9-1-2014

Theory and practice of inference in regression discontinuity: a fixed-bandwidth asymptotics approach

Otávio Bartalotti
Iowa State University, bartalot@iastate.edu

Theory and practice of inference in regression discontinuity: a fixed-bandwidth asymptotics approach

Abstract
In regression discontinuity design (RD), researchers use bandwidths around the discontinuity. For agiven bandwidth, one can estimate asymptotic variance based on the assumption that the bandwidthshrinks to zero as sample size increases (the traditional approach) or, alternatively, that the bandwidthis Â…xed. The main theoretical results for RD rely on the former, while most applications in the literaturetreat the estimates as parametric. This paper develops the Â“Â…xed-bandwidthÂ” alternative asymptotictheory for RD designs, bridging the gap between theorists and practitioners while shedding light onimplicit assumptions in both approaches. The Â…xed-bandwidth approach provides alternative formulas(approximations) for the bias and variance of RD estimators. Simulations indicate that Â…xed-bandwidthapproximations are usually better than traditional approximations, and improvements are nontrivialwhen there is heteroskedasticity. When there is no heteroskedasticity, both approximations are shown tobe equivalent, under some additional mild conditions. Feasible estimators of Â…xed-bandwidth standarderrors are easy to implement and improve coverage of conÂ…dence intervals compared to the traditionalapproach, especially in the presence of heteroskedasticity. Fixed-bandwidth approximations are akin totreating RD estimators as locally parametric, providing theoretical justiÂ…cation for the common empiricalpractice of using heteroskedasticity-robust standard errors in RD settings.

Keywords
average treatment effect, locally parametric inference, local polynomial estimators, fixed bandwidth

Disciplines
Economics

Theory and Practice of Inference in Regression Discontinuity: A Fixed-Bandwidth Asymptotics Approach

Otavio Bartalotti

Working Paper No. 14025
September 2014

IOWA STATE UNIVERSITY
Department of Economics
Ames, Iowa, 50011-1070

Theory and Practice of Inference in Regression Discontinuity: A Fixed-Bandwidth Asymptotics Approach[*,†]

Otávio Bartalotti[‡]

Iowa State University

This Draft: August 2014

First Draft: June 2011

Abstract

In regression discontinuity design (RD), researchers use bandwidths around the discontinuity. For a given bandwidth, one can estimate asymptotic variance based on the assumption that the bandwidth shrinks to zero as sample size increases (the traditional approach) or, alternatively, that the bandwidth is fixed. The main theoretical results for RD rely on the former, while most applications in the literature treat the estimates as parametric. This paper develops the "fixed-bandwidth" alternative asymptotic theory for RD designs, bridging the gap between theorists and practitioners while shedding light on implicit assumptions in both approaches. The fixed-bandwidth approach provides alternative formulas (approximations) for the bias and variance of RD estimators. Simulations indicate that fixed-bandwidth approximations are usually better than traditional approximations, and improvements are nontrivial when there is heteroskedasticity. When there is no heteroskedasticity, both approximations are shown to be equivalent, under some additional mild conditions. Feasible estimators of fixed-bandwidth standard errors are easy to implement and improve coverage of confidence intervals compared to the traditional approach, especially in the presence of heteroskedasticity. Fixed-bandwidth approximations are akin to treating RD estimators as *locally* parametric, providing theoretical justification for the common empirical practice of using heteroskedasticity-robust standard errors in RD settings.

[*]JEL: C12, C21; Keywords: Average Treatment Effect, Locally Parametric Inference, Local Polynomial Estimators, Fixed Bandwidth.

[†]I would like to thank Jeff Wooldridge, Tim Vogelsang, Gary Solon, Matias Cattaneo, Peter Schmidt, Steve Dieterle, Quentin Brummet, Valentin Verdier, Ilya Rahkovsky, Keith Finlay, Michael Darden, Brigham Frandsen and participants at several seminars and conferences where earlier versions of this paper were presented for valuable comments and discussions.

[‡]279 Heady Hall, Department of Economics, Iowa State University, Ames, IA, 50011. E-mail: bartalot@iastate.edu. Ph: +1 (515) 294-1085.

1 Introduction

Regression discontinuity (RD) designs have been propelled to the spotlight of economic analysis in recent years,[1] especially in 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.

Regarding estimation of the parameters, local polynomial estimators are the most common choice in empirical and theoretical work due to ease of implementation. 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. Implementation of the local polynomial estimators depends on the choice of the bandwidth, which can be aided by several bandwidth selectors available in the literature.

From the perspectives of both identification and estimation, the present paper follows this standard framework, relying on the same fairly weak nonparametric conditions in the literature to justify the validity of RD designs. However, when it comes to the issue of inference, improvements can be achieved by considering an alternative asymptotic approximation to the estimator's behavior. Interestingly, the alternative proposed here also reconciles current practice by researchers and theory.

The conventional "small-h" approach to obtaining the asymptotic properties of estimators in RD designs relies on the idea that the bandwidth shrinks towards zero asymptotically. In practice, implementation requires the researcher to use a particular bandwidth, which is necessarily greater than zero. One example is when the running variable is discrete, requiring a discretely positive bandwidth. Alternatively, even with a continuous running variable, sample sizes often are small enough that concerns about precision impel researchers to use a relatively large bandwidth. Hence, even though asymptotic approximations rely on $h \to 0$, in practice h is fixed.[2]

From that fact, there arises a disconnect between theory and practice for inference in RD since most practitioners use the standard parametric inference methods, even though asymptotic theory for RD is based on small-h asymptotic approximations.

It is crucial to separate the distinct roles of bandwidths and limiting arguments in the identification and inference procedures in RD. As pointed out, identification is nonparametric, and the estimators in this paper are identical to the standard local polynomial estimators in the literature.

This paper focuses on inference, presenting an alternative asymptotic approximation for the standard RD treatment effects estimator, using a fixed bandwidth for that end. This "fixed-h" approximation is intended, primarily, to provide expressions for the estimator's asymptotic variance that better incorporate

[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] 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."

the bandwidth used by the research, to refine the conventional approximation, and to lead naturally to the standard error formulas. Additionally, an alternative, fixed-h, approximation for the asymptotic bias is obtained and provides additional intuition on the robustness-precision trade-off facing the researcher in RD designs. This approach is akin to treating the model as parametric in the neighborhood of the cutoff.

The intuition fits nicely with the current practice in applied work, which has focused on the usual Huber–Eicker–White heteroskedasticity robust standard errors, essentially treating the estimates as locally parametric. Hence, this paper provides a theoretical framework that justifies such a choice, bridging the gap between theory and practice regarding inference in RD. It also provides evidence that asymptotic variance approximations that treat the estimators as locally parametric can improve inference and successfully adjust standard errors to reflect the bandwidth used by researchers.

Comparing fixed-h and small-h approximations provides additional clarity on the assumptions and on how the improvements in inference take place. Notably, the expressions for small-h standard errors implicitly impose homoskedasticity and constant probability density of the running variable around the discontinuity. The intuition is similar to the one put forth by Fan and Gijbels (1996) and Lee and Card (2008) for the case of discrete running variables. When a bandwidth is used, estimating the conditional expectation of the outcome at the threshold is akin to estimating a parametric polynomial model inside the bandwidth. For larger bandwidths, these parametric assumptions become more restrictive as we impose functional form and homoskedasticity to a larger support of the data, which can significantly affect inference performance.

Natural estimators for the asymptotic variance based on fixed-h results are presented, and Monte Carlo simulations provide evidence that feasible inference incorporates the improvements predicted by the theory. As expected, these improvements are more important when the bandwidth is larger and local heteroskedasticity is present. The variance estimators are analogous to the Huber–Eicker–White heteroskedasticity robust standard errors. 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 theoretical asymptotic approximations—see, for example, Lee and Lemieux (2009) and Calonico et al. (2014), among others. Finally, I provide an empirical application using Lee (2008), exemplifying with actual data the improvements obtained.

This paper provides theoretical justification for using these alternative variance estimators in the nonparametric context, a practice common in empirical applications. To be clear, the results present how to perform inference appropriately for any bandwidth, hence providing variance estimators which are "robust to the choice of bandwidths," not how to choose a bandwidth. These results validate current practice in applied work and bridge the gap between theory and practice, clarifying the assumptions necessary for the validity of this approach.

1.1 Brief Review of the Literature

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 alternative asymptotic approximations for the bias-corrected local polynomial RD estimator. Other studies about estimation, inference, and bandwidth choice in RD designs include Imbens and Kalyanaraman (2012) and Cattaneo et al. (2012), 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 proposed inferential framework using fixed bandwidths follows a growing literature that recognizes the potential for 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 by acknowledging that, with a finite sample, the bandwidth used is fixed. The author argued that the assumption equivalent to the bandwidth converging to zero "is a convenient assumption mathematically in that, in particular, it ensures consistency of the estimates, but it is unrealistic when such results are used as approximations to the finite case..." (Neave 1970, p.70). Neave's work was later extended by Hashimzade and Vogelsang (2008). Fan (1998) provided an alternative approximation for goodness-of-fit tests for density function estimates in which the bandwidth used in the test is fixed, obtaining improved approximations to the asymptotic behavior of the test and more appropriate critical values for inference.

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 regards 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. The discussion and extensions for the fuzzy design are presented in the online appendix.

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

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

[3] As discussed in Porter (2003), Imbens and Lemieux (2008), and Lee and Lemieux (2009), 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.

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. 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 of the outcome variable. The average causal effect of the treatment at the discontinuity is (Imbens and Lemieux, 2008)

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

The sharp regression discontinuity 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).

2.1 Local Polynomial Estimator

I focus on estimates of τ obtained using local polynomial estimation, which is the most common in applied work. This prevalence is partially due to the easy implementation, nice properties, and the fact that local linear estimators have been the focus of several papers that disseminated the technique (Hahn et al. 1999 and 2001; Imbens and Lemieux 2008; Lee and Lemieux 2009).

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:

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

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

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

3 Asymptotic Distributions

To derive the asymptotic distribution of the estimator for τ, the usual regularity and smoothness conditions in the literature are sufficient. Note that the fixed-h asymptotic distributions described do not require additional assumptions over what is used in the standard, small-h literature, e.g., Hahn et al. (2001), Porter (2003), etc.

In the following, let f_o denote the marginal density of x and $m(x)$ denote the conditional expectation of y given x minus the discontinuity, i.e., $m(x) = E[y \mid x] - \alpha 1[x \geq \overline{x}]$, where $\overline{x}$ is the value of the running variable in which the discontinuity occurs. Finally, define $\varepsilon = y - E[y \mid X = x] = y - m(x) - \alpha 1[x \geq \overline{x}]$.

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

Assumption 3 *(a) $\sigma^2(x) = E\left[\varepsilon^2 \mid X = x\right]$ is continuous for $x \neq \overline{x}$, $x \in \aleph$, and right- and left-hand limits at $\overline{x}$ exist.*

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

The asymptotic distribution for the local polynomial estimator of the average treatment effect for a fixed bandwidth, h, is given by the following theorem.[4]

Theorem 1 *Suppose assumptions 1 (a) and 3 hold. Suppose assumption 2 (a) holds with $l_m \geq p+1$ and l_f as any nonnegative integer. If h is fixed, positive, and such that all points of X within the bandwidth are in $\aleph$, as $n \to \infty$, then*

$$\sqrt{nh}(\widehat{\alpha}_p - \alpha_p^*) \xrightarrow{d} N\left(0, V_{fixed-h}\right), \tag{1}$$

where

$$V_{fixed-h} = e_1' \left[\left(\Gamma_+^*\right)^{-1} \Delta_+^* \left(\Gamma_+^*\right)^{-1} + \left(\Gamma_-^*\right)^{-1} \Delta_-^* \left(\Gamma_-^*\right)^{-1} \right] e_1 \tag{2}$$

$$\alpha_p^* = \alpha + B_{fixed-h}$$

$$B_{fixed-h} = e_1' \left\{ \begin{array}{c} \left(\Gamma_+^*\right)^{-1} \left[\int_0^\infty k(u) Z(\overline{x}+uh) m(\overline{x}+uh) f_o(\overline{x}+uh) du\right] - \\ - \left(\Gamma_-^*\right)^{-1} \left[\int_0^\infty k(u) Z(\overline{x}-uh) m(\overline{x}-uh) f_o(\overline{x}-uh) du\right] \end{array} \right\}, \tag{3}$$

and

$$\Gamma_{+(-)}^* = \begin{bmatrix} \gamma_0^{+(-)} & \cdots & \gamma_p^{+(-)} \\ \vdots & \ddots & \vdots \\ \gamma_p^{+(-)} & \cdots & \gamma_{2p}^{+(-)} \end{bmatrix}, \quad \Delta_{+(-)}^* = \begin{bmatrix} \delta_0^{+(-)} & \cdots & \delta_p^{+(-)} \\ \vdots & \ddots & \vdots \\ \delta_p^{+(-)} & \cdots & \delta_{2p}^{+(-)} \end{bmatrix},$$

$e_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}'$; $\gamma_j^+ = \int_0^\infty k(u) u^j f_o(\overline{x}+uh) du$; $\gamma_j^- = (-1)^j \int_0^\infty k(u) u^j f_o(\overline{x}-uh) du$;
$\delta_j^+ = \int_0^\infty k^2(u) u^j \sigma^2(\overline{x}+uh) f_o(\overline{x}+uh) du$; $\delta_j^- = (-1)^j \int_0^\infty k^2(u) u^j \sigma^2(\overline{x}-uh) f_o(\overline{x}-uh) du$.

[4] The proofs are collected in the online appendix.

The fixed-h approach used in theorem 1 explicitly takes into consideration the bandwidth used, without assuming $h \to 0$. It also captures the impact of h on the asymptotic variance, $V_{fixed-h}$, accounting for potential heteroskedasticity inside the bandwidth, which would be ignored by the conventional approximation, as will be shown in corollary 2.

As is well known, unless the true specification of the population model is known, the local polynomial estimator is biased in finite samples. 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. That correct specification depends directly on the bandwidth used, thus shedding light on the implicit parametric restriction that practitioners face when implementing RD.[5]

The bias is the difference of the (scaled) linear projection of $m(x)$ on Z, evaluated at $x = \overline{x}$ (i.e., the difference in intercepts) inside the bandwidth, both above and below the cutoff. Intuitively, the bias as highlighted in $B_{fixed-h}$ has two sources. The first, due to the potential lack of validity of the RD design, is the difference between the conditional expectation of the outcome above and below the cutoff that would have arisen in the absence of treatment, i.e., the difference that would have happened nevertheless and is erroneously attributed to the treatment or policy being analyzed. This potential bias is well known and usually the subject of careful efforts by practitioners to justify their identification strategy.

The second source of bias is due to the potential local model misspecification since Z might not be able to correctly capture $m(x)$'s features within the bandwidth. This potential source of bias is in general ignored by small-h approximations, and can therefore be misleading in RD applications. To clarify the working assumption practitioners impose when implementing RD, it is interesting to draw a parallel of the results in theorem 1 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. In this sense, the local polynomial estimator is a polynomial approximation to the unknown conditional expectation inside the bandwidth on each side, not different from standard parametric methods.[6] By using a relatively small bandwidth, we are fitting 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.

Hence, if one assumes that the conditional expectation is parametric and the polynomial of order p above correctly specifies the model in a certain window around the cutoff, the estimator will be asymptotically

[5] Developing a bias correction procedure is beyond the scope of this work. Recently Calonico et al. (2014) provided bias corrected approximations that incorporate the bias variability to the inference procedure in a small-h context. It seems that a natural extension of the results above would be to incorporate the Calonico et al. (2014) bias-correction to the framework presented here.

[6] As emphasized by Fan and Gijbels (1996), an advantage to using local polynomial approximations is the ability to rely on least square principles and have access to the statistical knowledge and generalizations connected to least square regression.

unbiased. This is the implicit assumption on the functional form being imposed by practitioners when using RD.

The fixed-h asymptotic approximation in theorem 1 bears a close connection to the small-h approximation in Porter (2003). Porter's relevant approximation for the local polynomial estimator is stated below so that the connection between the approximations can be analyzed.[7]

Theorem 2 *(Porter 2003, theorem 3(a)) Suppose assumptions 1 (a) and 3 hold. If assumption 2 (a) holds with* $l_m \geq p+1$ *and* l_f *as any nonnegative integer,* $nh \to \infty$, $h^{p+1}\sqrt{nh} \to C_a$, *where* $0 \leq C_a < \infty$, *then*

$$\sqrt{nh}(\widehat{\alpha}_p - \alpha^*_{small-h}) \xrightarrow{d} N\left(0, V_{small-h}\right),$$

where

$$V_{small-h} = \frac{\sigma^{2+}(\overline{x}) + \sigma^{2-}(\overline{x})}{f_o(\overline{x})} e_1'\Gamma^{-1}\Delta\Gamma^{-1}e_1 \tag{4}$$

$$\alpha^*_{small-h} = \alpha + B_{small-h}$$

$$B_{small-h} = \frac{h^{p+1}}{(p+1)!}\left[m^{(p+1)+}(\overline{x}) - (-1)^{p+1}m^{(p+1)-}(\overline{x})\right] e_1'\Gamma^{-1}\begin{bmatrix}\gamma_{p+1}\\ \vdots \\ \gamma_{2p+1}\end{bmatrix} \tag{5}$$

and

$$\Gamma = \begin{bmatrix}\gamma_0 & \cdots & \gamma_p\\ \vdots & \ddots & \vdots\\ \gamma_p & \cdots & \gamma_{2p}\end{bmatrix}, \Delta = \begin{bmatrix}\delta_0 & \cdots & \delta_p\\ \vdots & \ddots & \vdots\\ \delta_p & \cdots & \delta_{2p}\end{bmatrix},$$

$e_1 = \begin{bmatrix}1 & 0 & \cdots & 0\end{bmatrix}'$, $\gamma_j = \int_0^\infty k(u)u^j du$, $\delta_j = \int_0^\infty k^2(u)u^j du$, *and* $m^{(l)+(-)}(x)$ *is the* l^{th} *right- (left)-hand derivative of* $m(x)$ *at point* x.

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, the bias present in the estimate is given. Small-h asymptotics take advantage of a first order approximation of the asymptotic bias term to sustain the argument that the bias vanishes as the bandwidth shrinks; however, it is important to keep in mind that, in practice, the bias will still be present, even when the bandwidth was chosen to "undersmooth." For a discussion on bias correction and inference in the small-h framework, see Calonico et al. (2014).

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 of $\widehat{\alpha}$ in theorem 1 approach the asymptotic variance and bias of the small-h approximation in theorem 2.[8]

[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.

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

Corollary 1

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

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, $f_o(x)$ and $\sigma^2(x)$ 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.*

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

Corollary 2 makes clear that the refinements obtained by fixed-h approximation are due to incorporating the behavior of $f_o(x)$ and $\sigma^2(x)$ in the ranges around the cutoff, while small-h considers only its values at the cutoff, $f_o(\overline{x})$ and $\sigma^2(\overline{x})$. Hence, heteroskedasticity inside the bandwidth could lead to poor performance by the small-h variance approximation relative to that of fixed-h.[9]

4 Variance Estimators

To perform inference about α, appropriate estimates for the asymptotic variance formulas from theorem 1 are necessary. The components of the asymptotic variance of $\sqrt{nh}(\widehat{\alpha}_p - \alpha_p^*)$ can be written as

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

and similarly for γ_j^- and δ_j^-. A natural estimator for the asymptotic variance is given by their sample analogues,

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

[9] If the conditions in corollary 2 hold, the fixed-h and small-h approximations are the same; however, they could still differ in practice because they suggest different formulas for standard errors, as discussed in section 4.

which are consistent by standard arguments.

The 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 Huber–Eicker–White heteroskedasticity robust standard errors in a general weighted least squares framework, and it comes naturally from the fixed-h approximation. The use of such variance estimators in RD designs have been suggested based on intuitive or "pre-asymptotics" arguments that do not follow directly from the asymptotic approximations—see, for example, Imbens and Lemieux (2008), Lee and Lemieux (2009), and Calonico et al. (2014). The fixed-h approach provides a theoretical framework that justifies the use of such estimators by practitioners. When we treat the estimators as locally parametric, our variance estimators become "robust to the choice of bandwidths." This approach directly takes into consideration the impact of higher order polynomials on the estimator's variance and is flexible regarding the conditional variance and density of X around the cutoff. Note that,

$$\left[\left(\widehat{\Gamma}_{+}^{*}\right)^{-1}\widehat{\Delta}_{+}^{*}\left(\widehat{\Gamma}_{+}^{*}\right)^{-1}+\left(\widehat{\Gamma}_{-}^{*}\right)^{-1}\widehat{\Delta}_{-}^{*}\left(\widehat{\Gamma}_{-}^{*}\right)^{-1}\right]\xrightarrow{p}\left[\left(\Gamma_{+}^{*}\right)^{-1}\Delta_{+}^{*}\left(\Gamma_{+}^{*}\right)^{-1}+\left(\Gamma_{-}^{*}\right)^{-1}\Delta_{-}^{*}\left(\Gamma_{-}^{*}\right)^{-1}\right]=V_{fixed-h}$$

by standard asymptotic arguments, since each term is just the sample analogue of a population expectation and the law of large numbers holds.

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. This reinforces the intuition that, for a given bandwidth used, adequate inference can be obtained by dealing with the problem as if it is locally parametric. This intuition fits nicely with findings in Lee and Card (2008) that parametric assumptions are needed when discreteness is present in the running variable.

Variance estimators based on small-h asymptotics, 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 asymptotic variance matrix.

The matrix $\Gamma^{-1}\Delta\Gamma^{-1}$ can be calculated directly because it is a deterministic function of the kernel. An additional drawback of the variance estimator in formula (10) is the need to estimate $f_o(\overline{x})$, which is sidestepped in the fixed-h variance estimator in formula (6). 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.

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).

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. All simulations are based on a sharp RD design. The objective of the simulations is to evaluate the relative performance of tests based on the asymptotic variances obtained by fixed-h and small-h approximations. As shown in the previous results, it is expected that both approaches should 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, the focus in this paper is restricted to data generating processes for which the local linear estimator will have no or mild asymptotic bias. Obviously, if the bias in the local linear estimator were important, the inference on both approaches would suffer equally, since they use the same estimator, and would not allow an adequate comparison of the validity of both variance approximations. The fixed-h asymptotic bias approximation, even though more descriptive of the potential bias term, is not feasible; for a bias-correction procedure that relies on the small-h first order approximation of the bias term, see Calonico et al. (2014).

Evidence from the simulations presented below indicates that, both in the theoretical (unfeasible) and feasible cases, inference using the fixed-h approximation has better size behavior than the small-h approach, especially for larger bandwidths and when local heteroskedasticity is present.

The simulations consist of 2,000 replications with sample size n, equal to 750 observations; the actual number of observations included can be significantly smaller, depending on the bandwidth used.[10] For models 1 through 4, the running variable, X, is drawn from an $N(50, 100)$, with a cutoff set arbitrarily at $\overline{x} = 55$. The error term, u, is drawn from a normal distribution with mean 0 and standard error equal to 10 in the homoskedastic case; that value varies for different scenarios of heteroskedasticity. To exemplify the

[10] Even with the relatively large number of observations in the total sample, it is still the case that, for some samples, there are no observations inside the bandwidth for all choices of bandwidth considered. In such a case, the sample is dropped to guarantee that an estimator can be obtained.

distortions heteroskedasticity can create and how well the fixed-h asymptotic approximation can capture it, two heteroskedastic cases are considered, with the standard error for u defined as $\sigma(x) = \left(\frac{x}{17.4}\right)^2$ and $\sigma(x) = 10 + 0.25\,(x - \overline{x})^2$, respectively.[11] Finally, model 5 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 and using: a polynomial of order 5 (see below), $X \sim 2Beta(2,4) - 1$, and $\sigma(x) = 0.1295$ for the homoskedastic case and $\sigma(x) = 0.1295 + (5x)^2$ for the heteroskedastic case. Since this model introduces some relatively important bias for some bandwidths, it will allow us to compare the respective coverage obtained by both approaches in the presence of substantial bias. The bandwidths used range from 0.2 to 20, or from $\frac{1}{50}$ to 2 standard deviations of the running variable, which is well within the ranges used in most applications.

The empirical coverages presented are the fraction of rejections in the 2,000 repetitions for a test of size 5% (two-sided). The models that describe how the outcome variable is generated are given by:

- Model 1: $y_i = \mu + \beta_1 x_i + \alpha d_i + u_i$
- Model 2: $y_i = \mu + \beta_1 x_i + \beta_2 x_i^2 + \alpha d_i + u_i$
- Model 3: $y_i = \mu + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \alpha d_i + u_i$
- Model 4: $y_i = \exp\left(\frac{x}{20}\right) + \alpha d_i + u_i$
- Model 5: $$y_i = \begin{cases} 0.48 + 1.27x_i + 7.18x_i^2 + 20.21x_i^3 + 21.54x_i^4 + 7.33x_i^5 + u_i & \text{if } x < 0, \\ 0.52 + 0.84x_i - 3.00x_i^2 + 7.99x_i^3 - 9.01x_i^4 + 3.56x_i^5 + u_i & \text{if } x \geq 0. \end{cases}$$

The parameters in models 1 through 4 are $\mu = 3$, $\alpha = 10$, $\beta_1 = 0.5$, $\beta_2 = -0.005$, and $\beta_3 = 0.00002$.

The simulations use the local linear estimator ($p = 1$), since it is the preferred choice in applied work.[12]

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 section presents simulations that compare test coverages based on the theoretical fixed-h and small-h asymptotic variance formulas. While the results obtained are infeasible since they depend on knowledge of

[11] These examples aim to highlight the behavior of the fixed-h and small-h approximations in different heteroskedastic contexts. Note that all cases have the same $\sigma(\overline{x})$.

[12] 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 relationship of fixed-h inference and the standard heteroskedastic-robust standard errors. Similar results were obtained when using $p = 0, 2, 3$; Monte Carlo experiments using sample sizes equal to 500 and 1,000; and the Bartlett and truncated Gaussian kernel. They are available upon request.

$f_o(x)$ and $\sigma^2(x)$ around the cutoff, they nevertheless demonstrate the theoretical improvements provided by the fixed-h approximation.

These comparisons illustrate how conventional small-h inference, 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 hence it should not be expected to adequately describe the estimator's behavior away from the threshold.

The empirical coverage for model 1 is presented on figure 1, for which no bias is expected since the local linear estimator correctly specifies the relationship between Y and X inside any of the bandwidths used.

For smaller bandwidths, small-h and fixed-h asymptotic variances generate similar empirical coverages as expected, but there is a significant decrease in the small-h coverage as the bandwidth increases, while the fixed-h inference increasingly outperforms the standard approximation.

For the remaining models, X has a quadratic, cubic, or exponential relationship to Y. The bias is mitigated by the use of the local linear function and affects both approaches equally. The simulations focus on cases in which the bias is mild, so that the relative performance of tests based on each variance approximation can be more easily seen. Figure 2 compares the test coverages using fixed-h versus small-h standard error approximations when data is described by model 4.[13] The general pattern observed remains, with fixed-h outperforming small-h, especially for larger bandwidths.

Finally, model 5, based on data in Lee (2008), presents qualitatively similar results, as can be seen in figure 3. Note that the coverage varies severely depending on the bandwidth used due to the bias present in 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 better coverage even in the presence of bias. Furthermore, the improvements increase with bandwidth sizes, as predicted by the theory.

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, the improvements of the fixed-h approximation should be even more important.

Figures 4 and 5 present the coverages for the first (mild) heteroskedastic case using DGP models 1 and 4, while figures 6, 7, and 8 present the second (acute) heteroskedastic cases for models 1, 4, and 5, respectively.

The (infeasible) tests based on fixed-h asymptotics behave very well in both heteroskedastic cases, highlighting the robustness of the approach. In the first case, the small-h asymptotic approximation presents a slightly more pronounced pattern of decreasing coverage as the bandwidths increase, compared to the homoskedastic case; this finding is as expected, due to the effect of local heteroskedasticity. In contrast, for the acute heteroskedasticity case, the small-h-based test has a steep decline[14] in coverage as the bandwidth increases, since it is not able to properly capture the effect of the heteroskedasticity in its asymptotic variance for larger bandwidths.

[13] Models 2 and 3 provide qualitatively similar results and are omitted.

[14] Note the change in the scale of the y-axis, which now encompasses the interval from 0 to 1.

Comparing the small-h's performance in the two cases can provide useful intuition about when its weaknesses can prove most relevant. The acute heteroskedasticity situation 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 (4), 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. In the mild case, since $\sigma^2(\overline{x})$ is at the "middle" of the range for the conditional variance, the heteroskedasticity is less harmful than in the acute case.

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 asymptotic approximation can have very poor performance, while the fixed-h approach still provides a reliable asymptotic approximation for the estimator's behavior.

5.2 Simulations for Feasible Inference

As described in section 4, natural estimators for $V_{fixed-h}$ are readily available. This section presents simulations for the empirical coverage of the tests using two different estimated standard errors. The first is based on the fixed-h asymptotic distribution and is given by formula (6), which is akin to treating the estimates as locally parametric as discussed above. The second is proposed by Porter (2003) and described by formula (10).

For locally homoskedastic errors, the fixed-h standard errors' estimator incorporates the gains of improved inference as described in the theory and shown in the infeasible simulations. These results are seen in figures 9 through 11 (models 1, 4, and 5, respectively). Tests based on both approximations over-reject for very small bandwidths, due to the relatively small amount of data available in these cases. Perhaps surprisingly, in this case, tests obtained using small-h standard error estimators behave very similarly to those of fixed-h even at larger bandwidths, for which one would expect a significantly smaller coverage considering the evidence in section 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.

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.

As discussed above the presence of heteroskedasticity can generate substantial problems for the size of tests using small-h approximations.

In the mild heteroskedastic case (figures 12 and 13), the fixed-h variance estimator in formula (6) produces tests with better empirical size, which is in line with the theoretical results. Both approaches tend to over-reject for smaller bandwidths due to constrained data availability.

In figures 14, 15, and 16, where heteroskedasticity is more severe, the fixed-h variance estimator produces tests with coverage very close to the test's nominal size, while the coverage for small-h rapidly increases towards 1 as the bandwidth increases.[15]

Hence, there is evidence that heteroskedasticity can be accurately captured by tests based on fixed-h asymptotic approximations; small-h estimators, on the other hand, can produce tests with significant size distortion.

The comparison of the mild and acute heteroskedastic cases corroborates the theoretical result in theorem 1, in that the distortions caused by local heteroskedasticity in small-h's tests are more important for those patterns of heteroskedasticity in which the "weighted average" of the conditional variance on both sides of the cutoff does not approximate $\sigma^2(\overline{x})$.

Even though the empirical coverages obtained are similar when local homoskedasticity holds, it seems the fixed-h standard error estimator is 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. Using standard error estimates based on small-h asymptotics can lead to serious size distortions for larger bandwidths, especially in the presence of heteroskedasticity.

Furthermore, the fixed-h 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 exemplify this issue, figure 17 shows the simulated empirical coverages obtained by using the small-h variance estimator for model 1, with homoskedastic errors for five different scenarios. Each scenario differs by the choice of the bandwidth, h_f, used in formula (9) to obtain $\widehat{f}_o(\overline{x})$. The first reproduces the small-h result described above by choosing the same bandwidth used to estimate $\widehat{\tau}$, i.e., $h_f = h$; the other lines are the empirical coverages obtained by using a bandwidth of 1, 5, 10, and 20[16] for $\widehat{f}_o(\overline{x})$, independently of the bandwidth used for $\widehat{\tau}$.

The choice of bandwidth on the estimation of $\widehat{f}_o(\overline{x})$ can have a relevant impact on the test coverages. Choosing the same bandwidth used in estimating the parameter of interest provides more stable empirical coverages for a wide range of h relative to the cases in which the bandwidths are different.

To the empirical researcher, a useful conclusion can be drawn from these simulations. That is, by

[15] 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 since small-h does not incorporate the appropriate weights.

[16] $\frac{1}{20}$, $\frac{1}{4}$, $\frac{1}{2}$, and 1 standard deviations of the running variable, respectively.

performing inference using fixed-h standard error estimators, which is akin to treating the estimates as locally parametric and also simplifies to the standard heteroskedastic robust standard errors in the rectangular kernel case, one can feel relatively 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.[17] As pointed out in section 3, this issue similarly affects the empirical test coverage of both the fixed-h and small-h approaches. However, the fixed-h approximation has the benefit of clarifying that there are two sources of bias and 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, potentially requiring higher order polynomials of X or the inclusion of covariates.

The widespread practice in applied research to check the behavior of the estimates for different bandwidths seems like a sound practice in evaluating the impact of misspecification for different bandwidths when coupled with standard error estimators that are robust to the choice of h, such as the ones proposed here.

6 Empirical Example

To exemplify 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 .

As pointed out in Lee (2008), the U.S. Congressional electoral 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:

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

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.

[17] The small-h asymptotic bias approximation (Porter 2003) lends itself for estimation, since estimates for $m^{(p+1)}$ above and below the cutoff can be obtained. However, the results in section 3 indicate that this would be a relatively poor approximation, especially for larger h. For a detailed discussion on bias correction in RD designs and the inference adjustments needed when performing such correction, see Calonico et al. (2014).

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.[18] 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 Democrat 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 fixed-h and small-h standard errors given by formulas 6 and 10, respectively.

Table 1: 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)	(0.0114)	(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 using 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

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

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 that 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, thus 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."[19] See table 2 in the online appendix for similar results that include the pre-determined variables that Lee (2008) uses and the 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.[20] For the third entry in panel C, note that both standard errors become larger than those for the larger 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, by comparing panels along a column, note that the differences between fixed-h and small-h

[19] See Caughey and Sekhon (2011) for a compelling argument that close U.S. House election results may not be as "good as randomized."

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

estimated 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 one implied by the fixed-h approach, $\left(\widehat{\Gamma}_+^*\right)^{-1}\widehat{\Delta}_+^*\left(\widehat{\Gamma}_+^*\right)^{-1}$.

7 Conclusion

The use of RD designs to estimate the treatment effect, τ, has been widely used in recent years by researchers in economics. The standard literature on RD designs (Hahn et al. 2001; Porter 2003; Imbens and Lemieux 2008) approximates the behavior of the average treatment effect's estimator by assuming that the bandwidth around the discontinuity, h, shrinks fast enough towards zero, $h \to 0$ (i.e., small-h asymptotics), as the sample size increases. However, in practice, to obtain an estimate of the treatment effect and perform inference, the empiricist is required to use a particular value of the bandwidth that is necessarily greater than zero. Hence, a disconnect arises between theory and practice for inference in RD since most practitioners use the usual parametric asymptotic variance estimators, even though the asymptotic theory for RD is based on nonparametric, small-h asymptotic approximations.

This paper bridges the gap between theory and practice by providing a set of conditions under which the use of the usual parametric tools of inference would be locally valid; it also develops an alternative asymptotic approximation for the variance of the estimator by treating h as fixed. This fixed-h asymptotics approach explicitly acknowledges the fact that researchers must choose a bandwidth to implement the estimator, even though they are usually limited in their ability to reduce the bandwidth size because of data availability constraints. This approach is akin to treating the estimator as parametric in the neighborhood of the cutoff, and its intuition fits nicely with the current practice in applied work, which has focused on the usual Huber–Eicker–White heteroskedasticity robust standard errors—essentially treating the estimates as locally parametric. Hence, this paper provides a theoretical framework that justifies such a choice, while providing evidence that fixed-h variance estimators can improve inference and successfully adjust standard errors to reflect the choice of bandwidth by the researcher.

Fixed-h and small-h not only provide the same approximations as $h \to 0$, which one would expect, but also, interestingly, when local homoskedasticity and constant density of the running variable within the bandwidth around the cutoff are imposed. Hence, the improvements obtained by the fixed-h approximations (partially) derive from taking into account the local heteroskedasticity.

The simulations present evidence on inference using fixed-h variance estimators, which incorporate the theoretical gains of the improved approximations, are simple to implement, and simplify to the Huber–Eicker–White heteroskedasticity robust standard errors commonly used in applications in the rectangular kernel case. They also have the advantage of not requiring the estimation of the density of the running variable at the discontinuity. In line with the theoretical findings, the fixed-h variance estimators are markedly improved over the small-h estimators in the presence of heteroskedasticity and should generally be preferred.

A simple application to the analysis in Lee (2008) of the electoral advantages of incumbency was presented and confirms that the improvements, in terms of precision of the estimates, are relevant in practice.

Regarding the potential presence of bias, the analysis above sheds light on the fact that, in practice, estimates in the RD context will be asymptotically biased unless one is willing to assume that the conditional expectation is correctly specified around the cutoff. This is the implicit assumption on the functional form being imposed by practitioners when using RD. Naturally, for larger bandwidths, one would expect that the potential for misspecification increases, requiring higher order polynomials of X or the inclusion of covariates to guarantee the validity of estimates.

References

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

[2] Calonico, S., Cattaneo, M., and Titiunik, R. (2014), "Robust Nonparametric Confidence Intervals for Regression Discontinuity Designs," *Econometrica, forthcoming.*

[3] Cattaneo, M., Frandsen, B., and Titiunik, R. (2012), "Randomization Inference in the Regression Discontinuity Design," Working Paper, Department of Economics, University of Michigan.

[4] 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.

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

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

[7] 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.

[8] — — (2001), "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201–209.

[9] 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.

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

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

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

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

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

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

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

[17] Pagan, A., and Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press.

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

[19] 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 ad J. Stock, Cambridge University Press, pp. 109–120.

[20] White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–24.

[21] — —(1996), *Estimation Inference and Specification Analysis*, Cambridge University Press.

Figure 1: Simulation for Infeasible Inference - Linear GDP - Homoskedastic Errors

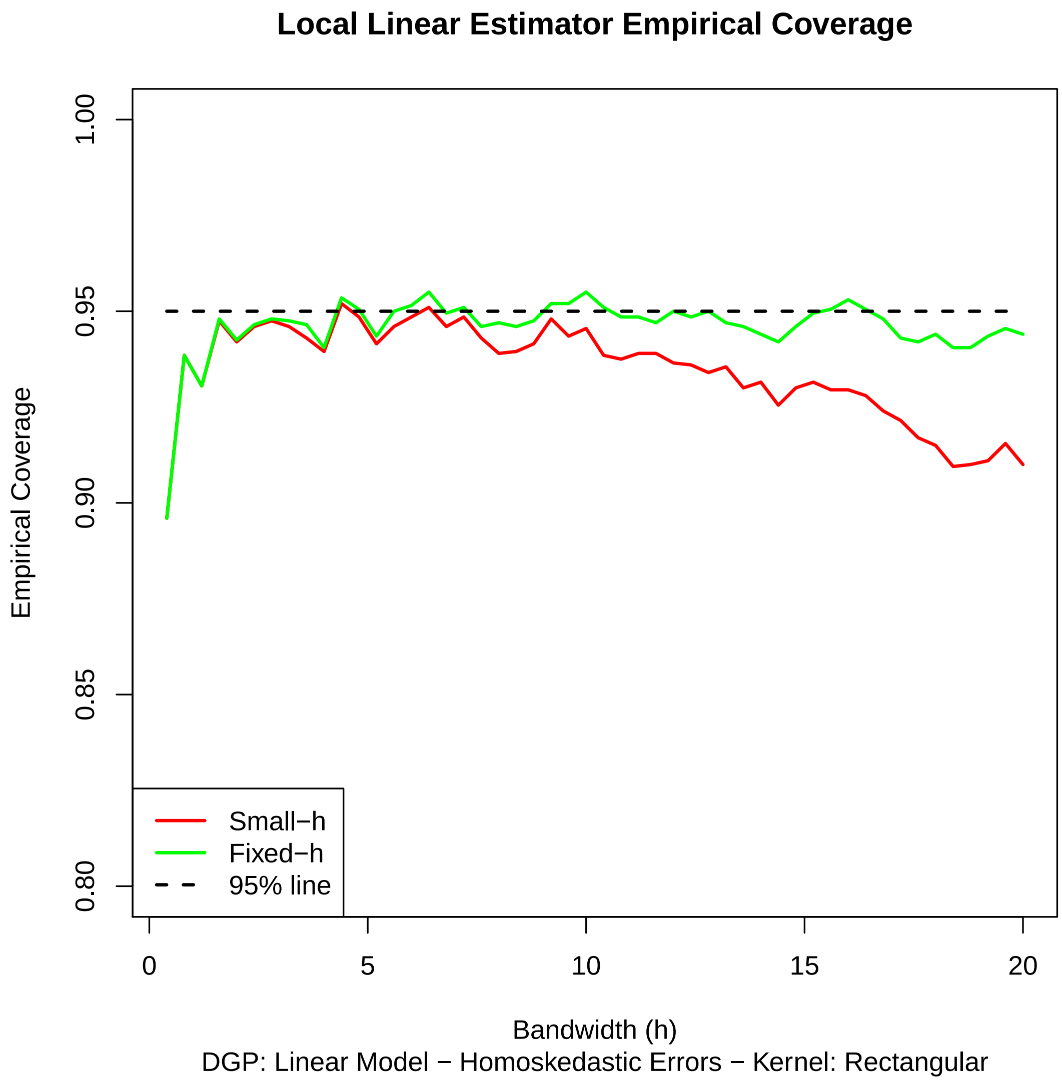

Figure 2: Simulation for Infeasible Inference - Exponential GDP - Homoskedastic Errors

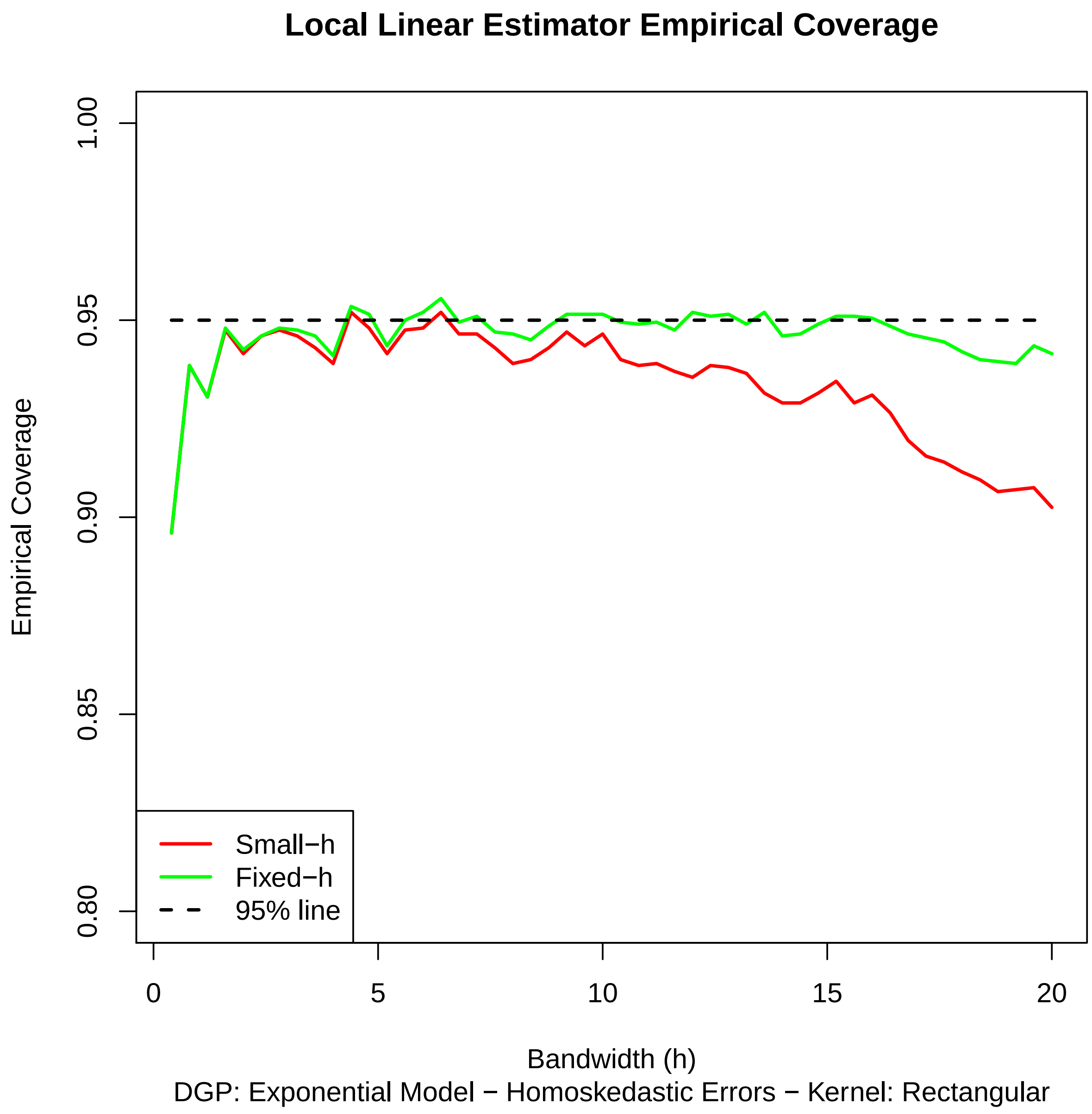

Figure 3: Simulation for Infeasible Inference - Lee (2008) GDP - Homoskedastic Errors

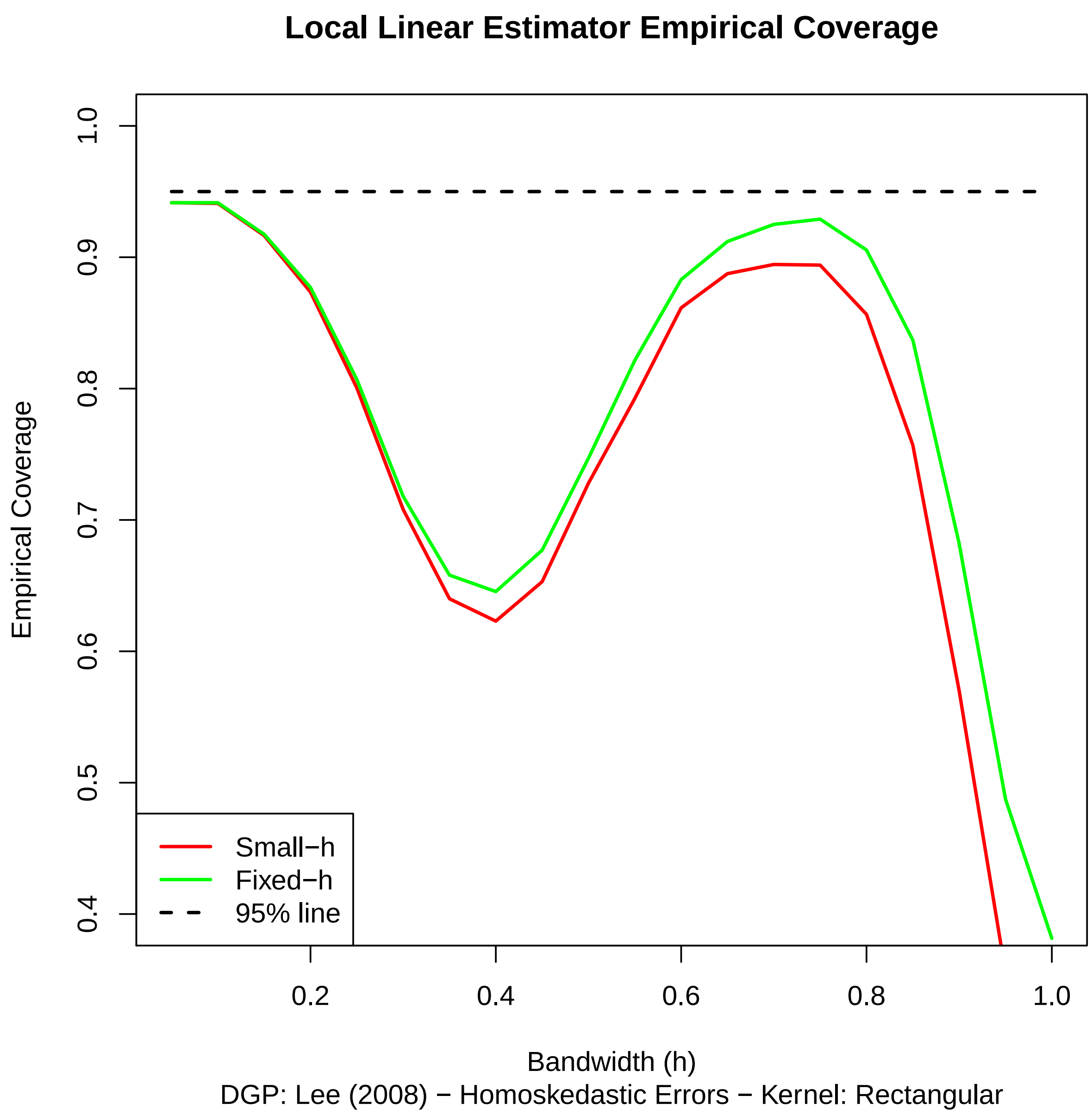

Figure 4: Simulation for Infeasible Inference - Linear GDP - Heteroskedastic Errors (Case 1)

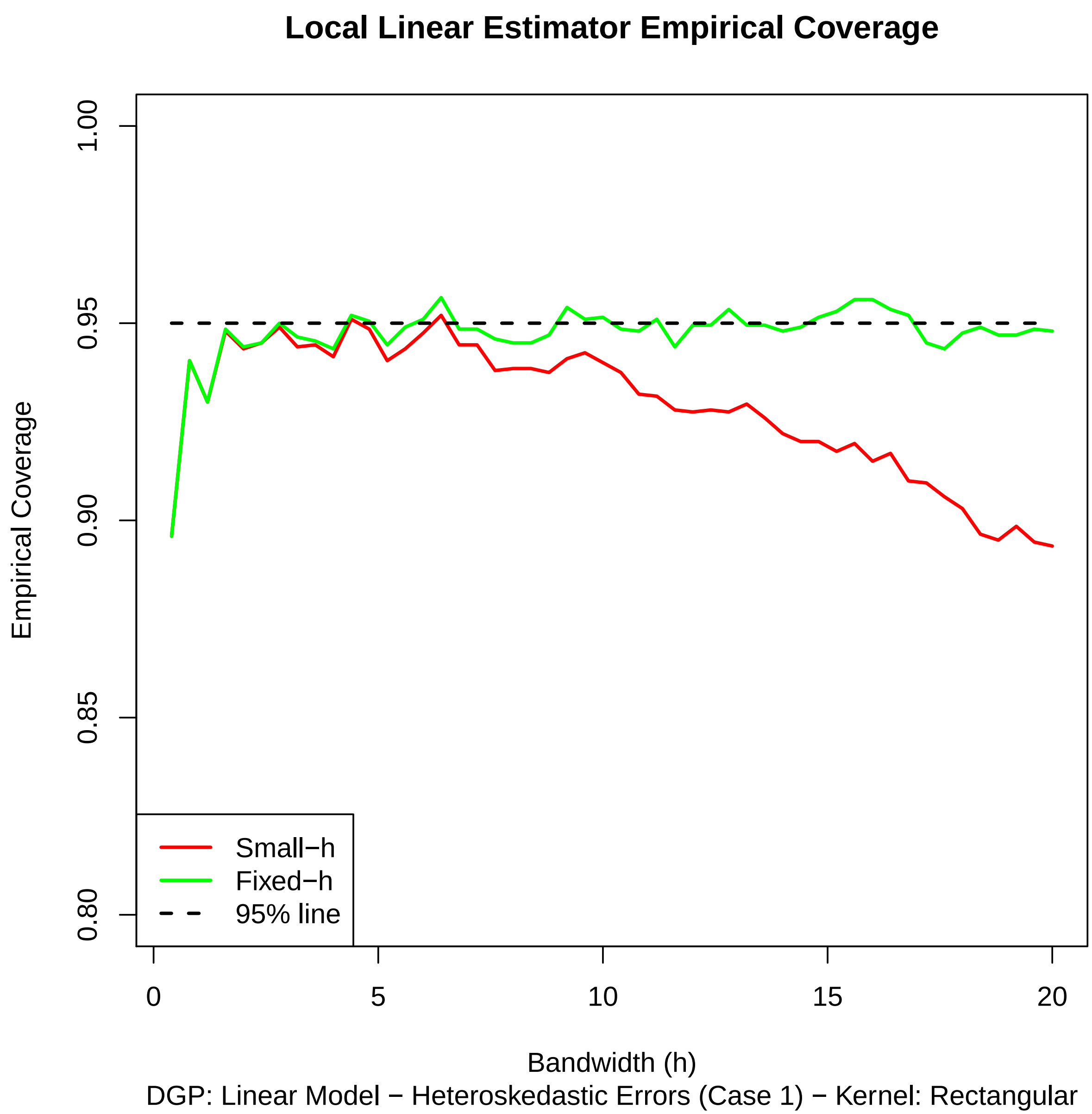

Figure 5: Simulation for Infeasible Inference - Exponential GDP - Heteroskedastic Errors (Case 1)

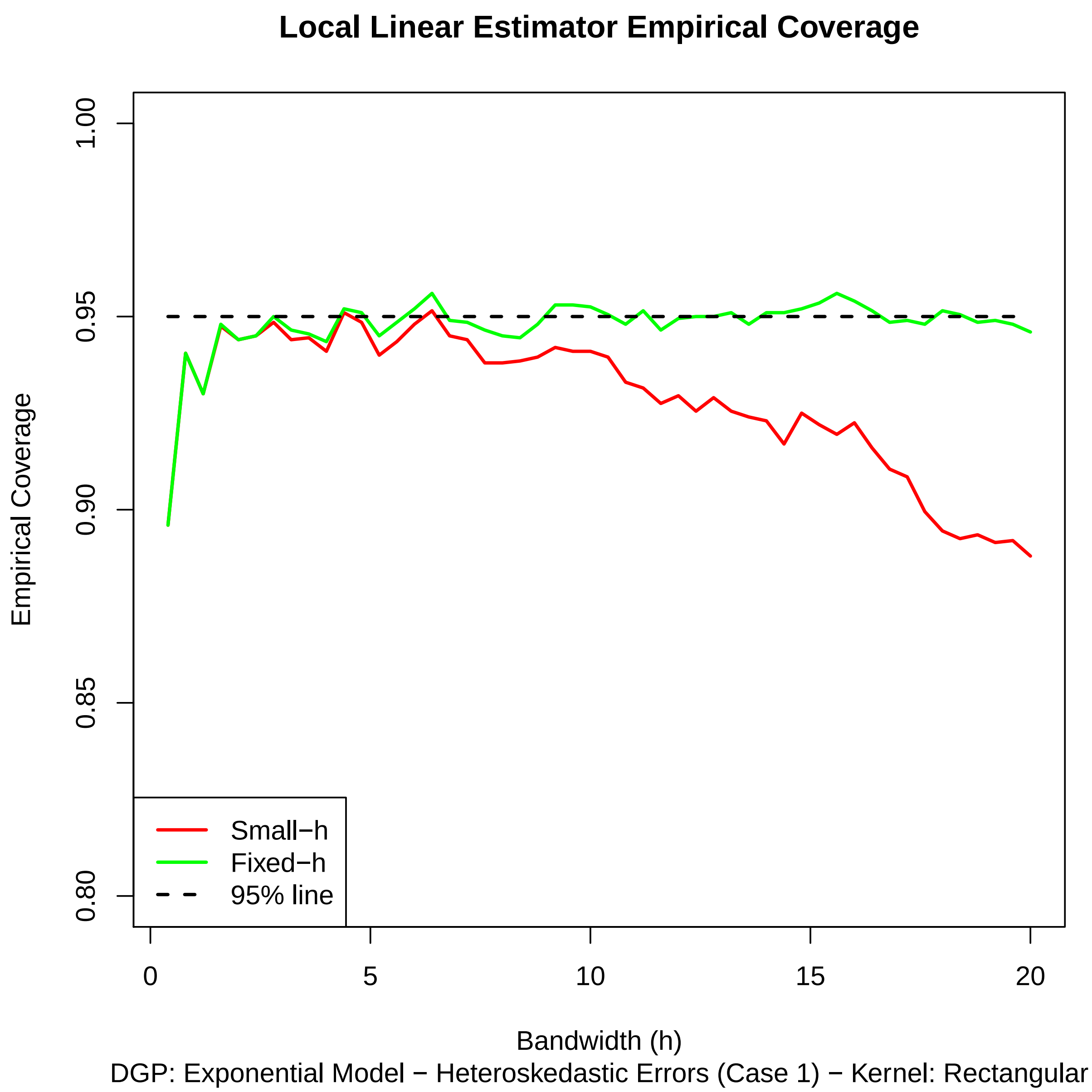

Figure 6: Simulation for Infeasible Inference - Linear GDP - Heteroskedastic Errors (Case 2)

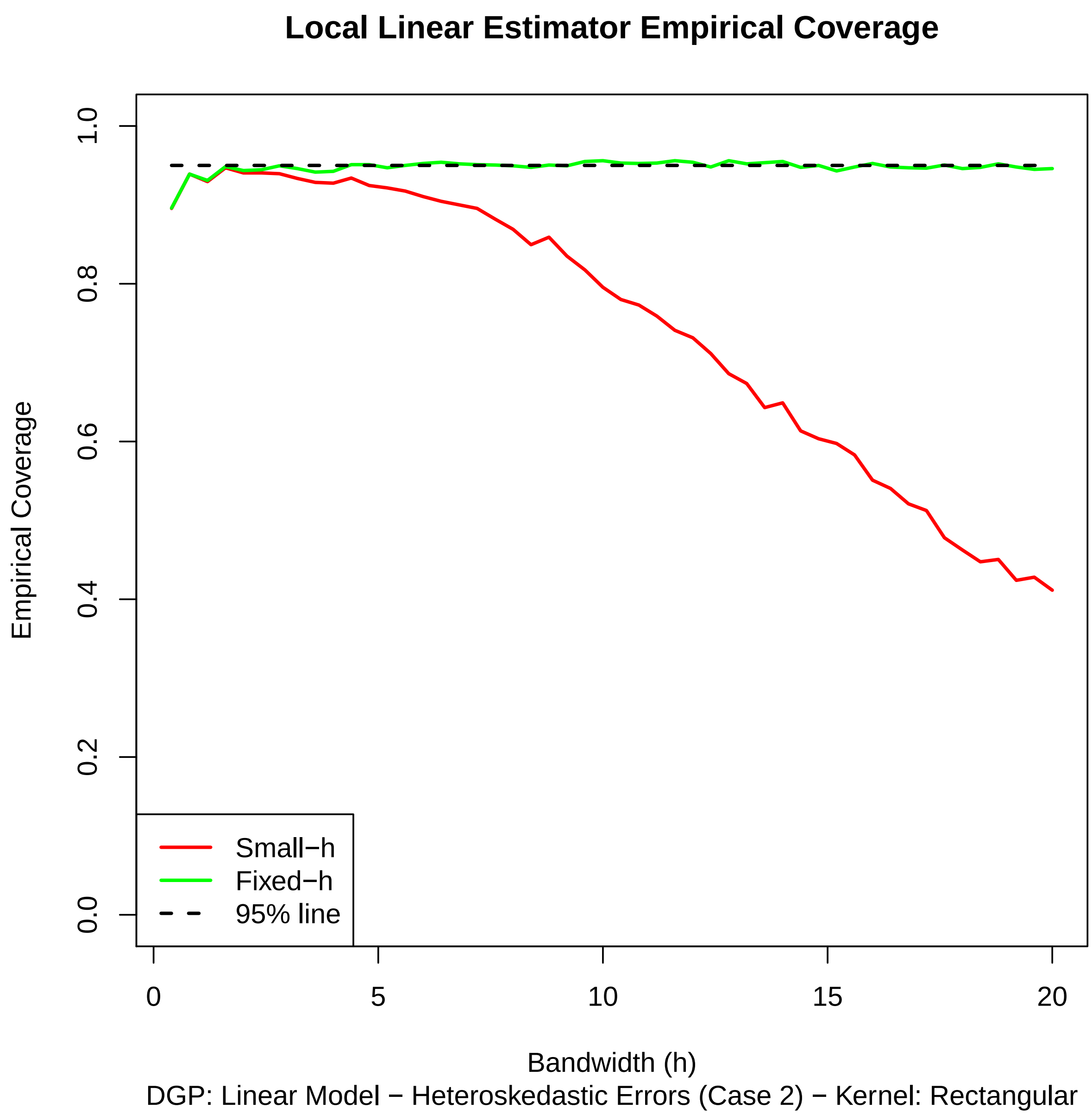

Figure 7: Simulation for Infeasible Inference - Exponential GDP - Heteroskedastic Errors (Case 2)

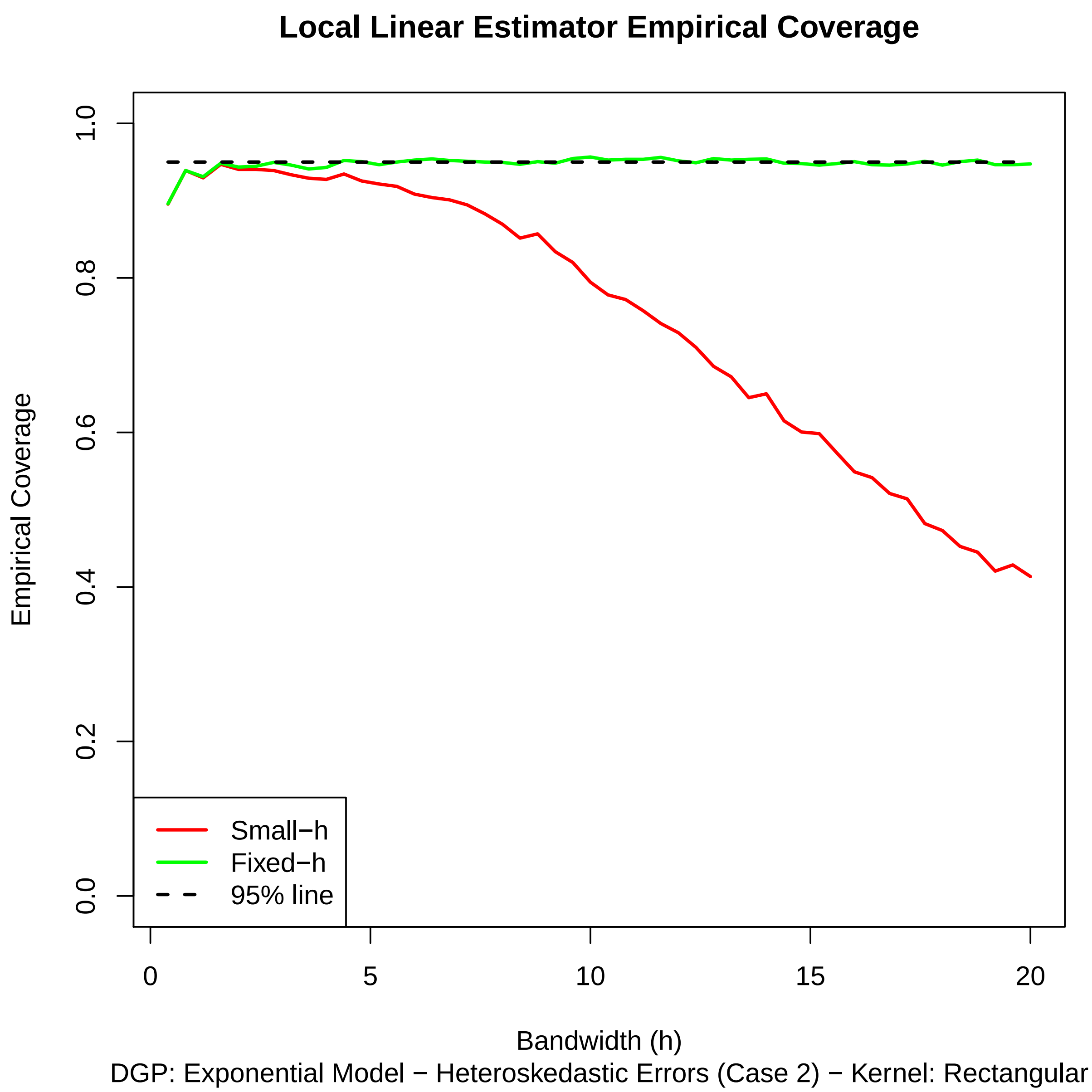

Figure 8: Simulation for Infeasible Inference - Lee (2008) GDP - Heteroskedastic Errors

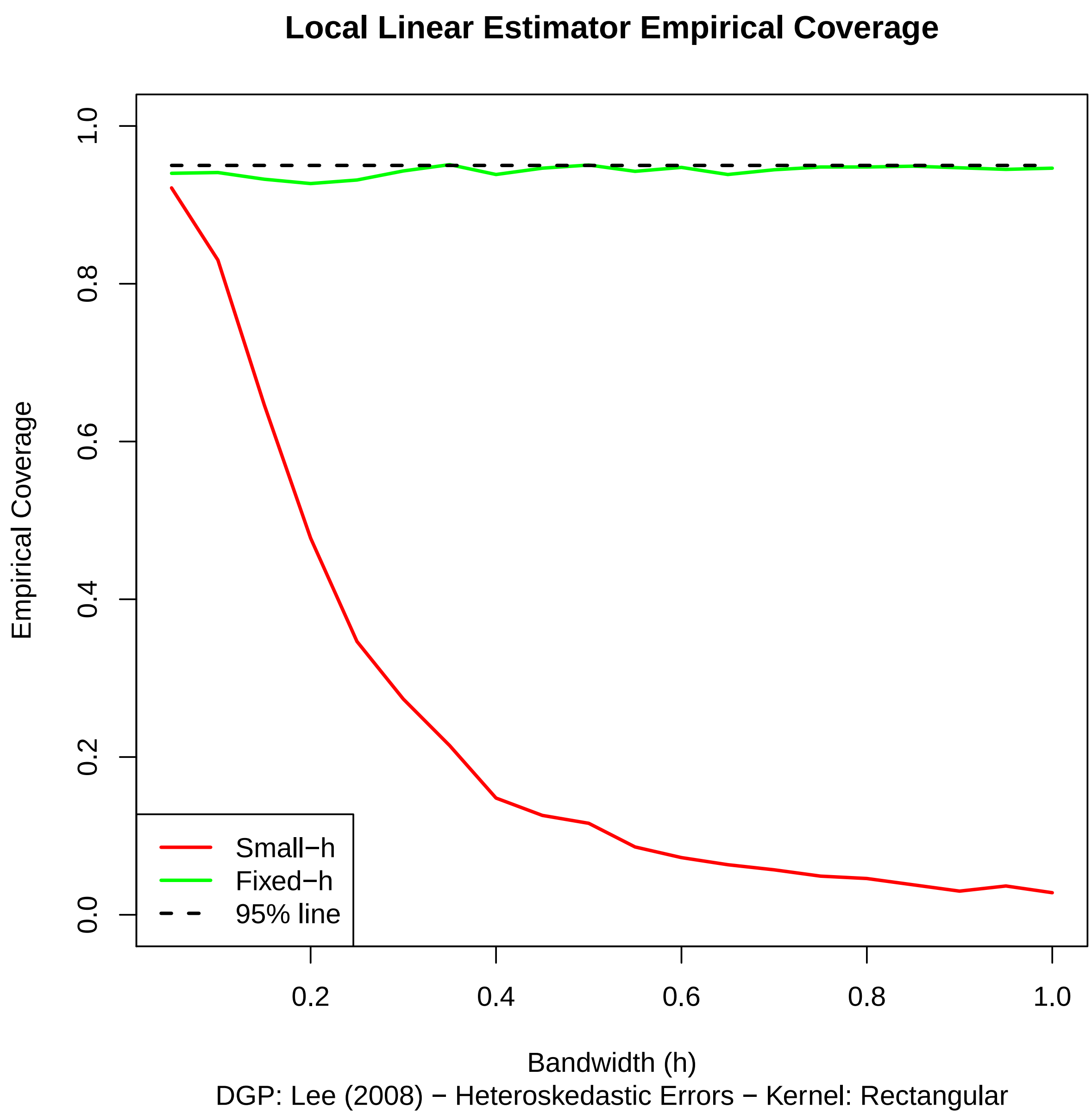

Figure 9: Simulation for Feasible Inference - Linear GDP - Homoskedastic Errors

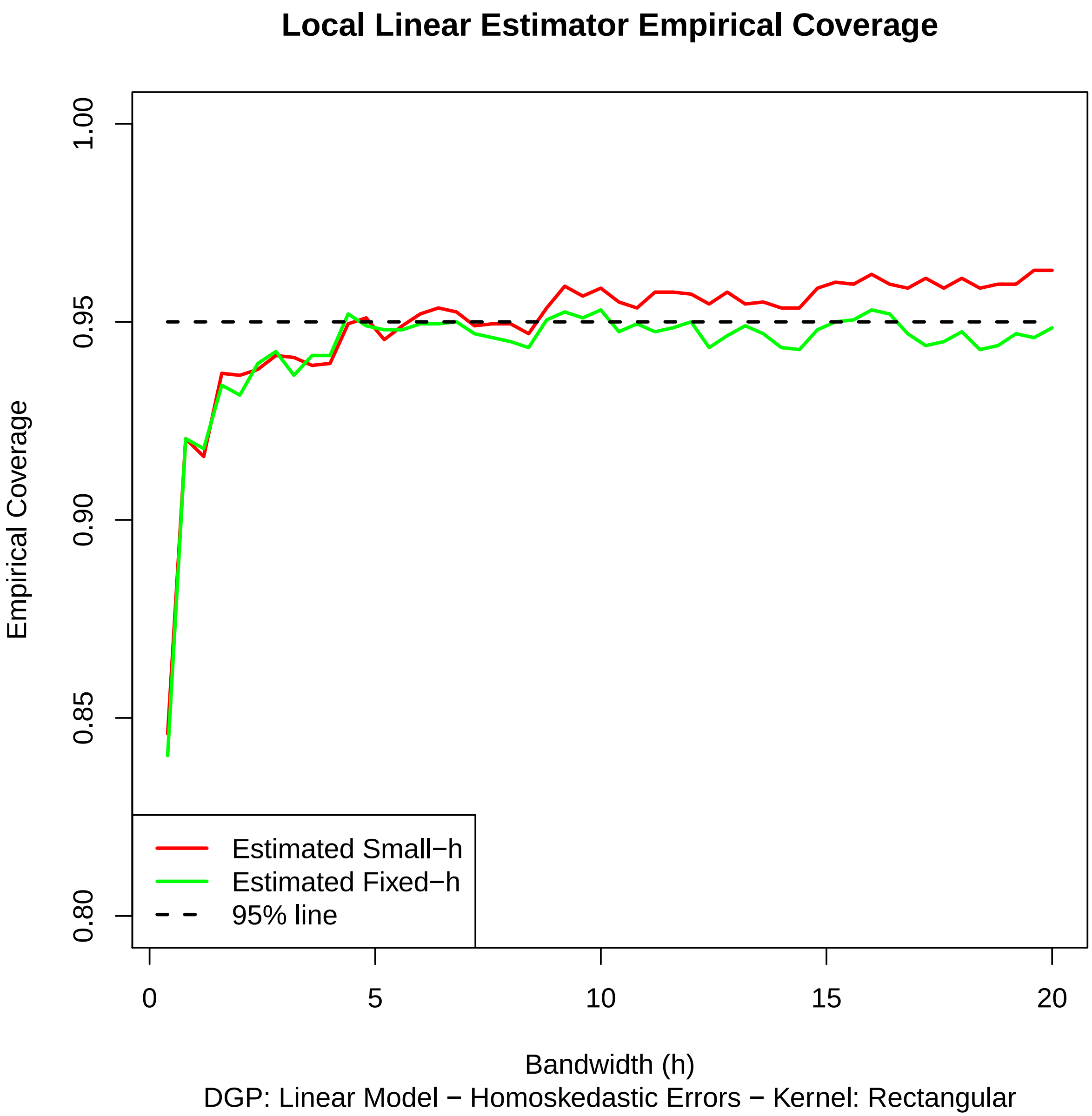

Figure 10: Simulation for Feasible Inference - Exponential GDP - Homoskedastic Errors

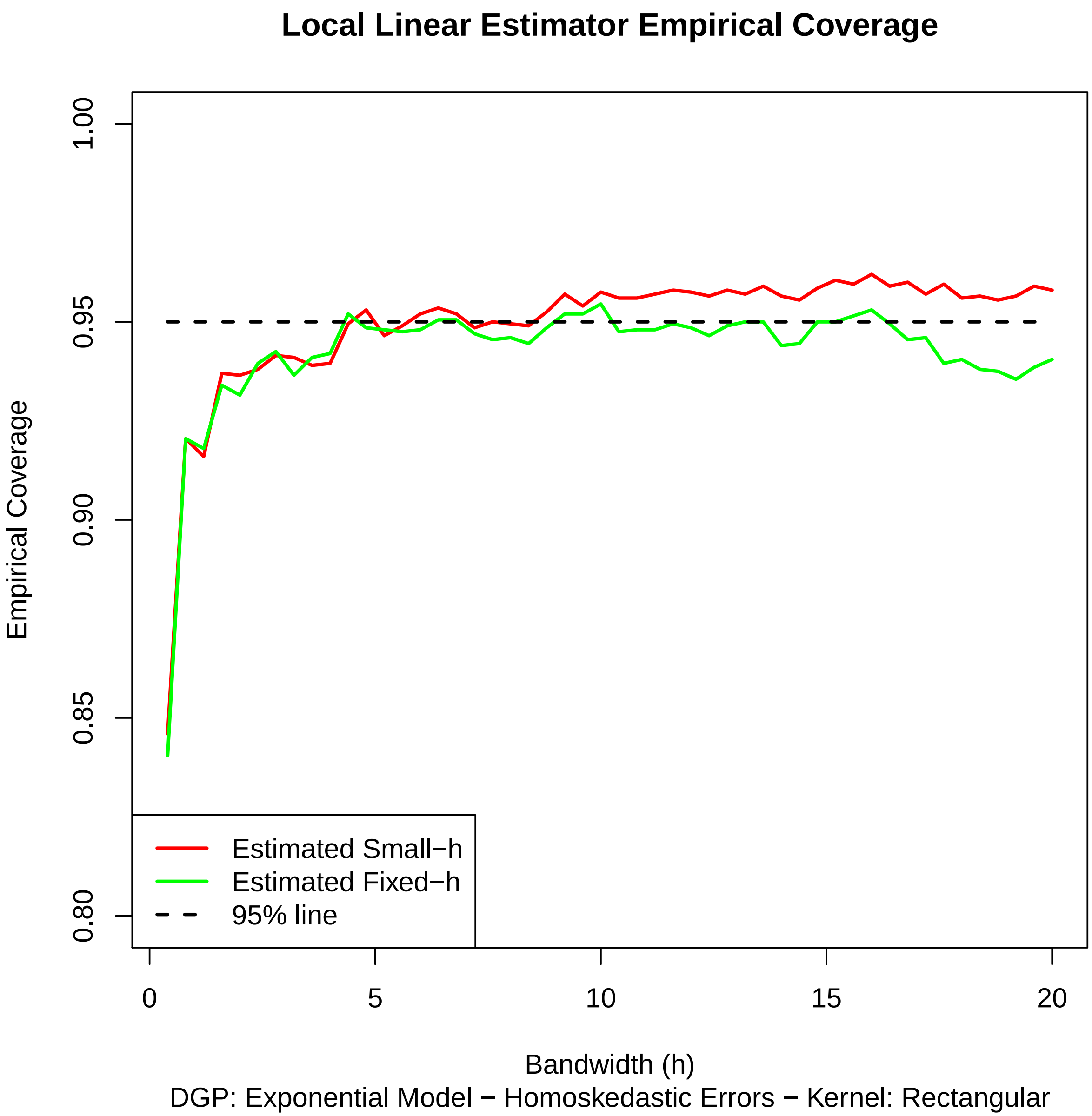

Figure 11: Simulation for Feasible Inference - Lee (2008) GDP - Homoskedastic Errors

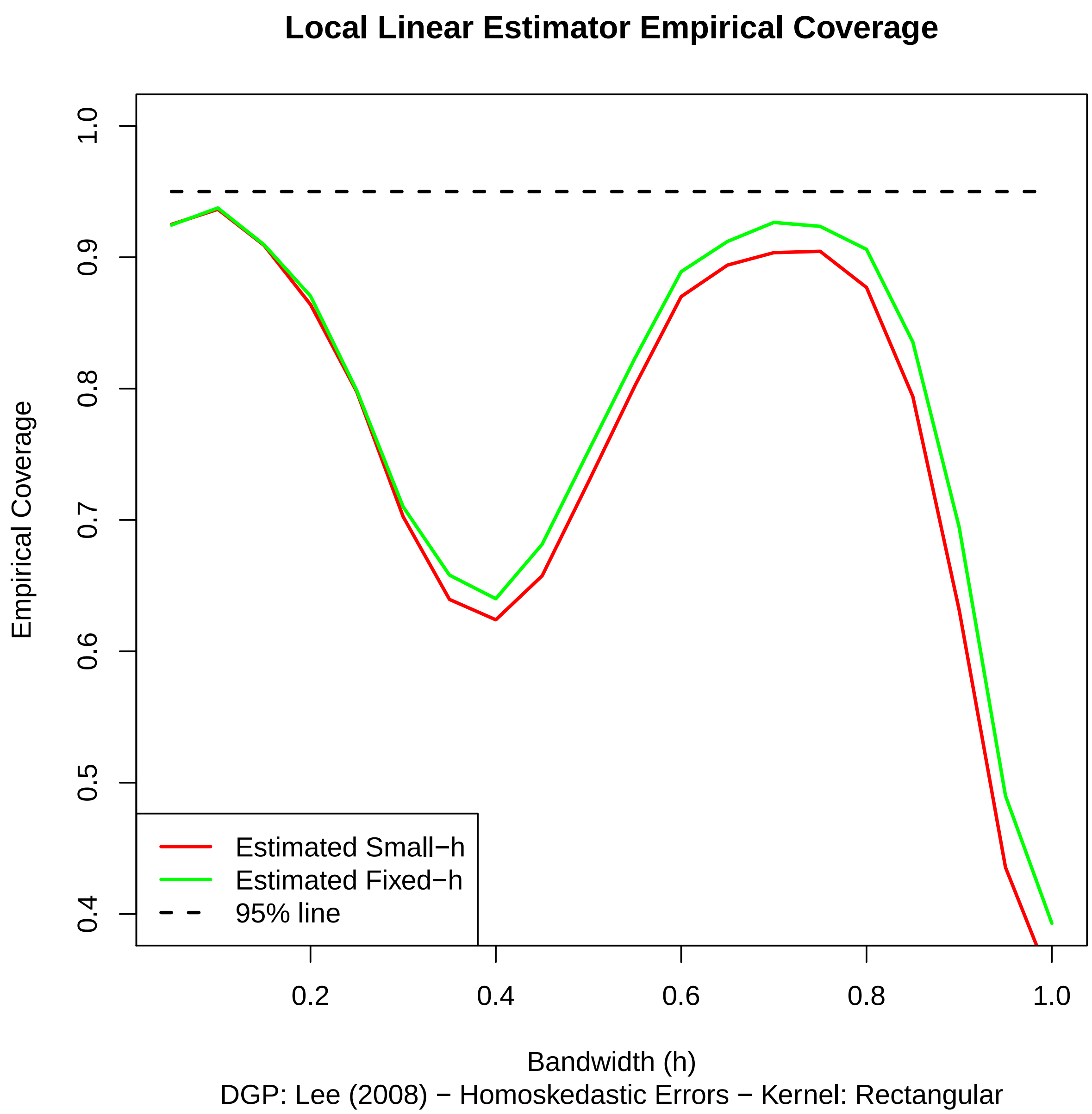

Figure 12: Simulation for Feasible Inference - Linear GDP - Heteroskedastic Errors (Case 1)

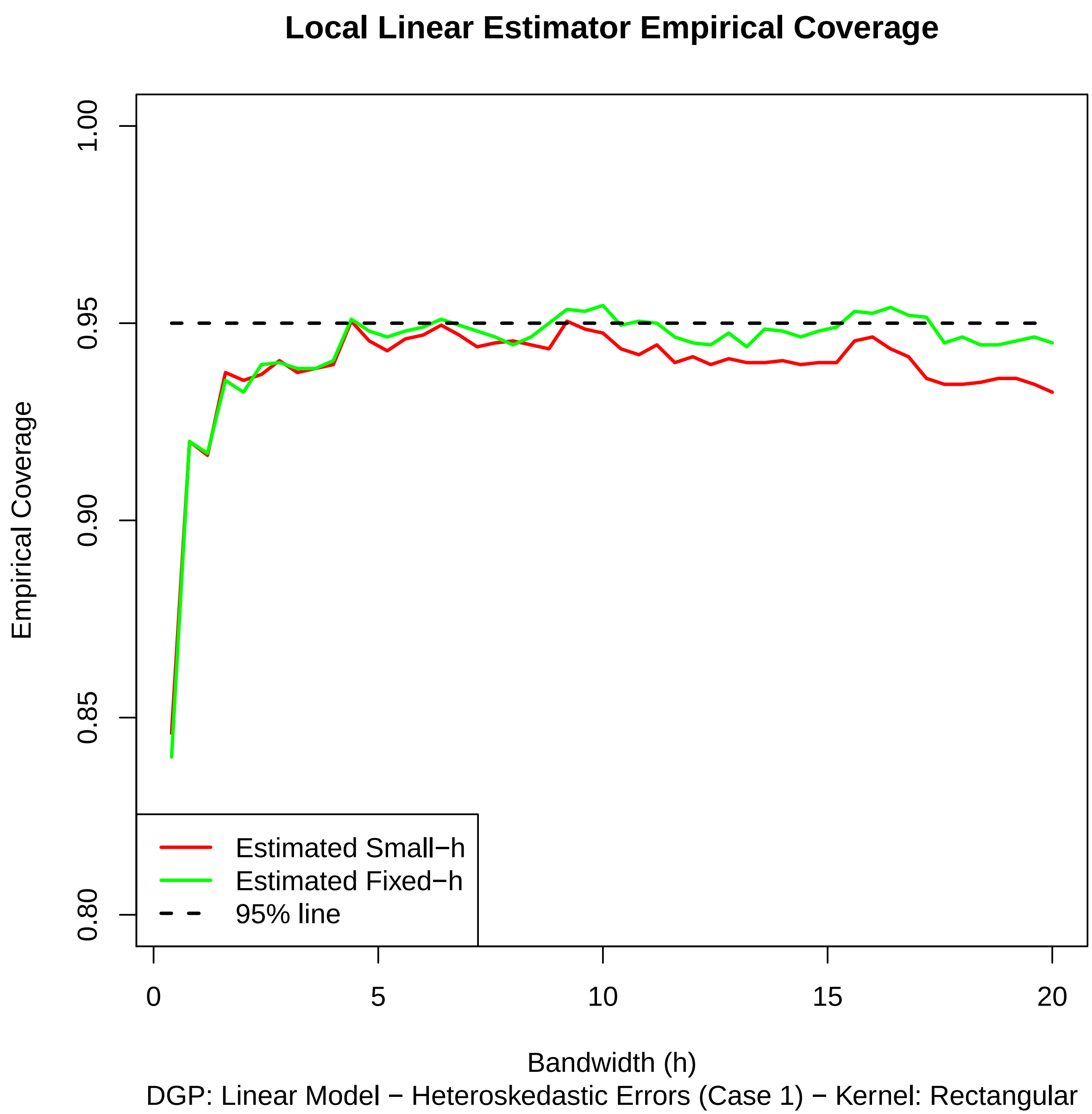

Figure 13: Simulation for Feasible Inference - Exponential GDP - Heteroskedastic Errors (Case 1)

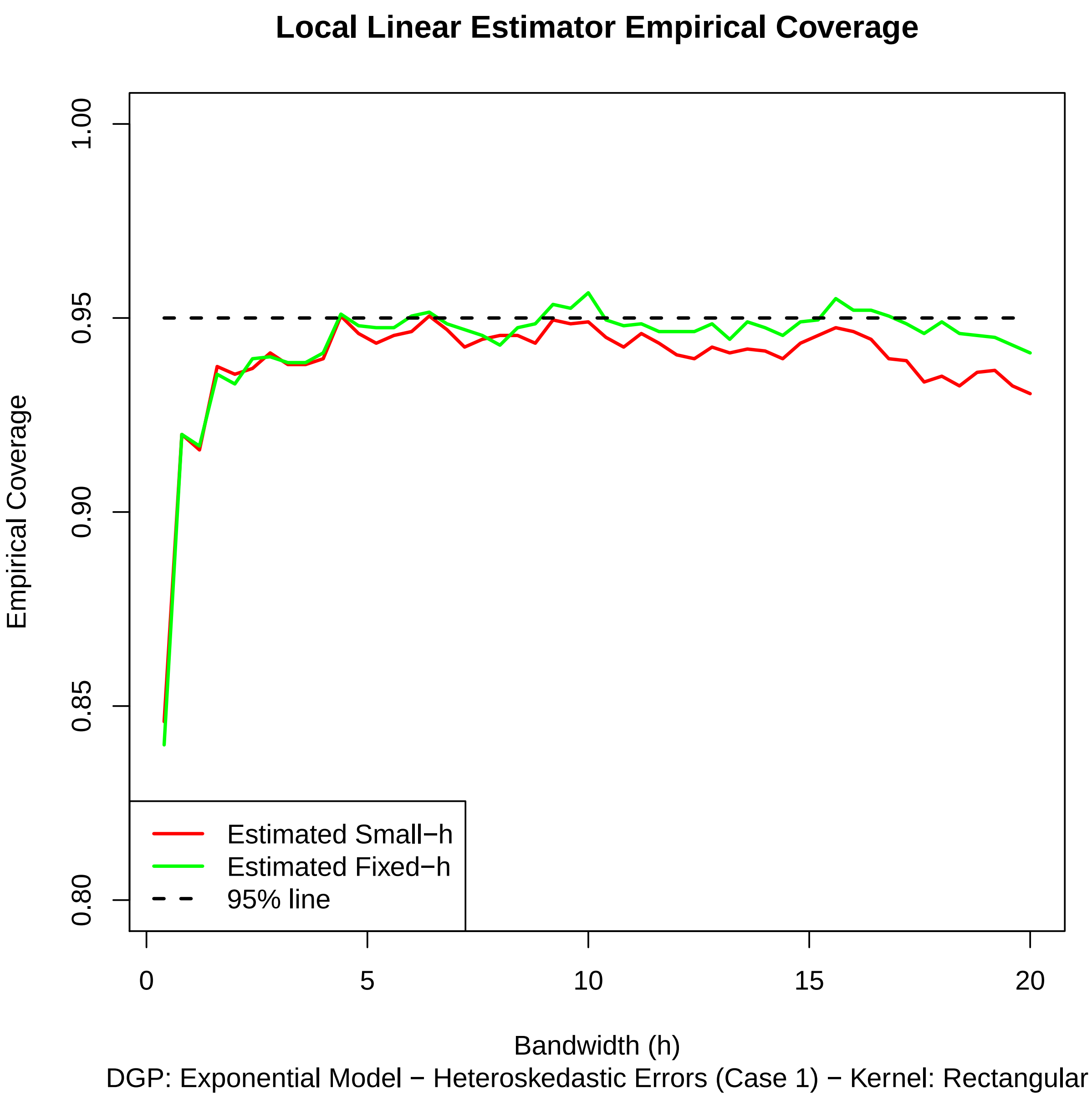

Figure 14: Simulation for Feasible Inference - Linear GDP - Heteroskedastic Errors (Case 2)

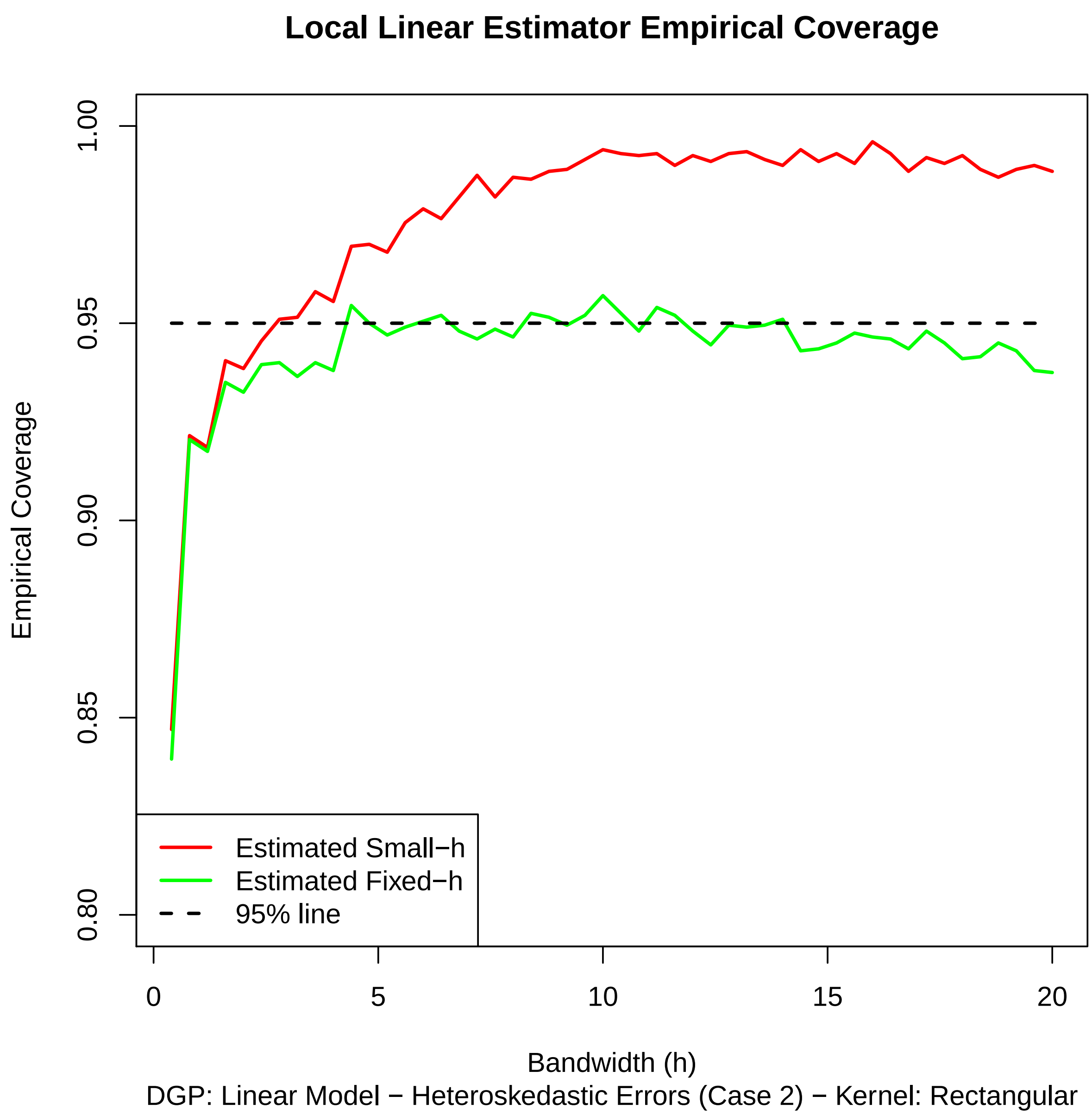

Figure 15: Simulation for Feasible Inference - Exponential GDP - Heteroskedastic Errors (Case 2)

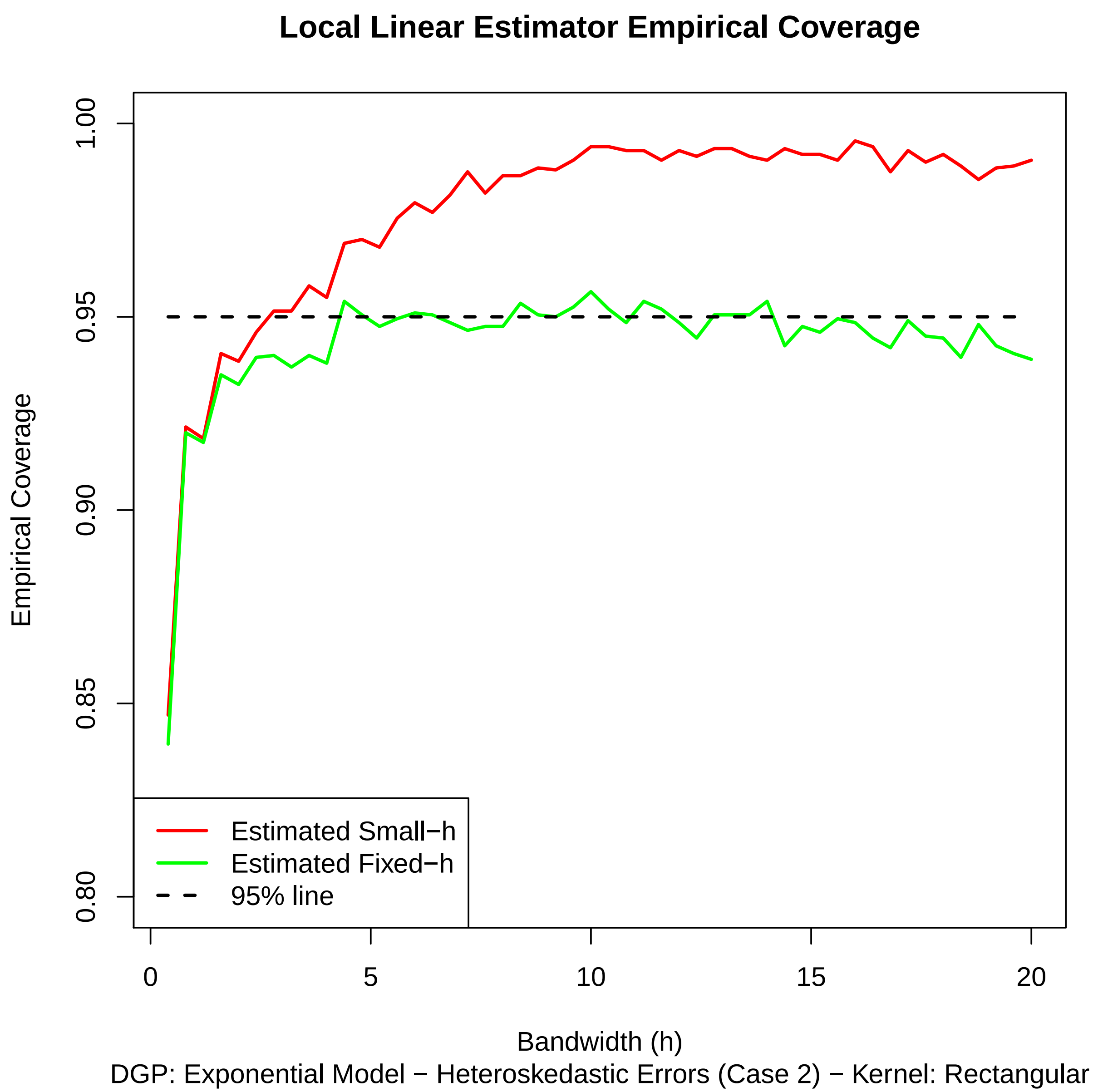

Figure 16: Simulation for Feasible Inference - Lee (2008) GDP - Heteroskedastic Errors

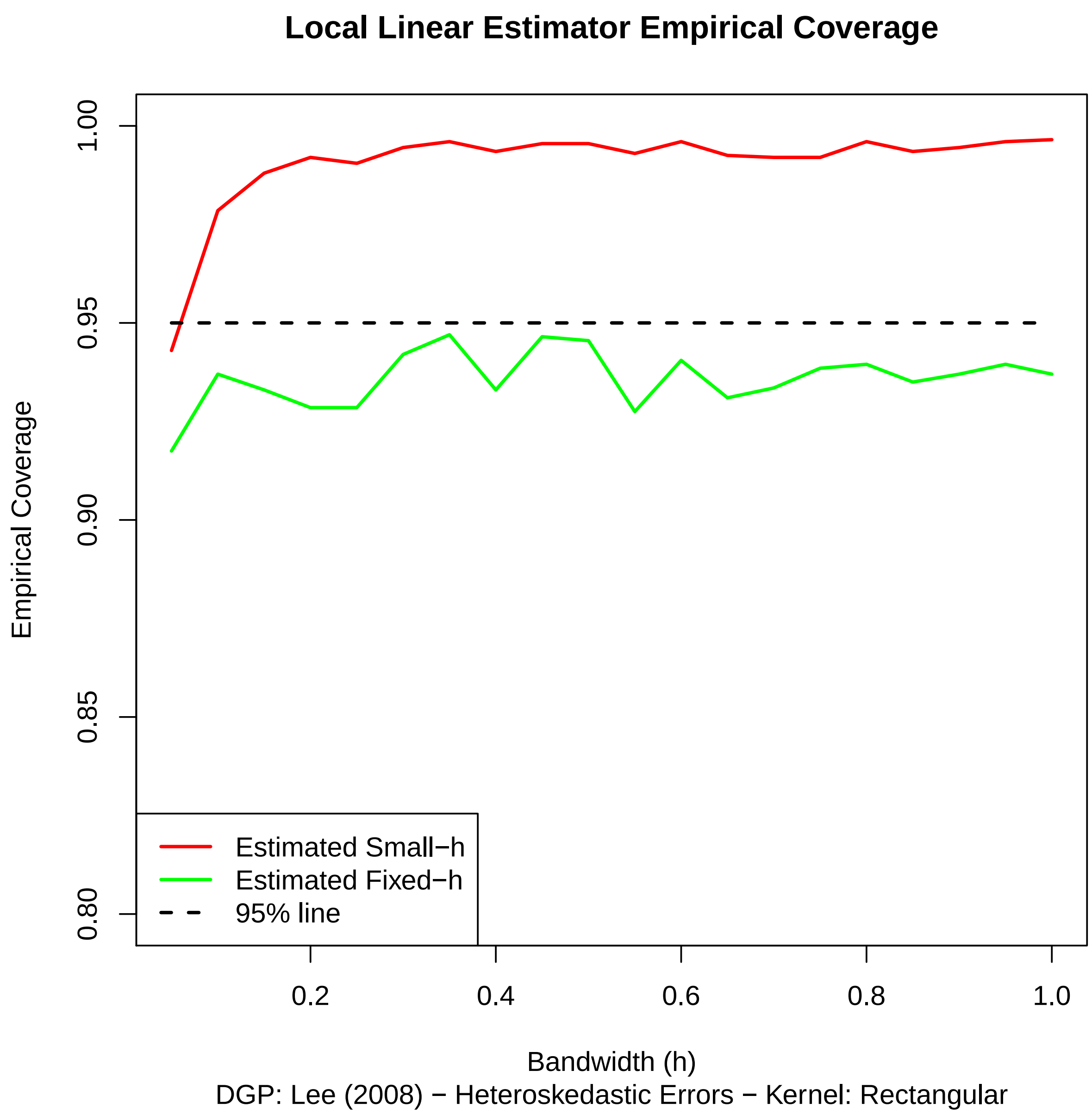

Figure 17: Simulation for Feasible Inference - Bandwidth Choice for $\hat{f}_o(\bar{x})$