285	0.0167	0.0017	0.0109	0.0111
	Adaptive	0.1980	0.0507	0.0009	0.0135	0.0135
N = 50,000	IK	0.1661	0.0089	0.0014	0.0054	0.0055
	Adaptive	0.1572	0.0410	0.0008	0.0063	0.0064

Table 3: Monte Carlo Simulations for Design 3

		$\hat{h}$		$\hat{\tau}$		
		Mean	SE	Bias	SE	RMSE
N = 100	IK	0.2289	0.0254	0.0234	0.1289	0.1310
	Adaptive	0.2105	0.0396	0.0169	0.1770	0.1778
N = 500	IK	0.1746	0.0160	0.0078	0.0582	0.0587
	Adaptive	0.1802	0.0263	0.0067	0.0581	0.0585
N = 1,000	IK	0.1563	0.0143	0.0057	0.0442	0.0445
	Adaptive	0.1643	0.0229	0.0049	0.0439	0.0442
N = 5,000	IK	0.1226	0.0107	0.0030	0.0210	0.0212
	Adaptive	0.1327	0.0182	0.0031	0.0209	0.0211
N = 10,000	IK	0.1106	0.0099	0.0021	0.0160	0.0162
	Adaptive	0.1209	0.0163	0.0022	0.0158	0.0159
N = 50,000	IK	0.0877	0.0077	0.0011	0.0076	0.0077
	Adaptive	0.0970	0.0134	0.0012	0.0075	0.0076

Table 4: Monte Carlo Simulations for Design 4

		$\hat{h}$		$\hat{\tau}$		
		Mean	SE	Bias	SE	RMSE
N = 100	IK	0.2238	0.0251	0.0192	0.1315	0.1329
	Adaptive	0.2157	0.0399	0.0155	0.1774	0.1781
N = 500	IK	0.1735	0.0165	0.0063	0.0581	0.0584
	Adaptive	0.1873	0.0283	0.0060	0.0573	0.0576
N = 1,000	IK	0.1556	0.0144	0.0047	0.0441	0.0444
	Adaptive	0.1710	0.0250	0.0044	0.0432	0.0435
N = 5,000	IK	0.1225	0.0108	0.0026	0.0210	0.0211
	Adaptive	0.1386	0.0208	0.0029	0.0205	0.0207
N = 10,000	IK	0.1105	0.0100	0.0018	0.0160	0.0161
	Adaptive	0.1262	0.0186	0.0021	0.0156	0.0157
N = 50,000	IK	0.0877	0.0077	0.0010	0.0076	0.0077
	Adaptive	0.1015	0.0154	0.0011	0.0074	0.0075