Supplement to "Optimal inference in a class of regression models"

Timothy B. Armstrong[*] Michal Kolesár[†]
Yale University Princeton University

November 22, 2017

This supplement provides appendices not included in the main text. Supplemental Appendix C compares our approach with other methods, and includes a Monte Carlo study. Supplemental Appendix D contains details for the results in Section 3 not included in the main text. Supplemental Appendix E contains details for the RD application. Supplemental Appendix F considers feasible versions of the procedures in Section 3 in the case with unknown error distribution and derives their asymptotic efficiency. Supplemental Appendix G gives some auxiliary results used for relative asymptotic efficiency comparisons. Supplemental Appendix H gives the proof of Theorem E.1.

Appendix C Comparison with other methods

This section compares the CIs developed in this paper to other approaches to inference in the RD application. We consider two popular approaches. The first approach is to form a nominal $100 \cdot (1-\alpha)\%$ CI by adding and subtracting the $1-\alpha/2$ quantile of the $\mathcal{N}(0,1)$ distribution times the standard error, thereby ignoring any bias. We refer to these CIs as "conventional." The second approach is the robust bias correction (RBC) method studied by Calonico et al. (2014), which subtracts an estimate of the bias, and then takes into account the estimation error in this bias correction in forming the interval.

The coverage of these CIs will depend on the smoothness class $\mathcal{F}$ as well as the choice of bandwidth. Since CIs reported in applied work are typically based on local linear esti-

[*]email: timothy.armstrong@yale.edu
[†]email: mkolesar@princeton.edu

mators, with relative efficiency results for minimax MSE in the class $\mathcal{F}_{T,2}(C, \mathbb{R}_+)$ for estimation of $f(0)$ due to Cheng et al. (1997) often cited as justification, we focus on the class $\mathcal{F}_{RDT,2}(C)$ when computing coverage (in Supplemental Appendix C.2, we consider classes that also impose bounds on smoothness away from the discontinuity point rather than just placing bounds on the error of the Taylor approximation around the discontinuity point). If the bandwidth choice is non-random, then finite sample coverage can be computed exactly when errors are normal with known variance.[1] We take this approach in Supplemental Appendix C.1. If a data-driven bandwidth is used, computing finite sample coverage exactly becomes computationally prohibitive. We examine the coverage and relative efficiency of CIs with data driven bandwidths in a Monte Carlo study in Supplemental Appendix C.2.

C.1 Exact coverage with nonrandom bandwidth

For a given CI, we examine coverage in the classes $\mathcal{F}_{RDT,2}(C)$ by asking "what is the largest value of C for which this CI has good coverage?" Since the conventional CI ignores bias, there will always be some undercoverage, so we formalize this by finding the largest value of C such that a nominal 95% CI has true coverage 90%. This calculation is easily done using the formulas in Section 3.2: the conventional approach uses the critical value $z_{0.975} = \text{cv}_{0.05}(0)$ to construct a nominal 95% CI, while a valid 90% CI uses $\text{cv}_{0.1}(\overline{\text{bias}}_{\mathcal{F}_{RDT,2}(C)}(\hat{L})/\text{se}(\hat{L}))$ (where $\hat{L}$ denotes the estimator and $\text{se}(\hat{L})$ denotes its standard error), so we equate these two critical values and solve for C.

The resulting value of C for which undercoverage is controlled will depend on the bandwidth. To provide a simple numerical comparison to commonly used procedures, we consider the (data-dependent) Imbens and Kalyanaraman (2012, IK) bandwidth $\hat{h}_{IK}$ in the context of the Lee application considered in Section 4, but treat it as if it were fixed a priori. The IK bandwidth selector leads to $\hat{h}_{IK} = 29.4$ for local linear regression with the triangular kernel. The conventional two-sided CI based on this bandwidth is given by 7.99 ± 1.71. Treating the bandwidth as nonrandom, it achieves coverage of at least 90% over $\mathcal{F}_{RDT,2}(C)$ as long as $C \leq C_{\text{conv}} = 0.0018$. This is a rather low value, lower than the lower bound estimate on C from Supplemental Appendix E.3. It implies that even when $x = 20\%$, the prediction error based on a linear Taylor approximation to f can be reduced by less than 1% by using the true conditional expectation.

[1] The resulting coverage calculations hold in an asymptotic sense with unknown error distribution in the same way that, for example, coverage calculations in Stock and Yogo (2005) are valid in an asymptotic sense in the instrumental variables setting.

As an alternative to the conventional approach, one can use the robust-bias correction method studied in Calonico et al. (2014). Calonico et al. (2014) show that if the pilot bandwidth and the kernel used by the bias estimator equal those used by the local linear estimator of Lf, this method is equivalent to running a quadratic instead of a linear local regression, and then using the usual CI. In the Lee application with IK bandwidth, this delivers the CI 6.68 ± 2.52, increasing the half-length substantially relative to the conventional CI. The maximum smoothness parameter under which these CIs have coverage at least 90% is given by $C_{RBC} = 0.0023 > C_{\text{conv}}$. By way of comparison, the optimal 95% fixed-length CIs at C_{RBC} leads to a much narrower CI given by 7.70 ± 2.11.

While the CCT CI maintains good coverage for a larger smoothness constant than the conventional CI, both constants are rather small (equivalently, coverage is bad for moderate values of C). This is an artifact of the large realized value of $\hat{h}_{IK}$: the CCT CI essentially "undersmooths" relative to a given bandwidth by making the bias-standard deviation ratio smaller. Since $\hat{h}_{IK}$ is large to begin with, the amount of undersmoothing is not enough to make the procedure robust to moderate values of C. In fact, the IK bandwidth is generally quite sensitive to tuning parameter choices: we show in a Monte Carlo study in Supplemental Appendix C.2 that the CCT implementation of the IK bandwidth yields smaller bandwidths and achieves good coverage over a much larger set of functions, at the cost of larger length. In finite samples, the tuning parameters drive the maximum bias of the estimator, and hence its coverage properties, even though under standard pointwise asymptotics, the tuning parameters shouldn't affect coverage.

In contrast, if one performs the CCT procedure starting from a minimax MSE optimal bandwidth based on a known smoothness constant C, the asymptotic coverage will be quite good (above 94%), although the CCT CI ends up being about 30% longer than the optimal CI (see Armstrong and Kolesár, 2016b). Thus, while using a data driven bandwidth selector such as IK for inference can lead to severe undercoverage for smoothness classes used in RD (even if one undersmooths or bias-corrects as in CCT), procedures such as RBC can have good coverage if based on an appropriate bandwidth choice that is fixed ex ante.

C.2 Monte Carlo evidence with random bandwidth

Corollaries 3.2 and 3.3 imply that confidence intervals based on data-driven bandwidths must either undercover or else cannot be shorter than fixed-length CIs that assume worst-case smoothness. We now illustrate this implication with a Monte Carlo study.

We consider the RD setup from Section 2. To help separate the difficulty in constructing

CIs for Lf due to unknown smoothness of f from that due to irregular design points or heteroskedasticity, for all designs below, the distribution of x_i is uniform on $[-1, 1]$, and u_i is independent of x_i, distributed $\mathcal{N}(0, \sigma^2)$. The sample size is $n = 500$ in each case.

For σ^2, we consider two values, $\sigma^2 = 0.1295$, and $\sigma^2 = 4 \times 0.1295 = 0.518$. We consider conditional mean functions f that lie in the smoothness class

$$\mathcal{F}_{RDH,2}(C) = \{f_+ - f_- \colon f_+ \in \mathcal{F}_{H,2}(C; \mathbb{R}_+),\ f_- \in \mathcal{F}_{H,2}(C; \mathbb{R}_-)\},$$

where $\mathcal{F}_{H,p}(C; \mathcal{X})$ is the second-order Hölder class, the closure of twice-differentiable functions with second derivative bounded by $2C$, uniformly over $\mathcal{X}$:

$$\mathcal{F}_{H,p}(C; \mathcal{X}) = \{f \colon |f'(x_1) - f'(x_2)| \leq 2C|x_1 - x_2| \text{ all } x_1, x_2 \in \mathcal{X}\}.$$

Unlike the class $\mathcal{F}_{RDT,2}(C)$, the class $\mathcal{F}_{RDH,2}(C)$ also imposes smoothness away from the cutoff, so that $\mathcal{F}_{RDH,2}(C) \subseteq \mathcal{F}_{RDT,2}(C)$. Imposing smoothness away from the cutoff is natural in many empirical applications. We consider $C = 1$ and $C = 3$, and for each C, we consider 4 different shapes for f. In each case, f is odd, $f_+ = -f_-$. In Designs 1 through 3, f_+ is given by a quadratic spline with two knots, at b_1 and b_2,

$$f_+(x) = 1(x \geq 0) \cdot C\left(x^2 - 2(x - b_1)_+^2 + 2(x - b_2)_+^2\right).$$

In Design 1 the knots are given by $(b_1, b_2) = (0.45, 0.75)$, in Design 2 by $(0.25, 0.65)$, and in Design 3 by $(0.4, 0.9)$. The function $f_+(x)$ is plotted in Figure S1 for $C = 1$. For $C = 3$, the function f is identical up to scale. It is clear from the figure that although locally to the cutoff, the functions are identical, they differ away from the cutoff (for $|x| \geq 0.25$), which, as we demonstrate below, affects the performance of data-driven methods. Finally, in Design 4, we consider $f(x) = 0$ to allow us to compare the performance of CIs when f is as smooth as possible.

We consider four methods for constructing CIs based on data-driven bandwidths, and two fixed-length CIs. All CIs are based on local polynomial regressions with a triangular kernel. The variance estimators used to construct the CIs are based on the nearest-neighbor method described in Remark 2.1. The results based on Eicker-Huber-White variance estimators are very similar and not reported here.

The first two methods correspond to conventional CIs based on local linear regression described in Supplemental Appendix C.1. The first CI uses Imbens and Kalyanaraman

(2012, IK) bandwidth selector $\hat{h}_{IK}$, and the second CI uses a bandwidth selector proposed in Calonico et al. (2014, CCT), $\hat{h}_{CCT}$. The third CI uses the robust bias correction (RBC) studied in CCT, with both the pilot and the main bandwidth given by $\hat{h}_{IK}$ (the main estimate is based on local linear regression, and the bias correction is based on local quadratic regression), so that the bandwidth ratio is given by $\rho = 1$. The fourth CI is also based on RBC, but with the main and pilot bandwidth potentially different and given by the Calonico et al. (2014) bandwidth selectors. Finally, we consider two fixed-length CIs with uniform coverage under the class $\mathcal{F}_{RDH,2}(C)$, with $C = 1, 3$, and bandwidth chosen to minimize their half-length. Their construction is similar to the CIs considered in Section 2.2, except they use the fact that under $\mathcal{F}_{RDH,2}(C)$, the maximum bias for local linear estimators based on a fixed bandwidth is attained at $g^*(x) = Cx^2 1(x \geq 0) - Cx^2 1(x < 0)$ (see Armstrong and Kolesár, 2016b, for derivation).

The results are reported in Table S1 for $C = 1$ and S2 for $C = 3$. One can see from the tables that CIs based on $\hat{h}_{IK}$ may undercover severely even at the higher level of smoothness, $C = 1$. In particular, the coverage of conventional CIs based on $\hat{h}_{IK}$ is as low as 10.1% for 95% nominal CIs in Design 1, and the coverage of RBC CIs is as low as 64.4%, again in Design 1. The undercoverage is even more severe when $C = 3$.

In contrast, CIs based on the CCT bandwidth selector perform much better in terms of coverage under $C = 1$, with coverage over 90% for all designs. These CIs only start undercovering once $C = 3$, with 80.7% coverage in Design 3 for conventional CIs, and 86.2% coverage for RBC CIs. The cost for the good coverage properties, as can be seen from the tables, is that the CIs are longer, sometimes much longer than optimal fixed-length CIs.

As discussed in Supplemental Appendix C.1, the dramatically different coverage properties of the CIs based on the IK and CCT bandwidths illustrates the point that the coverage of CIs based on data-driven bandwidths is governed by the tuning parameters used in defining the bandwidth selector. These results can also be interpreted as showing the limits of procedures that try to "estimate C" from the data. In particular, we show in Armstrong and Kolesár (2016b) that for inference at a point based on local linear regression under the second-order Hölder class, in large samples the MSE-optimal bandwidth (see Remark 2.2) differs from the usual (infeasible) bandwidth minimizing the large-sample MSE under pointwise asymptotics only in that it replaces $f''(0)$ with C. Thus, plug-in rules that estimate the infeasible pointwise bandwidth by plugging in an estimate of $f''(0)$ can be interpreted as data-driven bandwidths that try to estimate C from the data. Since the IK and CCT bandwidths are plug-in rules, to the extent that one can interpret them as trying to "estimate C"

from the data, these simulation results also illustrate the point that attempts to estimate C from the data cannot improve upon FLCIs (one can show that if these procedures were successful at estimating C, conventional CIs with 95% nominal level based on them should have coverage no less than 92.1% in large samples).

To assess sensitivity of these results to the normality and homoskedasticity of the errors, we also considered Designs 1–4 with heteroskedastic and log-normal errors. The results (not reported here) are similar in the sense that if a particular method achieved close to 95% coverage under normal homoskedastic errors, the coverage remained good under alternative error distributions. If a particular method undercovered in a given design, the amount of undercoverage could be more or less severe, depending on the form of heteroskedasticity. In particular, fixed-length CIs with $C = 3$ achieve excellent coverage for all designs and all error distributions considered.

Appendix D Additional details for Section 3

This section contains details for the results in Section 3 not included in the main text.

D.1 Special cases

In addition to regression discontinuity, the regression model (1) covers several other important models, including inference at a point ($Lf = f(x_0)$ with x_0 given) and average treatment effects under unconfoundedness (with $Lf = \frac{1}{n}\sum_{i=1}^{n}(f(w_i, 1) - f(w_i, 0))$ where $x_i = (w_i', d_i)'$, d_i is a treatment indicator and w_i are controls).

The setup (18) can also be used to study the linear regression model with restricted parameter space. For simplicity, consider the case with homoskedastic errors,

$$Y = X\theta + \sigma\varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I_n), \tag{S1}$$

where X is a fixed $n \times k$ design matrix and σ is known. This fits into our framework with $f = \theta$, X playing the role of K, taking $\theta \in \mathbb{R}^k$ to $X\theta \in \mathbb{R}^n$, and $\mathcal{Y} = \mathbb{R}^n$ with the Euclidean inner product $\langle x, y\rangle = x'y$. We are interested in a linear functional $L\theta = \ell'\theta$ where $\ell \in \mathbb{R}^k$. We consider this model in previous version of this paper (Armstrong and Kolesár, 2016a). Furthermore, (18) covers the multivariate normal location model $\hat{\theta} \sim \mathcal{N}(\theta, \Sigma)$, which obtains as a limiting experiment of regular parametric models. Our finite-sample results could thus be extended to local asymptotic results in regular parametric models with restricted parameter spaces.

In addition to the regression models (1) and (S1), the setup (18) includes other nonparametric and semiparametric regression models such as the partly linear model (where f takes the form $g(w_1)+\gamma' w_2$, and we are interested in a linear functional of g or γ). It also includes the Gaussian white noise model, which can be obtained as a limiting model for nonparametric density estimation (see Nussbaum, 1996) as well as nonparametric regression with fixed or random regressors (see Brown and Low, 1996; Reiß, 2008). These white noise equivalence results imply that our finite-sample results translate to asymptotic results in problems such as inference at a point in density estimation or regression with random regressors. We refer the reader to Donoho (1994, Section 9) for details of these and other models that fit into the general setup (18).

D.2 Derivative of the modulus

The class of optimal estimators $\hat{L}_{\delta,\mathcal{F},\mathcal{G}}$ involves the superdifferential of the modulus. In the case where the modulus is differentiable, the superdifferential is a singleton, so that $\hat{L}_{\delta,\mathcal{F},\mathcal{G}}$ is defined uniquely. In this section, we introduce a condition that guarantees differentiability and leads to a formula for the derivative. We also briefly discuss the case where the modulus is not differentiable.

Definition 1 (Translation Invariance)**.** *The function class $\mathcal{F}$ is translation invariant if there exists a function $\iota \in \mathcal{F}$ such that $L\iota = 1$ and $f + c\iota \in \mathcal{F}$ for all $c \in \mathbb{R}$ and $f \in \mathcal{F}$.*

Translation invariance will hold in most cases where the parameter of interest Lf is unrestricted. For example, if $Lf = f(0)$, it will hold with $\iota(x) = 1$ if $\mathcal{F}$ places monotonicity restrictions and/or restrictions on the derivatives of f. Under translation invariance, the modulus is differentiable, and we obtain an explicit expression for its derivative:

Lemma D.1. *Let f^* and g^* solve the modulus problem with $\delta_0 = \|K(g^* - f^*)\| > 0$, and suppose that $f^* + c\iota \in \mathcal{F}$ for all c in a neighborhood of zero, where $L\iota = 1$. Then the modulus is differentiable at δ_0 with $\omega'(\delta_0; \mathcal{F}, \mathcal{G}) = \delta_0/\langle K\iota, K(g^*_{\delta_0} - f^*_{\delta_0})\rangle$.*

Proof. Let $d \in \partial\omega(\delta_0; \mathcal{F}, \mathcal{G})$ and let $f_c = f^* - c\iota$. Let η be small enough so that $f_c \in \mathcal{F}$ for $|c| \le \eta$. Then, for $|c| \le \eta$,

$$L(g^* - f^*) + d\,[\|K(g^* - f_c)\| - \delta_0] \ge \omega(\|K(g^* - f_c)\|; \mathcal{F}, \mathcal{G}) \ge L(g^* - f_c) = L(g^* - f^*) + c$$

where the first inequality follows from the definition of the superdifferential and the second inequality follows from the definition of the modulus. Since the left-hand side of the above display is greater than or equal to the right-hand side for $|c| \leq \eta$, and the two sides are equal at $c = 0$, the derivatives of both sides with respect to c must be equal. Since

$$\frac{d\|K(g^* - f_c)\|}{dc}\bigg|_{c=0} = \frac{\frac{d}{dc}\|K(g^* - f_c)\|^2\big|_{c=0}}{2\delta_0} = \frac{\langle K(g^* - f^*), K\iota\rangle}{\delta_0},$$

result follows. □

The explicit expression for $\omega'(\delta; \mathcal{F}, \mathcal{G})$ is useful in simplifying the expressions (23) and (25) for the optimal estimators.

Translation invariance leads to a direct relation between optimal CIs and tests. In general, it can be seen from Lemma A.2 that the test that rejects L_0 when $L_0 \notin [\hat{c}_{\alpha,\delta,\mathcal{F},\mathcal{G}}, \infty)$ is minimax for $H_0 : Lf \leq L_0$ and $f \in \mathcal{F}$ against $H_1 : Lf \geq L_0 + \omega(\delta; \mathcal{F}, \mathcal{G})$ and $f \in \mathcal{G}$, where $L_0 = Lf_\delta^*$. If both $\mathcal{F}$ and $\mathcal{G}$ are translation invariant, $f_\delta^* + c\iota$ and $g_\delta^* + c\iota$ achieve the ordered modulus for any $c \in \mathbb{R}$, so that, varying c, this test can be seen to be minimax for any L_0. Thus, under translation invariance, the CI in Theorem 3.1 inverts minimax one sided tests with distance to the null given by $\omega(\delta)$ (in general, the test based on the CI in Theorem 3.1 is minimax only when $L_0 = Lf_\delta^*$).

If the modulus is not differentiable at some δ, the CIs defined in Sections 3.3 and 3.4 are valid with $\omega'(\delta, \mathcal{F}, \mathcal{G})$ given by any element of the superdifferential, so long as the same element of the superdifferential is used throughout the formula (in particular, the same element used in the estimator (23) must be used in the worst-case bias formula (24)). For the one-sided CI, Theorem 3.1 applies regardless of which element of the superdifferential is used. In the two-sided case, when computing the optimal fixed-length affine CI described in Section 3.4, the only additional detail in the case where the modulus is not everywhere differentiable is that one optimizes the half-length over both δ and over elements in the superdifferential.

Appendix E Additional details for RD

This section gives additional details for the RD application. Supplemental Appendix E.1 derives the worst-case bias formula given in (11). Supplemental Appendix E.2 derives the optimal estimator and the solution to the modulus problem. Supplemental Appendix E.3 discusses lower bounds for the smoothness constant C. Supplemental Appendix E.4 shows the asymptotic validity of the feasible version of the estimator in which the variance is estimated.

E.1 Worst-case bias for linear estimators

This section derives the worst-case bias formula (11) for linear estimators $\hat{L}_{h_+,h_-}$ defined in (10) in Section 2.2. We require the weights to satisfy $w_+(-x, h_+) = w_-(x, h_-) = 0$ for $x \geq 0$ and

$$\sum_{i=1}^{n} w_+(x_i, h_+) = \sum_{i=1}^{n} w_-(x_i, h_-) = 1,$$
$$\sum_{i=1}^{n} x_i^j w_-(x_i, h_-) = \sum_{i=1}^{n} x_i^j w_+(x_i, h_+) = 0 \text{ for } j = 1, \dots, p-1. \tag{S2}$$

Note that (S2) holds iff. $\hat{L}_{h_+,h_-}$ is unbiased at all $f = f_+ + f_-$ where f_+ and f_- are both polynomials of order $p-1$ or less, which is necessary to ensure that the worst-case bias is finite. This condition holds if $\hat{L}_{h_+,h_-}$ is based on a local polynomial estimator of order at least $p-1$.

We can write any function $f \in \mathcal{F}_{RDT,p}(C)$ as $f = f_+ + f_-$, where

$$f_+(x) = [\sum_{j=0}^{p-1} f_+^{(j)}(0)x^j/j! + r_+(x)]1(x \geq 0), \quad f_-(x) = [\sum_{j=0}^{p-1} f_-^{(j)}(0)x^j/j! + r_-(x)]1(x < 0),$$

and the remainder terms r_+ and r_- satisfy $|r_+(x)| \leq C|x|^p$ and $|r_-(x)| \leq C|x|^p$. Under (S2), we can therefore write

$$\text{bias}_f(\hat{L}_{h_+,h_-}) = \sum_{i=1}^{n} w_+(x_i, h_+)r_+(x) - \sum_{i=1}^{n} w_-(x_i, h_+)r_-(x),$$

which is maximized subject to the conditions $|r_+(x)| \leq C|x|^p$ and $|r_-(x)| \leq C|x|^p$ by taking $r_+(x_i) = C|x_i|^p \cdot \text{sign}(w_+(x_i, h_+))$ and $r_-(x_i) = -C|x_i|^p \cdot \text{sign}(w_-(x_i, h_-))$. This yields the worst-case bias formula Equation (11).

E.2 Solution to the modulus problem and optimal estimators

This section derives the form of the optimal estimators and CIs. To that end, we first need to find functions g_δ^* and f_δ^* that solve the modulus problem. Since the class $\mathcal{F}_{RDT,p}(C)$ is centrosymmetric, $f_\delta^* = -g_\delta^*$, and the (single-class) modulus of continuity $\omega(\delta; \mathcal{F}_{RDT,p}(C))$ is given by the value of the problem

$$\sup_{f_+ + f_- \in \mathcal{F}_{RDT,p}(C)} 2(f_+(0) - f_-(0)) \quad \text{st} \quad \sum_{i=1}^n \frac{f_-(x_i)^2}{\sigma^2(x_i)} + \sum_{i=1}^n \frac{f_+(x_i)^2}{\sigma^2(x_i)} \le \delta^2/4. \tag{S3}$$

Let $g^*_{\delta,C}$ denote the (unique up to the values at the x_is) solution to this problem. This solution can be obtained using a simple generalization of Theorem 1 of Sacks and Ylvisaker (1978). To describe it, define $g_{b,C}(x) = g_{+,b,C}(x) + g_{-,b,C}(x)$ by

$$g_{+,b,C}(x) = \Big((b - b_- + \textstyle\sum_{j=1}^{p-1} d_{+,j} x^j - C|x|^p)_+ - (b - b_- + \sum_{j=1}^{p-1} d_{+,j} x^j + C|x|^p)_- \Big) 1(x \ge 0),$$
$$g_{-,b,C}(x) = -\Big((b_- + \textstyle\sum_{j=1}^{p-1} d_- x^j - C|x|^p)_+ - (b_- + \sum_{j=1}^{p-1} d_{-,j} x^j + C|x|^p)_- \Big) 1(x < 0),$$

where we use the notation $(t)_+ = \max\{t, 0\}$ and $(t)_- = -\min\{t, 0\}$. The solution is given by $g^*_{\delta,C} = g_{b(\delta),C}$ where the coefficients $d_+ = (d_{+,1}, \dots, d_{-,p-1})$, $d_- = (d_{-,1}, \dots, d_{-,p-1})$, and $b(\delta)$ and b_- solve a system of equations given below. To see that the solution must take the form $g_{b,C}(x)$ for some b, b_-, d_+, d_-, note that any function $f_+ \in \mathcal{F}_{T,p}(C)$ can be written as

$$f_+(x) = b_+ + \sum_{j=1}^{p-1} d_{+,j} x^j + r_+(x), \qquad |r_+(x)| \le C|x|^p. \tag{S4}$$

Given b_+, d_+, in order to minimize $|f_+(x_i)|$ simultaneously for all i, it must be that

$$r_+(x) = \begin{cases} -C|x|^p & \text{if } b_+ + \sum_{j=1}^{p=1} d_{+,j} x^j \ge C|x|^p, \\ -b_+ - \sum_{j=1}^{p=1} d_{+,j} x^j & \text{if } |b_+ + \sum_{j=1}^{p=1} d_{+,j} x^j| < C|x|^p, \\ C|x|^p & \text{if } b_+ + \sum_{j=1}^{p=1} d_{+,j} x^j \le -C|x|^p. \end{cases}$$

This form of $r(x)$ is necessary for f_+ to solve (S3): otherwise, one could strictly decrease $\sum_{i=1}^n [f_-(x_i)^2/\sigma^2(x_i) + f_+(x_i)^2/\sigma^2(x_i)]$, thereby making this quantity strictly less than $\delta^2/4$. But this would allow for a strictly larger value of $2(f_+(0) + f_-(0))$ by increasing b_+ and leaving d_+ and r_+ the same. Plugging $r_+(x)$ from the above display into (S4) shows that $f_+(x) = g_{+,b,C}(x)$ for some b_+, d_+. Similar arguments apply for f_-.

Setting up the Lagrangian for the problem with f constrained to the class of functions that take the form $g_{b,C}$ for some b, b_-, d_+, d_-, and taking first order conditions with respect to b_-, d_+ and d_- gives

$$0 = \sum_{i=1}^{n} \frac{g_{-,b,C}(x_i)}{\sigma^2(x_i)} \left(x_i, \ldots, x_i^{p-1}\right)', \tag{S5}$$

$$0 = \sum_{i=1}^{n} \frac{g_{+,b,C}(x_i)}{\sigma^2(x_i)} \left(x_i, \ldots, x_i^{p-1}\right)', \tag{S6}$$

$$0 = \sum_{i=1}^{n} \frac{g_{+,b,C}(x_i)}{\sigma^2(x_i)} + \sum_{i=1}^{n} \frac{g_{-,b,C}(x_i)}{\sigma^2(x_i)}. \tag{S7}$$

The constraint in (S3) must be binding at the optimum, which gives the additional equation

$$\delta^2/4 = \sum_{i=1}^{n} \frac{g_{b,C}(x_i)^2}{\sigma^2(x_i)} = b\sum_{i=1}^{n} \frac{g_{+,b,C}(x_i)}{\sigma^2(x_i)} - C\sum_{i=1}^{n} \frac{|g_{b,C}(x_i)||x_i|^p}{\sigma^2(x_i)}, \tag{S8}$$

where the second equality follows from (S5)–(S6). Note also that, since $g^*_{\delta,C} = g_{b(\delta),C}$ solves the modulus problem and gives the modulus as $2b(\delta)$, it also gives the solution to the inverse modulus problem

$$\frac{\omega^{-1}(2b; \mathcal{F}_{RDT,p}(C))^2}{4} = \inf_{f_+ - f_- \in \mathcal{F}_{RDT,p}(C)} \sum_{i=1}^{n} \left(\frac{f_+^2(x_i)}{\sigma^2(x_i)} + \frac{f_-^2(x_i)}{\sigma^2(x_i)} \right) \text{ s.t. } 2(f_+(0) - f_-(0)) \geq 2b \tag{S9}$$

for $b = b(\delta)$. Since the objective for the inverse modulus is strictly convex, this shows that the solution is unique up to the values at the x_is.

Using the fact that the class $\mathcal{F}_{RDT,p}(C)$ is translation invariant as defined in Supplemental Appendix D.2 (we can take $\iota(x) = c_0 + 1(x \geq 0)$ for any c_0), so that the derivative of the modulus is given by Lemma D.1, along with (S7) implies that the class of estimators $\hat{L}_\delta$ can be written as

$$\hat{L}_\delta = \hat{L}_{\delta,\mathcal{F}_{RDT,p}(C)} = \frac{\sum_{i=1}^{n} g^*_{+,\delta,C}(x_i)y_i/\sigma^2(x_i)}{\sum_{i=1}^{n} g^*_{+,\delta,C}(x_i)/\sigma^2(x_i)} - \frac{\sum_{i=1}^{n} g^*_{-,\delta,C}(x_i)y_i/\sigma^2(x_i)}{\sum_{i=1}^{n} g^*_{-,\delta,C}(x_i)/\sigma^2(x_i)}. \tag{S10}$$

Note that Conditions (S5), (S6), and (S7) are simply the conditions (S2) applied to this class of estimators.

To write the estimator $\hat{L}_\delta$ in the form (10), let $w_-(x_i, h_-) = g_{-,b,C}(x_i)/\sum_{i=1}^{n} g_{-,b,C}(x_i)$ and $w_+(x_i, h_+) = g_{+,b,C}(x_i)/\sum_{i=1}^{n} g_{+,b,C}(x_i)$, where d_+ and d_- solve (S5) and (S6) with $b - b_- = Ch_+^p$ and $b_- = Ch_-^p$. Then $\hat{L}_\delta = \hat{L}_{h_+(\delta),h_-(\delta)}$ where $h_+(\delta)$ and $h_-(\delta)$ are determined by the additional conditions (S7) and (S8).

To find the optimal estimators as described in Section 2.2, one can use the estimator $\hat{L}_{h_+,h_-}$ and optimize h_+ and h_- for the given performance criterion, using the variance and worst-case bias formulas given in that section. Since the optimal estimator $\hat{L}_\delta$ (with δ determined by the performance criterion) takes this form for some h_+ and h_-, the resulting estimator and CI will be the same as the one obtained by computing $\hat{L}_\delta$ with δ determined by solving the additional equation that corresponds to the performance criterion of interest.

E.3 Lower bound on C

While it is not possible to consistently estimate the smoothness constant C from the data, it is possible to lower bound its value. Here we develop a simple estimator and lower CI for this bound, focusing on the case $f \in \mathcal{F}_{RDT,2}(C)$.

As noted in Supplemental Appendix E.2, we can write $f_+(x) = f_+(0) + f_+'(0)x + r_+(x)$, where $|r_+(x)| \leq Cx^2$. It therefore follows that for any three points $0 \leq x_1 \leq x_2 \leq x_3$,

$$\lambda f_+(x_1) + (1-\lambda)f_+(x_3) - f_+(x_2) = \lambda r_+(x_1) + (1-\lambda)r_+(x_3) - r_+(x_2),$$

where $\lambda = (x_3 - x_2)/(x_3 - x_1)$. The left-hand side measures the curvature of f by comparing $f(x_2)$ to an approximation based on linearly interpolating between $f(x_1)$ and $f(x_3)$. Since $|r_+(x)| \leq Cx^2$, the right-hand side is bounded by $C(\lambda x_1^2 + (1-\lambda)x_3^3 + x_2^2)$. Taking averages of the preceding display over intervals $I_k = [a_{k-1}, a_k)$ where $a_0 \leq a_1 \leq a_2 \leq a_3$ and applying this bound yields the lower bound

$$C \geq |\mu_+|, \qquad \mu_+ = \frac{\lambda E_{n,1}(f_+(x)) + (1-\lambda)E_{n,3}(f_+(x)) - E_{n,2}(f_+(x))}{\lambda E_{n,1}(x^2) + (1-\lambda)E_{n,3}(x^2) + E_{n,2}(x^2)},$$

where we use the notation $E_{n,k}(g(x)) = \sum_i 1(x_i \in I_k)g(x_i)/n_k$, $n_k = \sum_i 1(x_i \in I_k)g(x_i)$ to denote sample average over I_k. Replacing $E_{n,k}(f_+(x))$ with $E_{n,k}(y)$ yields the estimator of μ_+

$$Z = \frac{\lambda E_{n,1}(y) + (1-\lambda)E_{n,3}(y) - E_{n,2}(y)}{\lambda E_{n,1}(x^2) + (1-\lambda)E_{n,3}(x^2) + E_{n,2}(x^2)} \sim \mathcal{N}\left(\mu_+, \tau^2\right),$$

where $\tau^2 = \frac{\lambda^2 E_{n,1}(\sigma^2(x))/n_1 + (1-\lambda)^2 E_{n,3}(\sigma^2(x))/n_3 - E_{n,2}(\sigma^2(x))/n_2}{(\lambda E_{n,1}(x^2) + (1-\lambda)E_{n,3}(x^2) + E_{n,2}(x^2))^2}$. Inverting tests of the hypotheses $H_0\colon |\mu_+| \leq \mu_0$ against $H_1\colon |\mu_+| > \mu_0$ then yields a one-sided CI for $|\mu_+|$ of the form $[\hat{\mu}_{+,\alpha}, \infty)$, where $\hat{\mu}_{+,\alpha}$ solves $|Z/\tau| = \mathrm{cv}_\alpha(\mu/\tau)$, with the convention that $\hat{\mu}_{+,\alpha} = 0$ if $|Z/\tau| \leq \mathrm{cv}_\alpha(0)$. This CI can be used as a lower CI for C in model specification checks.

Since unbiased estimates of the lower bound $|\mu_+|$ do not exist, following Chernozhukov et al. (2013), we take $\hat{\mu}_{+,0.5}$ as an estimator of the lower bound, which has the property that it's half-median unbiased in the sense that $P(|\mu_+| \leq \hat{\mu}_{+,0.5}) \leq 0.5$. An analogous bound obtains by considering intervals below the cutoff. We leave the question of optimal choice of the intervals I_k to future research. In the Lee (2008) application, we set $a_0 = 0$, and set the remaining interval endpoints a_k such that each interval I_k contains 200 observations. This yields estimates $\hat{\mu}_{+,0.5} = 0.008$ and $\hat{\mu}_{-,0.5} = 0.017$.

E.4 Asymptotic validity

We now give a theorem showing asymptotic validity of CIs from Section 2.2 under an unknown error distribution. We consider uniform validity over regression functions in $\mathcal{F}$ and error distributions in a sequence $\mathcal{Q}_n$, and we index probability statements with $f \in \mathcal{F}$ and $Q \in \mathcal{Q}_n$. We make the following assumptions on the x_is and the class of error distributions $\mathcal{Q}_n$.

Assumption E.1. *For some $p_{X,+}(0) > 0$ and $p_{X,-}(0) > 0$, the sequence $\{x_i\}_{i=1}^n$ satisfies $\frac{1}{nh_n}\sum_{i=1}^n m(x_i/h_n)1(x_i \geq 0) \to p_{X,+}(0)\int_0^\infty m(u)\,du$ and $\frac{1}{nh_n}\sum_{i=1}^n m(x_i/h_n)1(x_i < 0) \to p_{X,-}(0)\int_{-\infty}^0 m(u)\,du$ for any bounded function m with bounded support and any h_n with $0 < \liminf_n h_n n^{1/(2p+1)} \leq \limsup_n h_n n^{1/(2p+1)} < \infty$.*

Assumption E.2. *For some $\sigma(x)$ with $\lim_{x\downarrow 0}\sigma(x) = \sigma_+(0) > 0$ and $\lim_{x\uparrow 0}\sigma(x) = \sigma_-(0) > 0$,*

(i) the u_is are independent under any $Q \in \mathcal{Q}_n$ with $E_Q u_i = 0$, $var_Q(u_i) = \sigma^2(x_i)$

(ii) for some $\eta > 0$, $E_Q|u_i|^{2+\eta}$ is bounded uniformly over n and $Q \in \mathcal{Q}_n$.

While the variance function $\sigma^2(x)$ is unknown, the definition of $\mathcal{Q}_n$ is such that the variance function is the same for all $Q \in \mathcal{Q}_n$. This is done for simplicity. One could consider uniformity over classes $\mathcal{Q}_n$ that place only smoothness conditions on $\sigma^2(x)$ at the cost of introducing additional notation and making the optimality statements more cumbersome.

The estimators and CIs that we consider in the sequel are based on an estimate $\hat{\sigma}(x)$ of the conditional variance in Step 1 of the procedure in Section 2.2. We make the following assumption on this estimate.

Assumption E.3. *The estimate $\hat{\sigma}(x)$ is given by $\hat{\sigma}(x) = \hat{\sigma}_+(0)1(x \geq 0) + \hat{\sigma}_-(0)1(x < 0)$ where $\hat{\sigma}_+(0)$ and $\hat{\sigma}_-(0)$ are consistent for $\sigma_+(0)$ and $\sigma_-(0)$ uniformly over $f \in \mathcal{F}$ and $Q \in \mathcal{Q}_n$.*

For asymptotic coverage, we consider uniformity over both $\mathcal{F}$ and $\mathcal{Q}_n$. Thus, a confidence set $\mathcal{C}$ is said to have asymptotic coverage at least $1-\alpha$ if

$$\liminf_{n\to\infty} \inf_{f\in\mathcal{F}, Q\in\mathcal{Q}_n} P_{f,Q}\left(Lf \in \mathcal{C}\right) \geq 1-\alpha.$$

Theorem E.1. *Under Assumptions E.1, E.2 and E.3, CIs given in Section 2.2 based on $\hat{L}_\delta$ have asymptotic coverage at least $1-\alpha$. CIs based on local polynomial estimators have asymptotic coverage at least $1-\alpha$ so long as the kernel is bounded and uniformly continuous with bounded support and the bandwidths h_+ and h_- satisfy $h_+ n^{1/(2p+1)} \to h_{+,\infty}$ and $h_- n^{1/(2p+1)} \to h_{-,\infty}$ for some $h_{+,\infty} > 0$ and $h_{-,\infty} > 0$.*

Let $\hat{\chi}$ denote the half-length of the optimal fixed-length CI based on $\hat{\sigma}(x)$. For χ_∞ given in Supplemental Appendix H, the scaled half-length $n^{p/(2p+1)}\hat{\chi}$ converges in probability to χ_∞ uniformly over $\mathcal{F}$ and $\mathcal{Q}_n$. If, in addition, each $\mathcal{Q}_n$ contains a distribution where the u_is are normal, then for any sequence of confidence sets $\mathcal{C}$ with asymptotic coverage at least $1-\alpha$, we have the following bound on the asymptotic efficiency improvement at any $f \in \mathcal{F}_{RDT,p}(0)$

$$\liminf_{n\to\infty} \sup_{Q\in\mathcal{Q}_n} \frac{n^{p/(2p+1)} E_{f,Q}\lambda(\mathcal{C})}{2\chi_\infty} \geq \frac{(1-\alpha)2^r E[(z_{1-\alpha}-Z)^r \mid Z \leq z_{1-\alpha}]}{2r \inf_{\delta>0} \mathrm{cv}_\alpha\left((\delta/2)(1/r-1)\right)\delta^{r-1}},$$

where $Z \sim \mathcal{N}(0,1)$ and $r = 2p/(2p+1)$.

Letting $\hat{c}_{\alpha,\delta}$ denote the lower endpoint of the one-sided CI corresponding to $\hat{L}_\delta$, the CI $[\hat{c}_{\alpha,\delta}, \infty)$ has asymptotic coverage at least $1-\alpha$. If δ is chosen to minimax the β quantile excess length, (i.e. $\delta = z_\beta + z_{1-\alpha}$), then, if each $\mathcal{Q}_n$ contains a distribution where the u_is are normal, any other one-sided CI $[\hat{c}, \infty)$ with asymptotic coverage at least $1-\alpha$ must satisfy the efficiency bound

$$\liminf_{n\to\infty} \frac{\sup_{f\in\mathcal{F}, Q\in\mathcal{Q}_n} q_{f,Q,\beta}\left(Lf - \hat{c}\right)}{\sup_{f\in\mathcal{F}, Q\in\mathcal{Q}_n} q_{f,Q,\beta}\left(Lf - \hat{c}_{\alpha,\delta}\right)} \geq 1.$$

In addition, we have the following bound on the asymptotic efficiency improvement at any $f \in \mathcal{F}_{RDT,p}(0)$:

$$\liminf_{n\to\infty} \frac{\sup_{Q\in\mathcal{Q}_n} q_{f,Q,\beta}\left(Lf - \hat{c}\right)}{\sup_{Q\in\mathcal{Q}_n} q_{f,Q,\beta}\left(Lf - \hat{c}_{\alpha,\delta}\right)} \geq \frac{2^r}{1+r}.$$

The proof of Theorem E.1 is given in Supplemental Appendix H. The asymptotic efficiency bounds correspond to those in Section 3 under (29) with $r = 2p/(2p+1)$.

Appendix F Unknown Error Distribution

The Gaussian regression model (1) makes the assumption of normal i.i.d. errors with a known variance conditional on the x_i's, which is often unrealistic. This section considers a model that relaxes these assumptions on the error distribution:

$$y_i = f(x_i) + u_i,\ \{u_i\}_{i=1}^n \sim Q,\ f \in \mathcal{F},\ Q \in \mathcal{Q}_n \tag{S11}$$

where $\mathcal{Q}_n$ denotes the set of possible joint distributions of $\{u_i\}_{i=1}^n$ and, as before, $\{x_i\}_{i=1}^n$ is deterministic and $\mathcal{F}$ is a convex set. We derive feasible versions of the optimal CIs in Section 3 and show their asymptotic validity (uniformly over $\mathcal{F}, \mathcal{Q}_n$) and asymptotic efficiency. As we discuss below, our results hold even in cases where the limiting form of the optimal estimator is unknown or may not exist, and where currently available methods for showing asymptotic efficiency, such as equivalence with Gaussian white noise, break down.

Since the distribution of the data $\{y_i\}_{i=1}^n$ now depends on both f and Q, we now index probability statements by both of these quantities: $P_{f,Q}$ denotes the distribution under (f, Q) and similarly for $E_{f,Q}$. The coverage requirements and definitions of minimax performance criteria in Section 3 are the same, but with infima and suprema over functions f now taken over both functions f and error distributions $Q \in \mathcal{Q}_n$. We will also consider asymptotic results. We use the notation $Z_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{d}{\to}} \mathcal{L}$ to mean that Z_n converges in distribution to $\mathcal{L}$ uniformly over $f \in \mathcal{F}$ and $Q \in \mathcal{Q}_n$, and similarly for $\underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}}$.

If the variance function is unknown, the estimator $\hat{L}_\delta$ is infeasible. However, we can form an estimate based on an estimate of the variance function, or based on some candidate variance function. For a candidate variance function $\tilde{\sigma}^2(\cdot)$, let $K_{\tilde{\sigma}(\cdot),n} f = (f(x_1)/\tilde{\sigma}(x_1), \dots, f(x_n)/\tilde{\sigma}(x_n))'$, and let $\omega_{\tilde{\sigma}(\cdot),n}(\delta)$ denote the modulus of continuity defined with this choice of K. Let $\hat{L}_{\delta,\tilde{\sigma}(\cdot)} = \hat{L}_{\delta,\mathcal{F},\mathcal{G},\tilde{\sigma}(\cdot)}$ denote the estimator defined in (23) with this choice of K and $Y = (y_1/\tilde{\sigma}(x_1), \dots, y_n/\tilde{\sigma}(x_n))'$, and let $f^*_{\tilde{\sigma}(\cdot),\delta}$ and $g^*_{\tilde{\sigma}(\cdot),\delta}$ denote the least favorable functions used in forming this estimate. We assume throughout this section that $\mathcal{G} \subseteq \mathcal{F}$. More generally, we will consider affine estimators, which, in this setting, take the form

$$\hat{L} = a_n + \sum_{i=1}^n w_{i,n} y_i \tag{S12}$$

where a_n and $w_{i,n}$ are a sequence and triangular array respectively. For now, we assume that a_n and $w_{i,n}$ are nonrandom, (which, in the case of the estimator $\hat{L}_{\delta,\tilde{\sigma}(\cdot)}$, requires that

$\tilde{\sigma}(\cdot)$ and δ be nonrandom). We provide conditions that allow for random a_n and $w_{i,n}$ after stating our result for nonrandom weights. For a class $\mathcal{G}$, the maximum and minimum bias are

$$\overline{\text{bias}}_{\mathcal{G}}(\hat{L}) = \sup_{f \in \mathcal{G}} \left[a_n + \sum_{i=1}^{n} w_{i,n} f(x_i) - Lf \right], \quad \underline{\text{bias}}_{\mathcal{G}}(\hat{L}) = \inf_{f \in \mathcal{G}} \left[a_n + \sum_{i=1}^{n} w_{i,n} f(x_i) - Lf \right].$$

By the arguments used to derive the formula (24), we have

$$\overline{\text{bias}}_{\mathcal{F}}(\hat{L}_{\delta,\mathcal{F},\mathcal{G},\tilde{\sigma}(\cdot)}) = -\underline{\text{bias}}_{\mathcal{G}}(\hat{L}_{\delta,\mathcal{F},\mathcal{G},\tilde{\sigma}(\cdot)}) = \frac{1}{2}(\omega_{n,\tilde{\sigma}(\cdot)}(\delta; \mathcal{F}, \mathcal{G}) - \delta \omega'_{n,\tilde{\sigma}(\cdot)}(\delta; \mathcal{F}, \mathcal{G})).$$

This holds regardless of whether $\tilde{\sigma}(\cdot)$ is equal to the actual variance function of the u_i's. In our results below, we allow for infeasible estimators in which a_n and $w_{i,n}$ depend on Q (for example, when the unknown variance $\sigma_Q(x_i) = \text{var}_Q(y_i)$ is used to compute the optimal weights), so that $\overline{\text{bias}}_{\mathcal{G}}(\hat{L})$ and $\underline{\text{bias}}_{\mathcal{G}}(\hat{L})$ may depend on Q. We leave this implicit in our notation.

Let $s_{n,Q}$ denote the (constant over f) standard deviation of $\hat{L}$ under Q and suppose that the uniform central limit theorem

$$\frac{\sum_{i=1}^{n} w_{i,n} u_i}{s_{n,Q}} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{d}{\to}} \mathcal{N}(0,1) \tag{S13}$$

holds. To form a feasible CI, we will require an estimate $\widehat{\text{se}}_n$ of $s_{n,Q}$ satisfying

$$\frac{\widehat{\text{se}}_n}{s_{n,Q}} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 1. \tag{S14}$$

The following theorem shows that using $\widehat{\text{se}}_n$ to form analogues of the CIs treated in Section 3 gives asymptotically valid CIs.

Theorem F.1. *Let $\hat{L}$ be an estimator of the form (S12), and suppose that (S13) and (S14) hold. Let $\hat{c} = \hat{L} - \overline{bias}_{\mathcal{F}}(\hat{L}) - \widehat{se}_n z_{1-\alpha}$, and let $b = \max\{|\overline{bias}_{\mathcal{F}}(\hat{L})|, |\underline{bias}_{\mathcal{F}}(\hat{L})|\}$. Then*

$$\liminf_{n \to \infty} \inf_{f \in \mathcal{F}, Q \in \mathcal{Q}_n} P_{f,Q}\left(Lf \in [\hat{c}, \infty)\right) \geq 1 - \alpha \tag{S15}$$

and

$$\liminf_{n \to \infty} \inf_{f \in \mathcal{F}, Q \in \mathcal{Q}_n} P_{f,Q}\left(Lf \in \left\{\hat{L} \pm \widehat{se}_n \, \text{cv}_\alpha\left(b/\widehat{se}_n\right)\right\}\right) \geq 1 - \alpha. \tag{S16}$$

The worst-case βth quantile excess length of the one-sided CI over $\mathcal{G}$ will satisfy

$$\limsup_{n\to\infty} \sup_{Q\in\mathcal{Q}_n} \frac{\sup_{g\in\mathcal{G}} q_{g,Q,\beta}(Lg-\hat{c})}{\overline{bias}_{\mathcal{F}}(\hat{L}) - \underline{bias}_{\mathcal{G}}(\hat{L}) + s_{n,Q}(z_{1-\alpha}+z_\beta)} \leq 1 \tag{S17}$$

and the length of the two-sided CI will satisfy

$$\frac{\text{cv}_\alpha\left(b/\widehat{se}_n\right)\widehat{se}_n}{\text{cv}_\alpha\left(b/s_{n,Q}\right)s_{n,Q}} \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 1.$$

Suppose, in addition, that $\hat{L} = \hat{L}_{\delta,\mathcal{F},\mathcal{G},\tilde{\sigma}(\cdot)}$ with $\tilde{\sigma}(\cdot) = \sigma_Q(\cdot)$ where $\sigma_Q^2(x_i) = var_Q(u_i)$ and, for each n, there exists a $Q_n \in \mathcal{Q}_n$ such that $\{u_i\}_{i=1}^n$ are independent and normal under Q_n. Then no one-sided CI satisfying (S15) can satisfy (S17) with the constant 1 replaced by a strictly smaller constant on the right-hand side.

Proof. Let $Z_n = \sum_{i=1}^n w_{i,n}u_i/\widehat{\text{se}}_n$, and let Z denote a standard normal random variable. To show asymptotic coverage of the one-sided CI, note that

$$P_{f,Q}\left(Lf\in[\hat{c},\infty)\right) = P_{f,Q}\left(\widehat{\text{se}}_n z_{1-\alpha} \geq \hat{L} - Lf - \overline{\text{bias}}_{\mathcal{F}}(\hat{L})\right) \geq P_{f,Q}\left(z_{1-\alpha}\geq Z_n\right)$$

using the fact that $\overline{\text{bias}}_{\mathcal{F}}(\hat{L}) + \sum_{i=1}^n w_{i,n}u_i \geq \hat{L} - Lf$ for all $f\in\mathcal{F}$ by the definition of $\overline{\text{bias}}_{\mathcal{F}}$. The right-hand side converges to $1-\alpha$ uniformly over $f\in\mathcal{F}$ and $Q\in\mathcal{Q}_n$ by (S13) and (S14). For the two-sided CI, first note that

$$\left|\frac{\text{cv}_\alpha\left(b/\widehat{\text{se}}_n\right)\widehat{\text{se}}_n}{\text{cv}_\alpha\left(b/s_{n,Q}\right)s_{n,Q}} - 1\right| = \left|\frac{\text{cv}_\alpha\left(b/\widehat{\text{se}}_n\right) - \text{cv}_\alpha\left(b/s_{n,Q}\right) + \text{cv}_\alpha\left(b/s_{n,Q}\right)\left(1 - s_{n,Q}/\widehat{\text{se}}_n\right)}{\text{cv}_\alpha\left(b/s_{n,Q}\right)\left(s_{n,Q}/\widehat{\text{se}}_n\right)}\right|$$

which converges to zero uniformly over $f\in\mathcal{F}, Q\in\mathcal{Q}_n$ since $\text{cv}_\alpha(t)$ is bounded from below and uniformly continuous with respect to t. Thus, $\frac{\text{cv}_\alpha(b/\widehat{\text{se}}_n)\widehat{\text{se}}_n}{\text{cv}_\alpha\left(b/s_{n,Q}\right)s_{n,Q}} \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 1$ as claimed. To show coverage of the two-sided CI, note that

$$P_{f,Q}\left(Lf\in\left\{\hat{L}\pm\text{cv}_\alpha\left(b/\widehat{\text{se}}_n\right)\widehat{\text{se}}_n\right\}\right) = P_{f,Q}\left(\left|\tilde{Z}_n + r\right| \leq \text{cv}_\alpha\left(b/s_{n,Q}\right)\cdot c_n\right)$$

where $c_n = \frac{\text{cv}_\alpha(b/\widehat{\text{se}}_n)\widehat{\text{se}}_n}{\text{cv}_\alpha\left(b/s_{n,Q}\right)s_{n,Q}}$, $\tilde{Z}_n = \sum_{i=1}^n w_{i,n}u_i/s_{n,Q}$ and $r = \left(a_n + \sum_{i=1}^n w_{i,n}f(x_i) - Lf\right)/s_{n,Q}$. By (S13) and the fact that $c_n \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 1$, this is equal to $P_{f,Q}\left(|Z+r|\leq\text{cv}_\alpha\left(b/s_{n,Q}\right)\right)$ (where $Z\sim\mathcal{N}(0,1)$) plus a term that converges to zero uniformly over f,Q (this can be seen by using the fact that convergence in distribution to a continuous distribution implies uniform convergence of the cdfs; see Lemma 2.11 in van der Vaart 1998). Since $|r|\leq b/s_{n,Q}$, (S16) follows.

To show (S17), note that,

$$
\begin{aligned}
Lg - \hat{c} = Lg - a_n - \sum_{i=1}^{n} w_{i,n} g(x_i) - \widehat{\text{se}}_n Z_n + \overline{\text{bias}}_{\mathcal{F}}(\hat{L}) + \widehat{\text{se}}_n z_{1-\alpha} \\
\leq \overline{\text{bias}}_{\mathcal{F}}(\hat{L}) - \underline{\text{bias}}_{\mathcal{G}}(\hat{L}) + \widehat{\text{se}}_n (z_{1-\alpha} - Z_n)
\end{aligned}
$$

for any $g \in \mathcal{G}$. Thus,

$$
\begin{aligned}
\frac{Lg - \hat{c}}{\overline{\text{bias}}_{\mathcal{F}}(\hat{L}) - \underline{\text{bias}}_{\mathcal{G}}(\hat{L}) + s_{n,Q}(z_{1-\alpha} + z_\beta)} - 1 &\leq \frac{\widehat{\text{se}}_n (z_{1-\alpha} - Z_n) - s_{n,Q}(z_{1-\alpha} + z_\beta)}{\overline{\text{bias}}_{\mathcal{F}}(\hat{L}) - \underline{\text{bias}}_{\mathcal{G}}(\hat{L}) + s_{n,Q}(z_{1-\alpha} + z_\beta)} \\
&= \frac{(\widehat{\text{se}}_n / s_{n,Q}) \cdot (z_{1-\alpha} - Z_n) - (z_{1-\alpha} + z_\beta)}{[\overline{\text{bias}}_{\mathcal{F}}(\hat{L}) - \underline{\text{bias}}_{\mathcal{G}}(\hat{L})]/s_{n,Q} + (z_{1-\alpha} + z_\beta)}.
\end{aligned}
$$

The β quantile of the above display converges to 0 uniformly over $f \in \mathcal{F}$ and $Q \in \mathcal{Q}_n$, which gives the result.

For the last statement, let $[\tilde{c}, \infty)$ be a sequence of CIs with asymptotic coverage $1-\alpha$. Let Q_n be the distribution from the conditions in the theorem, in which the u_i's are independent and normal. Then, by Theorem 3.1,

$$
\sup_{g \in \mathcal{F}} q_{f,Q_n,\beta}(\tilde{c} - Lg) \geq \omega_{\sigma_{Q_n}(\cdot),n}(\tilde{\delta}_n),
$$

where $\tilde{\delta}_n = z_{1-\alpha_n} + z_\beta$ and $1-\alpha_n$ is the coverage of $[\tilde{c}, \infty)$ over $\mathcal{F}, \mathcal{Q}_n$. When $\hat{L} = \hat{L}_{\delta,\mathcal{F},\mathcal{G},\sigma_Q(\cdot)}$, the denominator in (S17) for $Q = Q_n$ is equal to $\omega_{\sigma_{Q_n}(\cdot),n}(z_{1-\alpha} + z_\beta)$, which gives

$$
\frac{\sup_{g \in \mathcal{G}} q_{g,Q_n,\beta}(\hat{c} - Lg)}{\overline{\text{bias}}_{\mathcal{F}}(\hat{L}) - \underline{\text{bias}}_{\mathcal{G}}(\hat{L}) + s_{n,Q_n}(z_{1-\alpha} + z_\beta)} \geq \frac{\omega_{\sigma_{Q_n}(\cdot),n}(z_{1-\alpha_n} + z_\beta)}{\omega_{\sigma_{Q_n}(\cdot),n}(z_{1-\alpha} + z_\beta)}.
$$

If $\alpha_n \leq \alpha$, then $z_{1-\alpha_n} + z_\beta \geq z_{1-\alpha} - z_\beta$ so that the above display is greater than one by monotonicity of the modulus. If not, then by concavity, $\omega_{\sigma_{Q_n}(\cdot),n}(z_{1-\alpha_n} + z_\beta) \geq [\omega_{\sigma_{Q_n}(\cdot),n}(z_{1-\alpha} + z_\beta)/(z_{1-\alpha} + z_\beta)] \cdot (z_{1-\alpha_n} + z_\beta)$, so the above display is bounded from below by $(z_{1-\alpha_n} + z_\beta)/(z_{1-\alpha} + z_\beta)$, and the lim inf of this is at least one by the coverage requirement. □

The efficiency bounds in Theorem F.1 use the assumption that the class of possible distributions contains a normal law, as is often done in the literature on efficiency in nonparametric settings (see, e.g., Fan, 1993, pp. 205–206). We leave the topic of relaxing this assumption for future research.

Theorem F.1 requires that a known candidate variance function $\tilde{\sigma}(\cdot)$ and a known δ be used when forming CIs based on the estimate $\hat{L}_\delta$. However, the theorem does not require that the candidate variance function be correct in order to get asymptotic coverage, so long as the standard error $\widehat{\text{se}}_n$ is consistent. If it turns out that $\tilde{\sigma}(\cdot)$ is indeed the correct variance function, then it follows from the last part of the theorem that the resulting CI is efficient. In the special case where $\mathcal{F}$ imposes a (otherwise unconstrained) linear model, this corresponds to the common practice of using ordinary least squares with heteroskedasticity robust standard errors.

In some cases, one will want to use a data dependent $\tilde{\sigma}(\cdot)$ and δ in order to get efficient estimates with unknown variance. The asymptotic coverage and efficiency of the resulting CI can then be derived by showing equivalence with the infeasible estimator $\hat{L}_{\delta^*,\mathcal{F},\mathcal{G},\sigma_Q(\cdot)}$, where δ^* is chosen according to the desired performance criterion. The following theorem gives conditions for this asymptotic equivalence. We verify them for our regression discontinuity example in Supplemental Appendix H.

Theorem F.2. *Suppose that $\hat{L}$ and $\widehat{se}_n$ satisfy* (S13) *and* (S14). *Let $\tilde{L}$ and $\widetilde{se}_n$ be another estimator and standard error, and let $\widetilde{\overline{bias}}_n$ and $\widetilde{\underline{bias}}_n$ be (possibly data dependent) worst-case bias formulas for $\tilde{L}$ under $\mathcal{F}$. Suppose that*

$$\frac{\hat{L}-\tilde{L}}{s_{n,Q}} \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0, \qquad \frac{\overline{bias}_{\mathcal{F}}(\hat{L})-\widetilde{\overline{bias}}_n}{s_{n,Q}} \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0, \qquad \frac{\underline{bias}_{\mathcal{F}}(\hat{L})-\widetilde{\underline{bias}}_n}{s_{n,Q}} \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0, \qquad \frac{\widehat{se}_n}{\widetilde{se}_n} \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 1.$$

Let $\tilde{c} = \tilde{L} - \widetilde{\overline{bias}}_n - \widetilde{se}_n z_{1-\alpha}$, and let $\tilde{b} = \max\{|\widetilde{\overline{bias}}_n|, |\widetilde{\underline{bias}}_n|\}$. Then (S15) and (S16) hold with $\hat{c}$ replaced by $\tilde{c}$, $\hat{L}$ replaced by $\tilde{L}$, b replaced by $\tilde{b}$ and $\widehat{se}_n$ replaced by $\widetilde{se}_n$. Furthermore, the performance of the CIs is asymptotically equivalent in the sense that

$$\frac{\sup_{Q\in\mathcal{Q}_n}\sup_{g\in\mathcal{G}} q_{g,Q,\beta}(\tilde{c}-Lg)}{\sup_{Q\in\mathcal{Q}_n}\sup_{g\in\mathcal{G}} q_{g,Q,\beta}(\hat{c}-Lg)} \to 1 \text{ and } \frac{\text{cv}_\alpha(b/\widehat{se}_n)\widehat{se}_n}{\text{cv}_\alpha(\tilde{b}/\widetilde{se}_n)\widetilde{se}_n} \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 1.$$

Proof. By the conditions of the theorem, we have, for some c_n that converges in probability to zero uniformly over $\mathcal{F}$, $\mathcal{Q}_n$,

$$\begin{aligned}\tilde{c} - Lf = \tilde{L} - Lf - \widetilde{\overline{\text{bias}}}_n - \widetilde{\text{se}}_n z_{1-\alpha} &= \hat{L} - Lf - \overline{\text{bias}}_{\mathcal{F}}(\hat{L}) - s_{n,Q} z_{1-\alpha} + c_n s_{n,Q} \\ &\leq \sum_{i=1}^n w_{i,n} u_i - s_{n,Q} z_{1-\alpha} + c_n s_{n,Q}.\end{aligned}$$

Thus,

$$P_{f,Q}\left(Lf \in [\tilde{c}, \infty)\right) = P_{f,Q}\left(0 \geq \tilde{c} - Lf\right) \geq P_{f,Q}\left(0 \geq \frac{\sum_{i=1}^n w_{i,n} u_i}{s_{n,Q}} - z_{1-\alpha} + c_n\right),$$

which converges to $1-\alpha$ uniformly over $\mathcal{F}, \mathcal{Q}_n$. By Theorem F.1, $\sup_{g\in\mathcal{G}} q_{g,Q,\beta}(\hat{c} - Lg)$ is bounded from below by a constant times $s_{n,Q}$. Thus, $\left|\frac{\sup_{Q\in\mathcal{Q}_n}\sup_{g\in\mathcal{G}} q_{g,Q,\beta}(\tilde{c}-Lg)}{\sup_{Q\in\mathcal{Q}_n}\sup_{g\in\mathcal{G}} q_{g,Q,\beta}(\hat{c}-Lg)} - 1\right|$ is bounded from above by a constant times

$$\sup_{Q\in\mathcal{Q}_n}\sup_{g\in\mathcal{G}}\left|\frac{q_{g,Q,\beta}(\tilde{c} - Lg) - q_{g,Q,\beta}(\hat{c} - Lg)}{s_{n,Q}}\right| = \sup_{Q\in\mathcal{Q}_n}\sup_{g\in\mathcal{G}}\left|q_{g,Q,\beta}(\tilde{c}/s_{n,Q}) - q_{g,Q,\beta}(\hat{c}/s_{n,Q})\right|,$$

which converges to zero since $(\tilde{c} - \hat{c})/s_{n,Q} \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0$.

The claim that $\frac{\mathrm{cv}_\alpha(b/\widehat{\mathrm{se}}_n)\widehat{\mathrm{se}}_n}{\mathrm{cv}_\alpha(\tilde{b}/\widetilde{\mathrm{se}}_n)\widetilde{\mathrm{se}}_n} \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 1$ follows using similar arguments to the proof of Theorem F.1. To show coverage of the two-sided CI, note that

$$P_{f,Q}\left(Lf \in \left\{\tilde{L} \pm \mathrm{cv}_\alpha\left(\tilde{b}/\widetilde{\mathrm{se}}_n\right)\widetilde{\mathrm{se}}_n\right\}\right) = P_{f,Q}\left(\frac{|\tilde{L} - Lf|}{s_{n,Q}} \leq \mathrm{cv}_\alpha\left(b/s_{n,Q}\right)\cdot c_n\right),$$

where $c_n = \frac{\mathrm{cv}_\alpha(\tilde{b}/\widetilde{\mathrm{se}}_n)\widetilde{\mathrm{se}}_n}{\mathrm{cv}_\alpha(b/s_{n,Q})s_{n,Q}} \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 1$. Since $\frac{|\tilde{L}-Lf|}{s_{n,Q}} = |V_n + r|$ where $r = (a_n + \sum_{i=1}^n w_{i,n} f(x_i) - Lf)/s_{n,Q}$ and $V_n = \sum_{i=1}^n w_{i,n}u_i/s_{n,Q} + (\tilde{L} - \hat{L})/s_{n,Q} \xrightarrow[\mathcal{F},\mathcal{Q}_n]{d} \mathcal{N}(0,1)$, the result follows from arguments in the proof of Theorem F.1. □

The results above give high-level conditions that can be applied to a wide range of estimators and CIs. We now introduce an estimator and standard error formula that give asymptotic coverage for essentially arbitrary functionals L under generic low level conditions on $\mathcal{F}$ and the x_i's. The estimator is based on a nonrandom guess for the variance function and, if this guess is correct up to scale (e.g. if the researcher correctly guesses that the errors are homoskedastic), the one-sided CI based on this estimator will be asymptotically optimal for some quantile of excess length.

Let $\tilde{\sigma}(\cdot)$ be some nonrandom guess for the variance function bounded away from 0 and ∞, and let $\delta > 0$ be a deterministic constant specified by the researcher. Let $\hat{f}$ be an estimator of f. The variance of $\hat{L}_{\delta,\tilde{\sigma}(\cdot)}$ under some $Q \in \mathcal{Q}_n$ is equal to

$$\mathrm{var}_Q(\hat{L}_{\delta,\tilde{\sigma}(\cdot),n}) = \left(\frac{\omega'_{\tilde{\sigma}(\cdot),n}(\delta)}{\delta}\right)^2 \sum_{i=1}^n \frac{(g^*_{\tilde{\sigma}(\cdot),\delta}(x_i) - f^*_{\tilde{\sigma}(\cdot),\delta}(x_i))^2\sigma_Q^2(x_i)}{\tilde{\sigma}^4(x_i)}.$$

We consider the estimate

$$\widehat{\mathrm{se}}^2_{\delta,\tilde{\sigma}(\cdot),n} = \left(\frac{\omega'_{\tilde{\sigma}(\cdot),n}(\delta)}{\delta}\right)^2 \sum_{i=1}^{n} \frac{(g^*_{\tilde{\sigma}(\cdot),\delta}(x_i) - f^*_{\tilde{\sigma}(\cdot),\delta}(x_i))^2 (y_i - \hat{f}(x_i))^2}{\tilde{\sigma}^4(x_i)}.$$

Suppose that $f : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is a metric space with metric d_X such that the functions $f^*_{\tilde{\sigma}(\cdot),\delta}$ and $g^*_{\tilde{\sigma}(\cdot),\delta}$ satisfy the uniform continuity condition

$$\sup_{n} \sup_{x,x' \colon d_X(x,x') \leq \eta} \max\left\{\left|f^*_{\tilde{\sigma}(\cdot),\delta}(x) - f^*_{\tilde{\sigma}(\cdot),\delta}(x')\right|, \left|g^*_{\tilde{\sigma}(\cdot),\delta}(x) - g^*_{\tilde{\sigma}(\cdot),\delta}(x')\right|\right\} \leq \overline{g}(\eta), \tag{S18}$$

where $\lim_{\eta \to 0} \overline{g}(\eta) = 0$ and, for all $\eta > 0$,

$$\min_{1 \leq i \leq n} \sum_{j=1}^{n} I\left(d_X(x_j, x_i) \leq \eta\right) \to \infty. \tag{S19}$$

We also assume that the estimator $\hat{f}$ used to form the variance estimate satisfies the uniform convergence condition

$$\max_{1 \leq i \leq n} |\hat{f}(x_i) - f(x_i)| \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0. \tag{S20}$$

Finally, we impose conditions on the moments of the error distribution. Suppose that there exist K and $\eta > 0$ such that, for all n, $Q \in \mathcal{Q}_n$, the errors $\{u_i\}_{i=1}^n$ are independent with, for each i,

$$1/K \leq \sigma^2_Q(x_i) \leq K \text{ and } E_Q|u_i|^{2+\eta} \leq K. \tag{S21}$$

In cases where function class $\mathcal{F}$ imposes smoothness on f, (S18) will often follow directly from the definition of $\mathcal{F}$. For example, it holds for the Lipschitz class $\{f \colon |f(x) - f(x')| \leq C d_X(x, x')\}$. The condition (S19) will hold with probability one if the x_i's are sampled from a distribution with density bounded away from zero on a sufficiently regular bounded support. The condition (S20) will hold under regularity conditions for a variety of choices of $\hat{f}$. It is worth noting that smoothness assumptions on $\mathcal{F}$ needed for this assumption are typically weaker than those needed for asymptotic equivalence with Gaussian white noise. For example, if $\mathcal{X} = \mathbb{R}^k$ with the Euclidean norm, (S18) will hold automatically for Hölder classes with exponent less than or equal to 1, while equivalence with Gaussian white noise requires that the exponent be greater than $k/2$ (see Brown and Zhang, 1998). Furthermore, we do not require any explicit characterization of the limiting form of the optimal CI. In particular, we do not require that the weights for the optimal estimator converge to a limiting optimal kernel or efficient influence function.

The condition (S21) is used to verify a Lindeberg condition for the central limit theorem used to obtain (S13), which we do in the next lemma.

Lemma F.1. *Let $Z_{n,i}$ be a triangular array of independent random variables and let $a_{n,j}$, $1 \leq j \leq n$ be a triangular array of constants. Suppose that there exist constants K and $\eta > 0$ such that, for all i,*

$$1/K \leq \sigma_{n,i}^2 \leq K \text{ and } E|Z_{n,i}|^{2+\eta} \leq K$$

where $\sigma_{n,i}^2 = EZ_{n,i}^2$, and that

$$\lim_{n\to\infty} \frac{\max_{1\leq j\leq n} a_{n,j}^2}{\sum_{j=1}^n a_{n,j}^2} = 0.$$

Then

$$\frac{\sum_{i=1}^n a_{n,i} Z_{n,i}}{\sqrt{\sum_{i=1}^n a_{n,i}^2 \sigma_{n,i}^2}} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof. We verify the conditions of the Lindeberg-Feller theorem as stated on p. 116 in Durrett (1996), with $X_{n,i} = a_{n,i}Z_{n,i}/\sqrt{\sum_{j=1}^n a_{n,j}^2 \sigma_j^2}$. To verify the Lindeberg condition, note that

$$\sum_{i=1}^n E\left(|X_{n,m}|^2 1(|X_{n,m}| > \varepsilon)\right) = \frac{\sum_{i=1}^n E\left[|a_{n,i}Z_{n,i}|^2 I\left(|a_{n,i}Z_{n,i}| > \varepsilon\sqrt{\sum_{j=1}^n a_{n,j}^2\sigma_j^2}\right)\right]}{\sum_{i=1}^n a_{n,i}^2 \sigma_{n,i}^2}$$
$$\leq \frac{\sum_{i=1}^n E\left(|a_{n,i}Z_{n,i}|^{2+\eta}\right)}{\varepsilon^\eta \left(\sum_{i=1}^n a_{n,i}^2\sigma_{n,i}^2\right)^{1+\eta/2}} \leq \frac{K^{2+\eta/2}}{\varepsilon^\eta} \frac{\sum_{i=1}^n |a_{n,i}|^{2+\eta}}{\left(\sum_{i=1}^n a_{n,i}^2\right)^{1+\eta/2}} \leq \frac{K^{2+\eta/2}}{\varepsilon^\eta}\left(\frac{\max_{1\leq i\leq n} a_{n,i}^2}{\sum_{i=1}^n a_{n,i}^2}\right)^{1+\eta/2}.$$

This converges to zero under the conditions of the lemma. □

Theorem F.3. *Let $\hat{L}_{\delta,\tilde{\sigma}(\cdot)}$ and $\widehat{se}^2_{\delta,\tilde{\sigma}(\cdot),n}$ be defined above. Suppose that, for each n, $f^*_{\tilde{\sigma}(\cdot),\delta}$, $g^*_{\tilde{\sigma}(\cdot),\delta}$ achieve the modulus under $\tilde{\sigma}(\cdot)$ with $\|K_{\tilde{\sigma}(\cdot),n}(g^*_{\tilde{\sigma}(\cdot),\delta} - f^*_{\tilde{\sigma}(\cdot),\delta})\| = \delta$, and that (S18) and (S19) hold. Suppose the errors satisfy* (S21) *and are independent over i for all n and $Q \in \mathcal{Q}_n$. Then* (S13) *holds. If, in addition, the estimator $\hat{f}$ satisfies (S20), then* (S14) *holds with $\widehat{se}_n$ given by $\widehat{se}_{\delta,\tilde{\sigma}(\cdot),n}$.*

Proof. Condition (S13) will follow by applying Lemma F.1 to show convergence under arbitrary sequences $Q_n \in \mathcal{Q}_n$ so long as

$$\frac{\max_{1\le i\le n}(g^*_{\tilde\sigma(\cdot),\delta}(x_i)-f^*_{\tilde\sigma(\cdot),\delta}(x_i))^2/\tilde\sigma(x_i)^4}{\sum_{i=1}^n(f^*_{\tilde\sigma(\cdot),\delta}(x_i)-g^*_{\tilde\sigma(\cdot),\delta}(x_i))^2/\tilde\sigma(x_i)^4}\to 0.$$

Since the denominator is bounded from below by $\delta^2/\max_{1\le i\le n}\tilde\sigma^2(x_i)$, and $\tilde\sigma^2(x_i)$ is bounded away from 0 and ∞ over i, it suffices to show that $\max_{1\le i\le n}(g^*_{\tilde\sigma(\cdot),\delta}(x_i)-f^*_{\tilde\sigma(\cdot),\delta}(x_i))^2\to 0$. To this end, suppose, to the contrary, that there exists some $c>0$ such that $\max_{1\le i\le n}(g^*_{\tilde\sigma(\cdot),\delta}(x_i)-f^*_{\tilde\sigma(\cdot),\delta}(x_i))^2>c^2$ infinitely often. Let η be small enough so that $\overline{g}(\eta)\le c/4$. Then, for n such that this holds and k_n achieving this maximum,

$$\sum_{i=1}^n(g^*_{\tilde\sigma(\cdot),\delta}(x_i)-f^*_{\tilde\sigma(\cdot),\delta}(x_i))^2\ge\sum_{i=1}^n(c-c/2)^2 1(d_X(x_i,x_{k_n})\le\eta)\to\infty.$$

But this is a contradiction since $\sum_{i=1}^n(g^*_{\tilde\sigma(\cdot),\delta}(x_i)-f^*_{\tilde\sigma(\cdot),\delta}(x_i))^2$ is bounded by a constant times $\sum_{i=1}^n(g^*_{\tilde\sigma(\cdot),\delta}(x_i)-f^*_{\tilde\sigma(\cdot),\delta}(x_i))^2/\tilde\sigma^2(x_i)=\delta^2$.

To show convergence of $\widehat{\text{se}}^2_{\delta,\tilde\sigma(\cdot),n}/\text{var}_Q(\hat L_{\delta,\tilde\sigma(\cdot)})$, note that

$$\frac{\widehat{\text{se}}^2_{\delta,\tilde\sigma(\cdot),n}}{\text{var}_Q(\hat L_{\delta,\tilde\sigma(\cdot)})}-1=\frac{\sum_{i=1}^n a_{n,i}\left[(y_i-\hat f(x_i))^2-\sigma_Q^2(x_i)\right]}{\sum_{i=1}^n a_{n,i}\sigma_Q^2(x_i)}$$

where $a_{n,i}=\frac{(g^*_{\tilde\sigma(\cdot),\delta}(x_i)-f^*_{\tilde\sigma(\cdot),\delta}(x_i))^2}{\tilde\sigma^4(x_i)}$. Since the denominator is bounded from below by a constant times $\sum_{i=1}^n a_{n,i}\tilde\sigma^2(x_i)=\delta^2$, it suffices to show that the numerator, which can be written as

$$\sum_{i=1}^n a_{n,i}\left[u_i^2-\sigma_Q(x_i)^2\right]+\sum_{i=1}^n a_{n,i}(f(x_i)-\hat f(x_i))^2+2\sum_{i=1}^n a_{n,i}u_i(f(x_i)-\hat f(x_i)),$$

converges in probability to zero uniformly over f and Q. The second term is bounded by a constant times $\max_{1\le i\le n}(f(x_i)-\hat f(x_i))^2\sum_{i=1}^n a_{n,i}\tilde\sigma^2(x_i)=\max_{1\le i\le n}(f(x_i)-\hat f(x_i))^2\delta^2$, which converges in probability to zero uniformly over f and Q by assumption. Similarly, the last term is bounded by $\max_{1\le i\le n}|f(x_i)-\hat f(x_i)|$ times $2\sum_{i=1}^n a_{n,i}|u_i|$, and the expectation of the latter term is bounded uniformly over $\mathcal{F}$ and $\mathcal{Q}$. Thus, the last term converges in probability to zero uniformly over f and Q as well. For the first term in this display, an inequality of von Bahr and Esseen (1965) shows that the expectation of the absolute $1+\eta/2$ moment of this term is bounded by a constant times

$$\sum_{i=1}^n a_{n,i}^{1+\eta/2} E_Q \left| u_i^2 - \sigma_Q(x_i)^2 \right|^{1+\eta/2} \le \left(\max_{1\le i\le n} a_{n,i}^{\eta/2} \right) \max_{1\le i\le n} E_Q \left| \varepsilon_i^2 - \sigma_Q^2(x_i) \right|^{1+\eta/2} \sum_{i=1}^n a_{n,i},$$

which converges to zero since $\max_{1\le i\le n} a_{n,i} \to 0$ as shown earlier in the proof and $\sum_{i=1}^n a_{n,i}$ is bounded by a constant times $\sum_{i=1}^n a_{n,i}\tilde{\sigma}^2(x_i) = \delta^2$. □

If the variance function used by the researcher is correct up to scale (for example, if the variance function is known to be constant), the one-sided confidence intervals in (F.3) will be asymptotically optimal for some level β, which depends on δ and the magnitude of the true error variance relative to the one used by the researcher. We record this as a corollary.

Corollary F.1. *If, in addition to the conditions in Theorem F.3, $\sigma_Q^2(x) = \sigma^2 \cdot \tilde{\sigma}^2(x)$ for all n and $Q \in \mathcal{Q}_n$, then, letting $\beta = \Phi(\delta/\sigma - z_{1-\alpha})$, no CI satisfying* (S15) *can satisfy S17 with the constant* 1 *replaced by a strictly smaller constant on the right-hand side.*

Appendix G Asymptotics for the Modulus and Efficiency Bounds

As discussed in Section 3, asymptotic relative efficiency comparisons can often be performed by calculating the limit of the scaled modulus. Here, we state some lemmas that can be used to obtain asymptotic efficiency bounds and limiting behavior of the value of δ that optimizes a particular performance criterion. We use these results in the proof of Theorem E.1 in Supplemental Appendix H.

Before stating these results, we recall the characterization of minimax affine performance given in Donoho (1994). To describe the results, first consider the normal model $Z \sim \mathcal{N}(\mu, 1)$ where $\mu \in [-\tau, \tau]$. The minimax affine mean squared error for this problem is

$$\rho_A(\tau) = \min_{\delta(Y) \text{ affine}} \max_{\mu \in [-\tau,\tau]} E_\mu(\delta(Y) - \mu)^2.$$

The solution is achieved by shrinking Y toward 0, namely $\delta(Y) = c_\rho(\tau)Y$, with $c_\rho(\tau) = \tau^2/(1+\tau^2)$, which gives $\rho_A(\tau) = \tau^2/(1+\tau^2)$. The length of the smallest fixed-length affine

$100 \cdot (1-\alpha)\%$ confidence interval is

$$\chi_{A,\alpha}(\tau) = \min\left\{\chi\colon \text{there exists } \delta(Y) \text{ affine s.t. } \inf_{\mu\in[-\tau,\tau]} P_\mu(|\delta(Y)-\mu| \le \chi) \ge 1-\alpha\right\}.$$

The solution is achieved at some $\delta(Y) = c_\chi(\tau)Y$, and it is characterized in Drees (1999).

Using these definitions, the minimax affine root MSE is given by

$$\sup_{\delta>0} \frac{\omega(\delta)}{\delta}\sqrt{\rho_A\left(\frac{\delta}{2\sigma}\right)}\sigma,$$

and the MSE optimal estimate is given by $\hat{L}_{\delta,\chi}$ where χ maximizes the above display. Similarly, the optimal fixed-length affine CI has half-length

$$\sup_{\delta>0} \frac{\omega(\delta)}{\delta}\chi_{A,\alpha}\left(\frac{\delta}{2\sigma}\right)\sigma,$$

and is centered at $\hat{L}_{\delta_\chi}$ where δ_χ maximizes the above display (it follows from our results and those of Donoho 1994 that this leads to the same value of δ_χ as the one obtained by minimizing CI length as described in Section 3.4).

The results below give the limiting behavior of these quantities as well as the bound on expected length in Corollary 3.3 under pointwise convergence of a sequence of functions $\omega_n(\delta)$ that satisfy the conditions of a modulus scaled by a sequence of constants.

Lemma G.1. *Let $\omega_n(\delta)$ be a sequence of concave nondecreasing nonnegative functions on $[0,\infty)$ and let $\omega_\infty(\delta)$ be a concave nondecreasing function on $[0,\infty)$ with range $[0,\infty)$. Then the following are equivalent.*

(i) For all $\delta > 0$, $\lim_{n\to\infty}\omega_n(\delta) = \omega_\infty(\delta)$.

(ii) For all $b \in (0,\infty)$, b is in the range of ω_n for large enough n, and $\lim_{n\to\infty}\omega_n^{-1}(b) = \omega_\infty^{-1}(b)$.

(iii) For any $\overline{\delta} > 0$, $\lim_{n\to\infty}\sup_{\delta\in[0,\overline{\delta}]}|\omega_n(\delta)-\omega_\infty(\delta)| = 0$.

Proof. Clearly *(iii)* $\Longrightarrow$ *(i)*. To show *(i)* $\Longrightarrow$ *(iii)*, given $\varepsilon > 0$, let $0 < \delta_1 < \delta_2 < \cdots < \delta_k = \overline{\delta}$ be such that $\omega(\delta_j) - \omega(\delta_{j-1}) \le \varepsilon$ for each j. Then, using monotonicity of ω_n and ω_∞, we have $\sup_{\delta\in[0,\delta_1]} |\omega_n(\delta)-\omega_\infty(\delta)| \le \max\{|\omega_n(\delta_1)|, |\omega_n(0)-\omega_\infty(\delta_1)|\} \to \omega_\infty(\delta_1)$ and

$$\begin{aligned}&\sup_{\delta\in[\delta_{j-1},\delta_j]} |\omega_n(\delta)-\omega_\infty(\delta)| \le \max\{|\omega_n(\delta_j)-\omega_\infty(\delta_{j-1})|, |\omega_n(\delta_{j-1})-\omega_\infty(\delta_j)|\}\\ &\to |\omega_\infty(\delta_{j-1})-\omega_\infty(\delta_j)| \le \varepsilon.\end{aligned}$$

The result follows since ε can be chosen arbitrarily small. To show $(i) \implies (ii)$, let δ_ℓ and δ_u be such that $\omega_\infty(\delta_\ell) < b < \omega_\infty(\delta_u)$. For large enough n, we will have $\omega_n(\delta_\ell) < b < \omega_n(\delta_u)$ so that b will be in the range of ω_n and $\delta_\ell < \omega_n^{-1}(b) < \delta_u$. Since ω_∞ is strictly increasing, δ_ℓ and δ_u can be chosen arbitrarily close to $\omega_\infty^{-1}(b)$, which gives the result. To show $(ii) \implies (i)$, let b_ℓ and b_u be such that $\omega_\infty^{-1}(b_\ell) < \delta < \omega_\infty^{-1}(b_u)$. Then, for large enough n, $\omega_n^{-1}(b_\ell) < \delta < \omega_n^{-1}(b_u)$, so that $b_\ell < \omega_n(\delta) < b_u$, and the result follows since b_ℓ and b_u can be chosen arbitrarily close to $\omega_\infty(\delta)$ since ω_∞^{-1} is strictly increasing. □

Lemma G.2. *Suppose that the conditions of Lemma G.1 hold with* $\lim_{\delta\to 0}\omega_\infty(\delta) = 0$ *and* $\lim_{\delta\to\infty}\omega_\infty(\delta)/\delta = 0$. *Let* r *be a nonnegative function with* $0 \le r(\delta/2) \le \overline{r}\min\{\delta, 1\}$ *for some* $\overline{r} < \infty$. *Then*

$$\lim_{n\to\infty}\sup_{\delta>0}\frac{\omega_n(\delta)}{\delta}r\left(\frac{\delta}{2}\right) = \sup_{\delta>0}\frac{\omega_\infty(\delta)}{\delta}r\left(\frac{\delta}{2}\right).$$

If, in addition r *is continuous,* $\frac{\omega_\infty(\delta)}{\delta}r\left(\frac{\delta}{2}\right)$ *has a unique maximizer* δ^**, and, for each* n*,* δ_n *maximizes* $\frac{\omega_n(\delta)}{\delta}r\left(\frac{\delta}{2}\right)$*, then* $\delta_n \to \delta^*$ *and* $\omega_n(\delta_n) \to \omega_\infty(\delta^*)$. *In addition, for any* $\sigma > 0$ *and* $0 < \alpha < 1$ *and* Z *a standard normal variable,*

$$\lim_{n\to\infty}(1-\alpha)E[\omega_n(2\sigma(z_{1-\alpha}-Z))|Z \le z_{1-\alpha}] = (1-\alpha)E[\omega_\infty(2\sigma(z_{1-\alpha}-Z))|Z \le z_{1-\alpha}].$$

Proof. We will show that the objective can be made arbitrarily small for δ outside of $[\underline{\delta}, \overline{\delta}]$ for $\underline{\delta}$ small enough and $\overline{\delta}$ large enough, and then use uniform convergence over $[\underline{\delta}, \overline{\delta}]$. First, note that, if we choose $\underline{\delta} < 1$, then, for $\delta \le \underline{\delta}$,

$$\frac{\omega_n(\delta)}{\delta}r\left(\frac{\delta}{2}\right) \le \omega_n(\delta)\overline{r} \le \omega_n(\underline{\delta})\overline{r} \to \omega_\infty(\underline{\delta}),$$

which can be made arbitrarily small by making $\underline{\delta}$ small. Since $\omega_n(\delta)$ is concave and nonnegative, $\omega_n(\delta)/\delta$ is nonincreasing, so, for $\delta > \overline{\delta}$,

$$\frac{\omega_n(\delta)}{\delta} r\left(\frac{\delta}{2}\right) \leq \frac{\omega_n(\delta)}{\delta}\overline{r} \leq \frac{\omega_n(\overline{\delta})}{\overline{\delta}}\overline{r} \to \frac{\omega_\infty(\overline{\delta})}{\overline{\delta}}\overline{r},$$

which can be made arbitrarily small by making $\overline{\delta}$ large. Applying Lemma G.1 to show convergence over $[\underline{\delta}, \overline{\delta}]$ gives the first claim. The second claim follows since $\underline{\delta}$ and $\overline{\delta}$ can be chosen so that $\delta_n \in [\underline{\delta}, \overline{\delta}]$ for large enough n (the assumption that $\frac{\omega_\infty(\delta)}{\delta} r\left(\frac{\delta}{2}\right)$ has a unique maximizer means that it is not identically zero), and uniform convergence to a continuous function with a unique maximizer on a compact set implies convergence of the sequence of maximizers to the maximizer of the limiting function.

For the last statement, note that, by positivity and concavity of ω_n, we have, for large enough n, $0 \leq \omega_n(\delta) \leq \omega_n(1)\max\{\delta, 1\} \leq (\omega_\infty(1) + 1)\max\{\delta, 1\}$ for all $\delta > 0$. The result then follows from the dominated convergence theorem. □

Lemma G.3. *Let $\omega_n(\delta)$ be a sequence of nonnegative concave functions on $[0, \infty)$ and let $\omega_\infty(\delta)$ be a nonnegative concave differentiable function on $[0, \infty)$. Let $\delta_0 > 0$ and suppose that $\omega_n(\delta) \to \omega_\infty(\delta)$ for all δ in a neighborhood of δ_0. Then, for any sequence $d_n \in \partial\omega_n(\delta_0)$, we have $d_n \to \omega'_\infty(\delta_0)$. In particular, if $\omega_n(\delta) \to \omega_\infty(\delta)$ in a neighborhood of δ_0 and $2\delta_0$, then $\frac{\omega_n(2\delta_0)}{\omega_n(\delta_0)+\delta_0\omega'_n(\delta_0)} \to \frac{\omega_\infty(2\delta_0)}{\omega_\infty(\delta_0)+\delta_0\omega'_\infty(\delta_0)}$.*

Proof. By concavity, for $\eta > 0$ we have $[\omega_n(\delta_0) - \omega_n(\delta_0 - \eta)]/\eta \geq d_n \geq [\omega_n(\delta_0 + \eta) - \omega_n(\delta_0)]/\eta$. For small enough η, the left and right-hand sides converge, so that $[\omega_\infty(\delta_0) - \omega_\infty(\delta_0 - \eta)]/\eta \geq \limsup_n d_n \geq \liminf_n d_n \geq [\omega_\infty(\delta_0 + \eta) - \omega_\infty(\delta_0)]/\eta$. Taking the limit as $\eta \to 0$ gives the result. □

Appendix H Asymptotics for Regression Discontinuity

This section proves Theorem E.1. We first give a general result for linear estimators under high-level conditions in Supplemental Appendix H.1. We then consider local polynomial estimators in Supplemental Appendix H.2 and optimal estimators with a plug-in variance estimate in Supplemental Appendix H.3. Theorem E.1 follows immediately from the results in these sections.

Throughout this section, we consider the RD setup where the error distribution may be non-normal as in Supplemental Appendix E.4, using the conditions in that section. We repeat these conditions here for convenience.

Assumption H.1. *For some $p_{X,+}(0) > 0$ and $p_{X,-}(0) > 0$, the sequence $\{x_i\}_{i=1}^n$ satisfies $\frac{1}{nh_n}\sum_{i=1}^n m(x_i/h_n)1(x_i > 0) \to p_{X,+}(0)\int_0^\infty m(u)\,du$ and $\frac{1}{nh_n}\sum_{i=1}^n m(x_i/h_n)1(x_i < 0) \to p_{X,-}(0)\int_{-\infty}^0 m(u)\,du$ for any bounded function m with bounded support and any h_n with $0 < \liminf_n h_n n^{1/(2p+1)} \leq \limsup_n h_n n^{1/(2p+1)} < \infty$.*

Assumption H.2. *For some $\sigma(x)$ with $\lim_{x\downarrow 0}\sigma(x) = \sigma_+(0) > 0$ and $\lim_{x\uparrow 0}\sigma(x) = \sigma_-(0) > 0$, we have*

(i) the u_is are independent under any $Q \in \mathcal{Q}_n$ with $E_Q u_i = 0$, $var_Q(u_i) = \sigma^2(x_i)$

(ii) for some $\eta > 0$, $E_Q|u_i|^{2+\eta}$ is bounded uniformly over n and $Q \in \mathcal{Q}_n$.

Theorem E.1 considers affine estimators that are optimal under the assumption that the variance function is given by $\hat{\sigma}_+1(x > 0) + \hat{\sigma}_-1(x < 0)$, which covers the plug-in optimal affine estimators used in our application. Here, it will be convenient to generalize this slightly by considering the class of affine estimators that are optimal under a variance function $\tilde{\sigma}(x)$, which may be misspecified or data-dependent, but which may take some other form. We consider two possibilities for how $\tilde{\sigma}(\cdot)$ is calibrated.

Assumption H.3. *$\tilde{\sigma}(x) = \hat{\sigma}_+1(x > 0) + \hat{\sigma}_-1(x < 0)$ where $\hat{\sigma}_+ \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \tilde{\sigma}_+(0) > 0$ and $\hat{\sigma}_- \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \tilde{\sigma}_-(0) > 0$.*

Assumption H.4. *$\tilde{\sigma}(x)$ is a deterministic function with $\lim_{x\downarrow 0}\tilde{\sigma}(x) = \tilde{\sigma}_-(0) > 0$ and $\lim_{x\uparrow 0}\tilde{\sigma}(x) = \tilde{\sigma}_+(0) > 0$.*

Assumption H.3 corresponds to the estimate of the variance function used in the application. It generalizes Assumption E.3 slightly by allowing $\hat{\sigma}_+$ and $\hat{\sigma}_-$ to converge to something other than the left- and right-hand limits of the true variance function. Assumption H.4 is used in deriving bounds based on infeasible estimates that use the true variance function.

Note that, under Assumption H.3, $\tilde{\sigma}_+(0)$ is defined as the probability limit of $\hat{\sigma}_+$ as $n \to \infty$, and does not give the limit of $\tilde{\sigma}(x)$ as $x \downarrow 0$ (and similarly for $\tilde{\sigma}_-(0)$). We use this notation so that certain limiting quantities can be defined in the same way under each of the Assumptions H.4 and H.3.

H.1 General Results for Kernel Estimators

We first state results for affine estimators where the weights asymptotically take a kernel form. We consider a sequence of estimators of the form

$$\hat{L} = \frac{\sum_{i=1}^n k_n^+(x_i/h_n)1(x_i > 0)y_i}{\sum_{i=1}^n k_n^+(x_i/h_n)1(x_i > 0)} - \frac{\sum_{i=1}^n k_n^-(x_i/h_n)1(x_i < 0)y_i}{\sum_{i=1}^n k_n^-(x_i/h_n)1(x_i < 0)}$$

where k_n^+ and k_n^- are sequences of kernels. The assumption that the same bandwidth is used on each side of the discontinuity is a normalization: it can always be satisfied by redefining one of the kernels k_n^+ or k_n^-. We make the following assumption on the sequence of kernels.

Assumption H.5. *The sequences of kernels and bandwidths k_n^+ and h_n satisfy*

(i) k_n^+ *has support bounded uniformly over* n*. For a bounded kernel* k^+ *with* $\int k^+(u)\,du > 0$*, we have* $\sup_x |k_n^+(x) - k^+(x)| \to 0$

(ii) $\frac{1}{nh_n}\sum_{i=1}^n k_n^+(x_i/h_n)1(x_i > 0)(x_i, \ldots, x_i^{p-1})' = 0$ *for each* n

(iii) $h_n n^{1/(2p+1)} \to h_\infty$ *for some constant* $0 < h_\infty < \infty$*,*

and similarly for k_n^- *for some* k^-*.*

Let

$$\begin{aligned}\overline{\text{bias}}_n &= \frac{\sum_{i=1}^n |k_n^+(x_i/h_n)|1(x_i > 0)C|x_i|^p}{\sum_{i=1}^n k_n^+(x_i/h_n)1(x_i > 0)} + \frac{\sum_{i=1}^n |k_n^-(x_i/h_n)|1(x_i < 0)C|x_i|^p}{\sum_{i=1}^n k_n^-(x_i/h_n)1(x_i < 0)} \\ &= Ch_n^p\left(\frac{\sum_{i=1}^n |k_n^+(x_i/h_n)|1(x_i > 0)|x_i/h_n|^p}{\sum_{i=1}^n k_n^+(x_i/h_n)1(x_i > 0)} + \frac{\sum_{i=1}^n |k_n^-(x_i/h_n)|1(x_i < 0)|x_i/h_n|^p}{\sum_{i=1}^n k_n^-(x_i/h_n)1(x_i < 0)}\right)\end{aligned}$$

and

$$\begin{aligned}v_n &= \frac{\sum_{i=1}^n k_n^+(x_i/h_n)^2 1(x_i > 0)\sigma^2(x_i)}{\left[\sum_{i=1}^n k_n^+(x_i/h_n)1(x_i > 0)\right]^2} + \frac{\sum_{i=1}^n k_n^-(x_i/h_n)^2 1(x_i < 0)\sigma^2(x_i)}{\left[\sum_{i=1}^n k_n^-(x_i/h_n)1(x_i < 0)\right]^2} \\ &= \frac{1}{nh_n}\left(\frac{\frac{1}{nh_n}\sum_{i=1}^n k_n^+(x_i/h_n)^2 1(x_i > 0)\sigma^2(x_i)}{\left[\frac{1}{nh_n}\sum_{i=1}^n k_n^+(x_i/h_n)1(x_i > 0)\right]^2} + \frac{\frac{1}{nh_n}\sum_{i=1}^n k_n^-(x_i/h_n)^2 1(x_i < 0)\sigma^2(x_i)}{\left[\frac{1}{nh_n}\sum_{i=1}^n k_n^-(x_i/h_n)1(x_i < 0)\right]^2}\right)\end{aligned}$$

Note that v_n is the (constant over $Q \in \mathcal{Q}_n$) variance of $\hat{L}$, and that, by arguments in Supplemental Appendix E.1, $\overline{\text{bias}}_n = \sup_{f\in\mathcal{F}}(E_{f,Q}\hat{L} - Lf) = -\inf_{f\in\mathcal{F}}(E_{f,Q}\hat{L} - Lf)$ for any $Q \in \mathcal{Q}_n$ under Assumption H.5 (ii).

To form a feasible CI, we need an estimate of v_n. While the results below go through with any variance estimate that is consistent uniformly over $f, \mathcal{Q}_n$, we propose one here for concreteness. For a possibly data-dependent guess $\tilde{\sigma}(\cdot)$ of the variance function, let $\tilde{v}_n$ denote v_n with $\sigma(\cdot)$ replaced by $\tilde{\sigma}(\cdot)$. We record the limiting behavior of $\overline{\text{bias}}_n$, v_n and $\tilde{v}_n$ in the following lemma. Let

$$\overline{\text{bias}}_\infty = Ch_\infty^p \left(\frac{\int_0^\infty |k^+(u)||u|^p \, du}{\int_0^\infty k^+(u) \, du} + \frac{\int_{-\infty}^0 |k^-(u)||u|^p \, du}{\int_{-\infty}^0 k^-(u) \, du} \right)$$

and

$$v_\infty = \frac{1}{h_\infty} \left(\frac{\sigma_+^2(0) \int_0^\infty k^+(u)^2 \, du}{p_{X,+}(0) \left[\int_0^\infty k^+(u) \, du \right]^2} + \frac{\sigma_-^2(0) \int_{-\infty}^0 k^-(u)^2 \, du}{p_{X,-}(0) \left[\int_{-\infty}^0 k^-(u) \, du \right]^2} \right).$$

Lemma H.1. *Suppose that Assumption H.1 holds. If Assumption H.5 also holds, then* $\lim_{n\to\infty} n^{p/(2p+1)} \overline{bias}_n = \overline{bias}_\infty$ *and* $\lim_{n\to\infty} n^{2p/(2p+1)} v_n = v_\infty$*. If, in addition,* $\tilde{\sigma}(\cdot)$ *satisfies Assumption H.3 or Assumption H.4 with* $\tilde{\sigma}_+(0) = \sigma_+(0)$ *and* $\tilde{\sigma}_-(0) = \sigma_-(0)$*, then* $n^{2p/(2p+1)} \tilde{v}_n \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} v_\infty$ *under Assumption H.3 and* $\lim_{n\to\infty} n^{2p/(2p+1)} \tilde{v}_n = v_\infty$ *under Assumption H.4.*

Proof. The results follow from applying the convergence in Assumption H.1 along with Assumption H.5(i) to the relevant terms in $\overline{\text{bias}}_n$ and $\tilde{v}_n$. □

Theorem H.1. *Suppose that Assumptions H.1, H.2 and H.5 hold, and that* $\tilde{v}_n$ *is formed using a variance function* $\tilde{\sigma}(\cdot)$ *that satisfies Assumption H.3 or H.4 with* $\tilde{\sigma}_+(0) = \sigma_+(0)$ *and* $\tilde{\sigma}_-(0) = \sigma_-(0)$*. Then*

$$\liminf_{n\to\infty} \inf_{f \in \mathcal{F}_{RDT,p}(C), Q \in \mathcal{Q}_n} P_{f,Q} \left(Lf \in \left\{ \hat{L} \pm \text{cv}_\alpha \left(\overline{bias}_n / \tilde{v}_n \right) \sqrt{\tilde{v}_n} \right\} \right) \geq 1 - \alpha$$

and, letting $\hat{c} = \hat{L} - \overline{bias}_n - z_{1-\alpha}\sqrt{\tilde{v}_n}$*,*

$$\liminf_{n\to\infty} \inf_{f \in \mathcal{F}_{RDT,p}(C), Q \in \mathcal{Q}_n} P_{f,Q} \left(Lf \in [\hat{c}, \infty) \right) \geq 1 - \alpha.$$

In addition, $n^{p/(2p+1)} \text{cv}_\alpha(\overline{bias}_n / \tilde{v}_n) \tilde{v}_n \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} \text{cv}_\alpha(\overline{bias}_\infty / v_\infty) v_\infty$ *if* $\tilde{\sigma}(\cdot)$ *satisfies Assumption H.3 and* $n^{p/(2p+1)} \text{cv}_\alpha(\overline{bias}_n / \tilde{v}_n) \tilde{v}_n \to \text{cv}_\alpha(\overline{bias}_\infty / v_\infty) v_\infty$ *if* $\tilde{\sigma}(\cdot)$ *satisfies Assumption H.4. The* *minimax β quantile of the one-sided CI satisfies*

$$\limsup_{n\to\infty} n^{p/(2p+1)} \sup_{f\in\mathcal{F}_{RDT,p}(C),Q\in\mathcal{Q}_n} q_{f,Q,\beta}(Lf-\hat{c}) \leq 2\overline{bias}_\infty + (z_\beta + z_{1-\alpha})\sqrt{v_\infty}.$$

The worst-case β quantile over $\mathcal{F}_{RDT,p}(0)$ satisfies

$$\limsup_{n\to\infty} n^{p/(2p+1)} \sup_{f\in\mathcal{F}_{RDT,p}(0),Q\in\mathcal{Q}_n} q_{f,Q,\beta}(Lf-\hat{c}) \leq \overline{bias}_\infty + (z_\beta + z_{1-\alpha})\sqrt{v_\infty}.$$

Furthermore, the same holds with $\hat{L}$, $\overline{bias}_n$ and $\tilde{v}_n$ replaced by any $\hat{L}^*$, $\overline{bias}_n^*$ and $\tilde{v}_n^*$ such that

$$n^{p/(2p+1)}\left(\hat{L}-\hat{L}^*\right) \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0, \qquad n^{p/(2p+1)}\left(\overline{bias}_n - \overline{bias}_n^*\right) \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0, \qquad \frac{\tilde{v}_n}{\tilde{v}_n^*} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 1.$$

Proof. We verify the conditions of Theorem F.1. Condition (S14) follows from Lemma H.1. To verify (S13), note that $\hat{L}$ takes the general form in Theorem F.1 with $w_{n,i}$ given by $w_{n,i} = k_n^+(x_i/h_n)/\sum_{j=1}^n k_n^+(x_j/h_n)1(x_j > 0)$ for $x_i > 0$ and $w_{n,i} = k_n^-(x_i/h_n)/\sum_{j=1}^n k_n^-(x_j/h_n)\cdot 1(x_j < 0)$ for $x_i < 0$. The uniform central limit theorem in (S13) with $w_{n,i}$ taking this form follows from Lemma F.1. This gives the asymptotic coverage statements.

For the asymptotic formulas for excess length of the one-sided CI and length of the two-sided CI, we apply Theorem F.2 with $n^{-p/(2p+1)}\overline{\text{bias}}_\infty$ playing the role of $\widetilde{\overline{\text{bias}}}_n$ and $n^{-p/(2p+1)}v_\infty$ playing the role of $\widetilde{\text{se}}_n$. Finally, the last statement of the theorem is immediate from Theorem F.2. □

H.2 Local Polynomial Estimators

The $(p-1)$th order local polynomial estimator of $f_+(0)$ based on kernel k_+^* and bandwidth $h_{+,n}$ is given by

$$\begin{aligned}\hat{f}_+(0) =& e_1'\left(\sum_{i=1}^n p(x_i/h_{+,n})p(x_i/h_{+,n})'k_+^*(x_i/h_{+,n})1(x_i>0)\right)^{-1} \\ & \sum_{i=1}^n k_+^*(x_i/h_{+,n})1(x_i>0)p(x_i/h_{+,n})y_i\end{aligned}$$

where $e_1 = (1,0,\ldots,0)'$ and $p(x) = (1,x,x^2,\ldots,x^{p-1})'$. Letting the local polynomial estimator of $f_-(0)$ be defined analogously for some kernel k_-^* and bandwidth $h_{-,n}$, the local polynomial estimator of $Lf = f_+(0) - f_-(0)$ is given by

$$\hat{L} = \hat{f}_+(0) - \hat{f}_-(0).$$

This takes the form given in Supplemental Appendix H.1, with $h_n = h_{n,+}$,

$$k_n^+(u) = e_1' \left(\frac{1}{nh_n} \sum_{i=1}^n p(x_i/h_{+,n}) p(x_i/h_{+,n})' k_+^*(x_i/h_{+,n}) 1(x_i > 0) \right)^{-1} k_+^*(u) p(u) 1(u > 0)$$

and

$$\begin{aligned} k_n^-(u) = & e_1' \left(\frac{1}{nh_n} \sum_{i=1}^n p(x_i/h_{-,n}) p(x_i/h_{-,n})' k_+^*(x_i/h_{-,n}) 1(x_i < 0) \right)^{-1} \\ & k_+^*(u(h_{n,+}/h_{n,-})) p(u(h_{n,+}/h_{n,-})) 1(u < 0). \end{aligned}$$

Let M^+ be the $(p-1)\times(p-1)$ matrix with $\int_0^\infty u^{j+k-2} k_+^*(u)$ as the i,jth entry, and let M^- be the $(p-1)\times(p-1)$ matrix with $\int_{-\infty}^0 u^{j+k-2} k_-^*(u)$ as the i,jth entry. Under Assumption H.1, for k_+^* and k_-^* bounded with bounded support, $\frac{1}{nh_n}\sum_{i=1}^n p(x_i/h_{+,n})p(x_i/h_{+,n})'k_+^*(x_i/h_{+,n}) \cdot 1(x_i > 0) \to M^+ p_{X,+}(0)$ and similarly $\frac{1}{nh_n}\sum_{i=1}^n p(x_i/h_{-,n})p(x_i/h_{-,n})'k_+^*(x_i/h_{-,n}) \cdot 1(x_i < 0) \to M^- p_{X,-}(0)$. Furthermore, Assumption H.5 (ii) follows immediately from the normal equations for the local polynomial estimator. This gives the following result.

Theorem H.2. *Let k_+^* and k_-^* be bounded and uniformly continuous with bounded support. Let $h_{n,+}n^{1/(2p+1)} \to h_\infty > 0$ and suppose $h_{n,-}/h_{n,+}$ converges to a strictly positive constant. Then Assumption H.5 holds for the local polynomial estimator so long as Assumption H.1 holds.*

H.3 Optimal Affine Estimators

We now consider the class of affine estimators that are optimal under the assumption that the variance function is given by $\tilde{\sigma}(\cdot)$, which satisfies either Assumption H.3 or Assumption H.4. We use the same notation as in Supplemental Appendix E, except that n and/or $\tilde{\sigma}(\cdot)$ are added as subscripts for many of the objects under consideration to make the dependence on $\{x_i\}_{i=1}^n$ and $\tilde{\sigma}(\cdot)$ explicit.

The modulus problem is given by Equation (S3) in Supplemental Appendix E.2 with $\tilde{\sigma}(\cdot)$ in place of $\sigma(\cdot)$. We use $\omega_{\tilde{\sigma}(\cdot),n}(\delta)$ to denote the modulus, or $\omega_n(\delta)$ when the context is clear. The corresponding estimator $\hat{L}_{\delta,\tilde{\sigma}(\cdot)}$ is then given by Equation (S10) in Supplemental Appendix E.2 with $\tilde{\sigma}(\cdot)$ in place of $\sigma(\cdot)$.

We will deal with the inverse modulus, and use Lemma G.1 to obtain results for the modulus itself. The inverse modulus $\omega^{-1}_{\tilde{\sigma}(\cdot),n}(2b)$ is given by Equation (S9) in Supplemental Appendix E.2, with $\tilde{\sigma}^2(x_i)$ in place of $\sigma^2(x_i)$, and the solution takes the form given in that section. Let $h_n = n^{-1/(2p+1)}$. We will consider a sequence $b = b_n$, and will define $\tilde{b}_n = n^{p/(2p+1)}b_n = h_n^{-p}b_n$. Under Assumption H.4, we will assume that $\tilde{b}_n \to \tilde{b}_\infty$ for some $\tilde{b}_\infty > 0$. Under Assumption H.3, we will assume that $\tilde{b}_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \tilde{b}_\infty$ for some $\tilde{b}_\infty > 0$. We will then show that this indeed holds for $2b_n = \omega_{\tilde{\sigma}(\cdot),n}(\delta_n)$ with δ_n chosen as in Theorem H.3 below.

Let $\tilde{b}_n = n^{p/(2p+1)}b_n = h_n^{-p}b_n$, $\tilde{b}_{-,n} = n^{p/(2p+1)}b_{-,n} = h_n^{-p}b_{-,n}$, $\tilde{d}_{+,j,n} = n^{(p-j)/(2p+1)}d_{+,j,n} = h_n^{j-p}d_{+,j,n}$ and $\tilde{d}_{-,j,n} = n^{(p-j)/(2p+1)}d_{-,j,n} = h_n^{j-p}d_{-,j,n}$ for $j = 1,\dots,p-1$, where b_n, $b_{-,n}$, $d_{+,n}$, and d_{-n} correspond to the function $g_{b,C}$ that solves the inverse modulus problem, given in Supplemental Appendix E.2. These values of $\tilde{b}_{+,n}$, $\tilde{b}_{-,n}$, $\tilde{d}_{+,n}$ and $\tilde{d}_{-,n}$ minimize $G_n(b_+, b_-, d_+, d_-)$ subject to $b_+ + b_- = \tilde{b}_n$ where, letting $\mathcal{A}(x_i, b, d) = h_n^p b + \sum_{j=1}^{p-1} h_n^{p-j} d_j x_i^j$,

$$
\begin{aligned}
&G_n(b_+, b_-, d_+, d_-) = \\
&\qquad \sum_{i=1}^n \tilde{\sigma}^{-2}(x_i)\left((\mathcal{A}(x_i, b_+, d_+) - C|x_i^p|)_+ + (\mathcal{A}(x_i, b_+, d_+) + C|x_i|^p)_-\right)^2 1(x_i > 0) \\
&\qquad + \sum_{i=1}^n \tilde{\sigma}^{-2}(x_i)\left((\mathcal{A}(x_i, b_-, d_-) - C|x_i|^p) + (\mathcal{A}(x_i, b_-, d_-) + C|x_i|^p)_-\right)^2 1(x_i < 0) \\
&\qquad\qquad = \frac{1}{nh_n}\sum_{i=1}^n k^+_{\tilde{\sigma}(\cdot)}(x_i/h_n; b_+, d_+)^2\tilde{\sigma}^2(x_i) + \frac{1}{nh_n}\sum_{i=1}^n k^-_{\tilde{\sigma}(\cdot)}(x_i/h_n; b_-, d_-)^2\tilde{\sigma}^2(x_i)
\end{aligned}
$$

with

$$
k^+_{\tilde{\sigma}(\cdot)}(u; b, d) = \tilde{\sigma}^{-2}(uh_n)\left(\left(b + \sum_{j=1}^{p-1} d_j u^j - C|u|^p\right)_+ - \left(b + \sum_{j=1}^{p-1} d_j u^j + C|u|^p\right)_-\right) 1(u > 0),
$$

$$
k^-_{\tilde{\sigma}(\cdot)}(u; b, d) = \tilde{\sigma}^{-2}(uh_n)\left(\left(b + \sum_{j=1}^{p-1} d_j u^j - C|u|^p\right)_+ - \left(b + \sum_{j=1}^{p-1} d_j u^j + C|u|^p\right)_-\right) 1(u < 0).
$$

We use the notation k^+_c for a scalar c to denote $k^+_{\tilde{\sigma}(\cdot)}$ where $\tilde{\sigma}(\cdot)$ is given by the constant function $\tilde{\sigma}(x) = c$.

With these definitions, the estimator $\hat{L}_{\delta,\tilde{\sigma}(\cdot)}$ with $\omega_{\tilde{\sigma}(\cdot),n}(\delta) = 2b_n$ takes the general kernel form in Supplemental Appendix H.1 with $k_n^+(u) = k^+_{\tilde{\sigma}(\cdot)}(u; \tilde{b}_{+,n}, \tilde{d}_{+,n})$ and similarly for k_n^-. In the notation of Supplemental Appendix H.1, $\overline{\text{bias}}_n$ is given by $\frac{1}{2}(\omega_{\tilde{\sigma}(\cdot),n}(\delta) - \delta\omega'_{\tilde{\sigma}(\cdot),n}(\delta))$ and $\tilde{v}_n$ is given by $\omega'_{\tilde{\sigma}(\cdot),n}(\delta)^2$ (see Equation (24) in the main text). If δ is chosen to minimize the length of the fixed-length CI, the half-length will be given by

$$\text{cv}_\alpha(\overline{\text{bias}}_n/\sqrt{\tilde{v}_n})\sqrt{\tilde{v}_n} = \inf_{\delta>0} \text{cv}_\alpha \left(\frac{\omega_{\tilde{\sigma}(\cdot),n}(\delta)}{2\omega'_{\tilde{\sigma}(\cdot),n}(\delta)} - \frac{\delta}{2} \right) \omega'_{\tilde{\sigma}(\cdot),n}(\delta),$$

and δ will achieve the minimum in the above display. Similarly, if the MSE criterion is used, δ will minimize $\overline{\text{bias}}_n^2 + v_n$.

We proceed by verifying the conditions of Theorem H.1 for the case where $\tilde{\sigma}(\cdot)$ is nonrandom and satisfies Assumption H.4, and then verifying the conditions in the last display of Theorem H.1 for the case where $\tilde{\sigma}(\cdot)$ satisfies Assumption H.3. The limiting kernel k^+ and k^- in Assumption H.5 will correspond to an asymptotic version of the modulus problem, which we now describe. Let

$$G_\infty(b_+, b_-, d_+, d_-) = p_{X,+}(0) \int_0^\infty \tilde{\sigma}_+^2(0) k^+_{\tilde{\sigma}_+(0)}(u; b_+, d_+)^2 \, du + p_{X,-}(0) \int_0^\infty \tilde{\sigma}_-^2(0) k^+_{\tilde{\sigma}_-(0)}(u; b_+, d_+)^2 \, du.$$

Consider the limiting inverse modulus problem

$$\omega^{-1}_{\tilde{\sigma}_+(0),\tilde{\sigma}_-(0),\infty}(2\tilde{b}_\infty) = \min_{f_+, f_- \in \mathcal{F}_{RDT,p}(C)} \sqrt{\frac{p_{X,+}(0)}{\tilde{\sigma}_+^2(0)} \int_0^\infty f_+(u)^2 \, du + \frac{p_{X,-}(0)}{\tilde{\sigma}_-^2(0)} \int_{-\infty}^0 f_-(u)^2 \, du}$$
$$\text{s.t. } f_+(0) + f_-(0) \geq \tilde{b}_\infty.$$

We use $\omega_\infty(\delta) = \omega_{\tilde{\sigma}_+(0),\tilde{\sigma}_-(0),\infty}(\delta)$ to denote the limiting modulus corresponding to this inverse modulus. The limiting inverse modulus problem is solved by the functions $f_+(u) = \tilde{\sigma}_+^2(0)k^+_{\tilde{\sigma}_+(0)}(u; b_+, d_+) = k_1^+(u; b_+, d_+)$ and $f_-(u) = \tilde{\sigma}_-^2(0)k^+_{\tilde{\sigma}_-(0)}(u; b_-, d_-) = k_1^-(u; b_+, d_+)$ for some (b_+, b_-, d_+, d_-) with $b_+ + b_- = \tilde{b}_\infty$ (this holds by the same arguments as for the modulus problem in Supplemental Appendix E.2). Thus, for any minimizer of G_∞, the functions $k_1^+(\cdot; b_+, d_+)$ and $k_1^+(\cdot; b_+, d_+)$ must solve the above inverse modulus problem. The solution to this problem is unique by strict convexity, which implies that G_∞ has a unique minimizer. Similarly, the minimizer of G_n is unique for each n. Let $(\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty})$ denote the minimizer of G_∞. The limiting kernel k^+ in Assumption H.5 will be given by

$k^+_{\tilde\sigma_+(0)}(\cdot;\tilde b_{+,\infty},\tilde d_{+,\infty})$ and similarly for k^-.

To derive the form of the limiting modulus of continuity, we argue as in Donoho and Low (1992). Let $k^+_1(\cdot;\tilde b_{+,\infty,1},\tilde d_{+,\infty,1})$ and $k^+_1(\cdot;\tilde b_{+,\infty,1},\tilde d_{+,\infty,1})$ solve the inverse modulus problem $\omega_\infty^{-1}(2\tilde b_\infty)$ for $\tilde b_\infty=1$. The feasible set for a given $\tilde b_\infty$ consists of all b_+,b_-,d_+,d_- such that $b_++b_-\geq\tilde b_\infty$, and a given b_+,b_-,d_+,d_- in this set achieves the value

$$\begin{aligned}&\sqrt{\frac{p_{X,+}(0)}{\tilde\sigma_+^2(0)}\int_0^\infty k_1^+(u;b_+,d_+)^2\,du+\frac{p_{X,-}(0)}{\tilde\sigma_-^2(0)}\int_{-\infty}^0 k_1^-(u;b_-,d_-)^2\,du}\\
&=\sqrt{\frac{p_{X,+}(0)}{\tilde\sigma_+^2(0)}\int_0^\infty k_1^+(vb_\infty^{1/p};b_+,d_+)^2\,d(vb_\infty^{1/p})+\frac{p_{X,-}(0)}{\tilde\sigma_-^2(0)}\int_{-\infty}^0 k_1^-(vb_\infty^{1/p};b_-,d_-)^2\,d(vb_\infty^{1/p})}\\
&=\sqrt{\frac{p_{X,+}(0)}{\tilde\sigma_+^2(0)}\tilde b_\infty^{1/p}\int_0^\infty \tilde b_\infty^2 k_1^+(v;b_+/\tilde b_\infty,\bar d_+)^2\,dv+\frac{p_{X,-}(0)}{\tilde\sigma_-^2(0)}\tilde b_\infty^{1/p}\int_{-\infty}^0 \tilde b_\infty^2 k_1^-(v;b_-/\tilde b_\infty,\bar d_-)^2\,dv},\end{aligned}$$

where $\bar d_+=(d_{+,1}/\tilde b_\infty^{(p-1)/p},\dots,d_{+,p-1}/\tilde b_\infty^{1/p})'$ and similarly for $\bar d_-$. This uses the fact that, for any $h>0$, $h^pk_1^+(u/h;b_+,d_+)=k_1^+(u;b_+h^p,d_{+,1}h^{p-1},d_{+,2}h^{p-2},\dots,d_{+,p-1}h)$ and similarly for k_1^-. This can be seen to be $\tilde b_\infty^{(2p+1)/(2p)}$ times the objective evaluated at $(b_+/\tilde b_\infty,b_-/\tilde b_\infty,\bar d_+,\bar d_-)$, which is feasible under $\tilde b_\infty=1$. Similarly, for any feasible function under $\tilde b_\infty=1$, there is a feasible function under a given $\tilde b_\infty$ that achieves $\tilde b_\infty^{(2p+1)/(2p)}$ times the value of under $\tilde b_\infty=1$. It follows that $\omega_\infty^{-1}(2b)=b^{(2p+1)/(2p)}\omega_\infty(2)$. Thus, ω_∞^{-1} is invertible and the inverse ω_∞ satisfies $\omega_\infty(\delta)=\omega_{\tilde\sigma_+(0),\tilde\sigma_-(0),\infty}(\delta)=\delta^{2p/(2p+1)}\omega_{\tilde\sigma_+(0),\tilde\sigma_-(0),\infty}(1)$.

If $\tilde b_\infty=\omega_\infty(\delta_\infty)$ for some δ_∞, then it can be checked that the limiting variance and worst-case bias defined in Supplemental Appendix H.1 correspond to the limiting modulus problem:

$$\overline{\text{bias}}_\infty=\frac12\left(\omega_\infty(\delta_\infty)-\delta_\infty\omega_\infty'(\delta_\infty)\right),\quad \sqrt{v_\infty}=\omega_\infty'(\delta_\infty). \tag{S22}$$

Furthermore, we will show that, if δ is chosen to optimize FLCI length for $\omega_{\tilde\sigma(\cdot),n}$, then this will hold with δ_∞ optimizing $\text{cv}_\alpha(\overline{\text{bias}}_\infty/\sqrt{v_\infty})\sqrt{v_\infty}$. Similarly, if δ is chosen to optimize MSE for $\omega_{\tilde\sigma(\cdot),n}$, then this will hold with δ_∞ optimizing $\overline{\text{bias}}_\infty^2+v_\infty$.

We are now ready to state the main result concerning the asymptotic validity and efficiency of feasible CIs based on the estimator given in this section.

Theorem H.3. *Suppose Assumptions H.1 and H.2 hold. Let $\hat L=\hat L_{\delta_n,\tilde\sigma(\cdot)}$ where δ_n is chosen to optimize one of the performance criteria for $\omega_{\tilde\sigma(\cdot),n}$ (FLCI length, RMSE, or a given quantile of excess length), and suppose that $\tilde\sigma(\cdot)$ satisfies Assumption H.3 or Assumption H.4* *with* $\tilde{\sigma}_+(0) = \sigma_+(0)$ *and* $\tilde{\sigma}_+(0) = \sigma_-(0)$. *Let* $\overline{bias}_n = \frac{1}{2}(\omega_{\tilde{\sigma}(\cdot),n}(\delta_n) - \delta_n \omega'_{\tilde{\sigma}(\cdot),n}(\delta_n))$ *and* $\tilde{v}_n = \omega'_{\tilde{\sigma}(\cdot),n}(\delta_n)^2$ *denote the worst-case bias and variance formulas. Let* $\hat{c}_{\alpha,\delta_n} = \hat{L} - \overline{bias}_n - z_{1-\alpha}\sqrt{\tilde{v}_n}$ *and* $\hat{\chi} = \mathrm{cv}_\alpha(\overline{bias}_n/\sqrt{\tilde{v}_n})\sqrt{\tilde{v}_n}$ *so that* $[\hat{c}_{\alpha,\delta_n}, \infty)$ *and* $[\hat{L} - \hat{\chi}, \hat{L} + \hat{\chi}]$ *give the corresponding CIs.*

The CIs $[\hat{c}_{\alpha,\delta_n}, \infty)$ *and* $[\hat{L} - \hat{\chi}, \hat{L} + \hat{\chi}]$ *have uniform asymptotic coverage at least* $1 - \alpha$. *In addition,* $n^{p/(2p+1)}\hat{\chi} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \chi_\infty$ *where* $\chi_\infty = \mathrm{cv}_\alpha(\overline{bias}_\infty/\sqrt{v_\infty})\sqrt{v_\infty}$ *with* $\overline{bias}_\infty$ *and* $\sqrt{v_\infty}$ *given in* (S22) *and* $\delta_\infty = z_\beta + z_{1-\alpha}$ *if excess length is the criterion,* $\delta_\infty = \arg\min_\delta \mathrm{cv}_\alpha(\frac{\omega_\infty(\delta)}{2\omega'_\infty(\delta)} - \frac{\delta}{2})\omega'_\infty(\delta)$ *if FLCI length is the criterion, and* $\delta_\infty = \arg\min_\delta[\frac{1}{4}\left(\omega_\infty(\delta_\infty) - \delta_\infty\omega'_\infty(\delta_\infty)\right)^2 + \omega'_\infty(\delta)^2]$ *if RMSE is the criterion.*

Suppose, in addition, that each $\mathcal{Q}_n$ *contains a distribution where the* u_i*s are normal. If the FLCI criterion is used, then no other sequence of linear estimators* $\tilde{L}$ *can satisfy*

$$\liminf_{n\to\infty} \inf_{f\in\mathcal{F},Q\in\mathcal{Q}_n} P_{f,Q}\left(Lf \in \left\{\tilde{L} \pm n^{-p/(2p+1)}\chi\right\}\right) \geq 1 - \alpha$$

with χ *a constant with* $\chi < \chi_\infty$. *In addition, for any sequence of confidence sets* $\mathcal{C}$ *with* $\liminf_{n\to\infty}\inf_{f\in\mathcal{F},Q\in\mathcal{Q}_n} P_{f,Q}(Lf \in \mathcal{C}) \geq 1 - \alpha$, *we have the following bound on the asymptotic efficiency improvement at any* $f \in \mathcal{F}_{RDT,p}(0)$:

$$\liminf_{n\to\infty} \sup_{Q\in\mathcal{Q}_n} \frac{n^{p/(2p+1)} E_{f,Q}\lambda(\mathcal{C})}{2\chi_\infty} \geq \frac{(1-\alpha)2^{2p/(2p+1)} E[(z_{1-\alpha} - Z)^{2p/(2p+1)} \mid Z \leq z_{1-\alpha}]}{\frac{4p}{2p+1}\inf_{\delta>0}\mathrm{cv}_\alpha\left(\delta/(4p)\right)\delta^{-1/(2p+1)}}$$

where $Z \sim \mathcal{N}(0,1)$.

If the excess length criterion is used with quantile β *(i.e.* $\delta_n = z_\beta + z_{1-\alpha}$*), then any one-sided CI* $[\hat{c}, \infty)$ *with*

$$\liminf_{n\to\infty} \inf_{f\in\mathcal{F},Q\in\mathcal{Q}_n} P_{f,Q}\left(Lf \in [\hat{c}, \infty)\right) \geq 1 - \alpha$$

must satisfy

$$\liminf_{n\to\infty} \frac{\sup_{f\in\mathcal{F},Q\in\mathcal{Q}_n} q_{f,Q,\beta}\left(Lf - \hat{c}\right)}{\sup_{f\in\mathcal{F},Q\in\mathcal{Q}_n} q_{f,Q,\beta}\left(Lf - \hat{c}_{\alpha,\delta_n}\right)} \geq 1$$

and, for any $f \in \mathcal{F}_{RDT,p}(0)$,

$$\liminf_{n\to\infty} \frac{\sup_{Q\in\mathcal{Q}_n} q_{f,Q,\beta}\left(Lf - \hat{c}\right)}{\sup_{Q\in\mathcal{Q}_n} q_{f,Q,\beta}\left(Lf - \hat{c}_{\alpha,\delta_n}\right)} \geq \frac{2^{2p/(2p+1)}}{1 + 2p/(2p+1)}.$$

To prove this theorem, we first prove a series of lemmas. To deal with the case where δ

is chosen to optimize FLCI length or MSE, we will use the characterization of the optimal δ for these criteria from Donoho (1994), which is described at the beginning of Supplemental Appendix G. In particular, for ρ_A and $\chi_{A,\alpha}$ given in Supplemental Appendix G, the δ that optimizes FLCI length is given by the δ that maximizes $\omega_{\tilde{\sigma}(\cdot),n}(\delta)\chi_{A,\alpha}(\delta)/\delta$, and the resulting FLCI half-length is given by $\sup_{\delta>0}\omega_{\tilde{\sigma}(\cdot),n}(\delta)\chi_{A,\alpha}(\delta)/\delta$. In addition, when δ is chosen to optimize FLCI length, χ_∞ in Theorem H.3 is given by $\sup_{\delta>0}\omega_\infty(\delta)\chi_{A,\alpha}(\delta)/\delta$, and δ_∞ maximizes this expression. If δ is chosen according to the MSE criterion, then δ maximizes $\omega_{\tilde{\sigma}(\cdot),n}(\delta)\sqrt{\rho_A(\delta)}/\delta$ and δ_∞ maximizes $\omega_\infty(\delta)\sqrt{\rho_A(\delta)}/\delta$.

Lemma H.2. *For any constant B, the following holds. Under Assumption H.4,*

$$\lim_{n\to\infty}\sup_{\|(b_+,b_-,d_+,d_-)\|\le B}|G_n(b_+,b_-,d_+,d_-)-G_\infty(b_+,b_-,d_+,d_-)|=0.$$

Under Assumption H.3,

$$\sup_{\|(b_+,b_-,d_+,d_-)\|\le B}|G_n(b_+,b_-,d_+,d_-)-G_\infty(b_+,b_-,d_+,d_-)|\underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}}0.$$

Proof. Define $\tilde{G}_n^+(b_+,d_+)=\frac{1}{nh_n}\sum_{i=1}^n k_1^+(x_i/h_n;b_+,d_+)^2$, and define $\tilde{G}_n^-$ analogously. Also, $\tilde{G}_\infty^+(b_+,d_+)=p_{X,+}(0)\int_0^\infty k_1^+(u;b_+,d_+)^2\,du$, with G_∞^- defined analogously. For each (b_+,d_+), $\tilde{G}_n(b_+,d_+)\to G_\infty(b_+,d_+)$ by Assumption H.1. To show uniform convergence, first note that, for some constant K_1, the support of $k_1^+(\cdot;b_+,d_+)$ is bounded by K_1 uniformly over $\|(b_+,d_+)\|\le B$ and similarly for $k_1^-(\cdot;b_-,d_-)$. Thus, for any (b_+,d_+) and $(\bar{b}_+,\bar{d}_+)$,

$$|G_n^+(b_+,d_+)-G_n^+(\bar{b}_+,\bar{d}_+)|\le\left[\frac{1}{nh_n}\sum_{i=1}^n 1(|x_i/h_n|\le K_1)\right]\sup_{|u|\le K_1}|k_1^+(u;b_+,d_+)-k_1^+(u;\bar{b}_+,\bar{d}_+)|.$$

Since the term in brackets converges to a finite constant by Assumption H.1 and k_1^+ is Lipschitz continuous on any bounded set, it follows that there exists a constant K_2 such that $|G_n^+(b_+,d_+)-G_n^+(\bar{b}_+,\bar{d}_+)|\le K_2\|(b_+,d_+)-(\bar{b}_+,\bar{d}_+)\|$ for all n. Using this and applying pointwise convergence of $G_n^+(b_+,d_+)$ on a small enough grid along with uniform continuity of $G_\infty(b_+,d_+)$ on compact sets, it follows that

$$\lim_{n\to\infty}\sup_{\|(b_+,b_-,d_+,d_-)\|\le B}|\tilde{G}_n(b_+,d_+)-\tilde{G}_\infty(b_+,d_+)|=0,$$

and similar arguments give the same statement for $\tilde{G}_n^-$ and $\tilde{G}_\infty^-$. Under Assumption H.4,

$$\left|G_n(b_+, b_-, d_+, d_-) - \left[\tilde{G}_n(b_+, d_+)\tilde{\sigma}_+^2(0) + \tilde{G}_n(b_-, d_-)\tilde{\sigma}_-^2(0)\right]\right| \leq$$
$$\overline{k} \cdot \left[\frac{1}{nh_n}\sum_{i=1}^{n} 1(|x_i/h_n| \leq K_1)\right]\left[\sup_{0<x\leq K_1 h_n}\left|\tilde{\sigma}_+^2(0) - \tilde{\sigma}_+^2(x)\right| + \sup_{-K_1 h_n \leq x<0}\left|\tilde{\sigma}_-^2(0) - \tilde{\sigma}_-^2(x)\right|\right]$$

where $\overline{k}$ is an upper bound for $|k_1^+(x)|$ and $|k_1^-(x)|$. This converges to zero by left- and right-continuity of $\tilde{\sigma}$ at 0. The result then follows since $G_\infty(b_+, b_-, d_+, d_-) = \tilde{\sigma}_+^2(0)\tilde{G}_\infty^+(b_+, d_+) + \tilde{\sigma}_-^2(0)\tilde{G}_\infty^-(b_-, d_-)$. Under Assumption H.3, we have $G_n(b_+, b_-, d_+, d_-) = \tilde{G}_n^+(b_+, d_+)\hat{\sigma}_+^2 + \tilde{G}_n^+(b_-, d_-)\hat{\sigma}_-^2$, and the result follows from uniform convergence in probability of $\hat{\sigma}_+^2$ and $\hat{\sigma}_-^2$ to $\tilde{\sigma}_+^2(0)$ and $\tilde{\sigma}_-^2(0)$. □

Lemma H.3. *Under Assumption H.4, $\|(\tilde{b}_{+,n}, \tilde{b}_{-,n}, \tilde{d}_{+,n}, \tilde{d}_{-,n})\| \leq B$ for some constant B and n large enough. Under Assumption H.3, the same statement holds with probability approaching one uniformly over $\mathcal{F}, \mathcal{Q}_n$.*

Proof. Let $\mathcal{A}(x, b, d) = b + \sum_{i=1}^{p-1} d(x/h_n)^j$, where $d = (d_1, \ldots, d_{p-1})$. Note $G_n(b_+, b_-, d_+, d_-)$ is bounded from below by $1/\sup_{|x|\leq h_n}\tilde{\sigma}^2(x)$ times

$$\frac{1}{nh_n}\sum_{i:0<x_i\leq h_n}(|\mathcal{A}(x_i, b_+, d_+)| - C)_+^2 + \frac{1}{nh_n}\sum_{i:-h_n\leq x_i<0}(|\mathcal{A}(x_i, b_-, d_-)| - C)_+^2$$
$$\geq \frac{1}{4nh_n}\sum_{i:0<x_i\leq h_n}\left[\mathcal{A}(x_i, b_+, d_+)^2 - 4C^2\right] + \frac{1}{4nh_n}\sum_{i:-h_n\leq x_i<0}\left[\mathcal{A}(x_i, b_-, d_-)^2 - 4C^2\right]$$

(the inequality follows since, for any $s \geq 2C$, $(s - C)^2 \geq s^2/4 \geq s^2/4 - C^2$ and, for $2C \geq s \geq C$, $(s - C)^2 \geq 0 \geq s^2/4 - C^2$). Note that, for any $B > 0$

$$\inf_{\max\{|b_+|,|d_{+,1}|,\ldots,|d_{+,p-1}|\}\geq B}\frac{1}{4nh_n}\sum_{i:0<x_i\leq h_n}\mathcal{A}(x_i, b_+, d_+)^2$$
$$= B^2 \inf_{\max\{|b_+|,|d_{+,1}|,\ldots,|d_{+,p-1}|\}\geq 1}\frac{1}{4nh_n}\sum_{i:0<x_i\leq h_n}\mathcal{A}(x_i, b_+, d_+)^2$$
$$\to \frac{p_{X,+}(0)}{4}B^2 \inf_{\max\{|b_+|,|d_{+,1}|,\ldots,|d_{+,p-1}|\}\geq 1}\int_0^\infty\left(b_+ + \sum_{i=1}^{p-1} d_{+,j}u^j\right)^2 du$$

and similarly for the term involving $\mathcal{A}(x_i, b_-, d_-)$ (the convergence follows since the infimum is taken on the compact set where $\max\{|b_+|, |d_{+,1}|, \ldots, |d_{+,p-1}|\} = 1$). Combining this with the previous display and the fact that $\frac{1}{nh}\sum_{i:|x_i|\leq h_n} C^2$ converges to a finite constant, it follows that, for some $\eta > 0$, $\inf_{\max\{|b_+|,|d_{+,1}|,\ldots,|d_{+,p-1}|\}\geq B} G_n(b_+, b_-, d_+, d_-) \geq (B^2\eta - \eta^{-1})/\sup_{|x|\leq h_n}\tilde{\sigma}^2(x)$ for large enough n. Let K be such that $G_\infty(\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty}) \leq K/2$ and $\max\{\tilde{\sigma}_+^2(0), \tilde{\sigma}_-^2(0)\} \leq K/2$. Under Assumption H.4, $G_n(\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty}) < K$ and $\sup_{|x|\leq h_n}\tilde{\sigma}^2(x) \leq K$ for large enough n. Under Assumption H.3, $G_n(\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty}) < K$ and $\sup_{|x|\leq h_n}\tilde{\sigma}^2(x) \leq K$ with probability approaching one uniformly over $\mathcal{F}, \mathcal{Q}_n$. Let B be large enough so that $(B^2\eta - \eta^{-1})/K > K$. Then, when $G_n(\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty}) \leq K$ and $\sup_{|x|\leq h_n}\tilde{\sigma}^2(x) \leq K$, $(\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty})$ will give a lower value of G_n than any (b_+, b_-, d_+, d_-) with $\max\{|b_+|, |d_{+,1}|, \ldots, |d_{+,p-1}|, |b_-|, |d_{-,1}|, \ldots, |d_{-,p-1}|\} \geq B$. The result follows from the fact that the max norm on $\mathbb{R}^{2p}$ is bounded from below by a constant times the Euclidean norm. □

Lemma H.4. *If Assumption H.4 holds and* $\tilde{b}_n \to \tilde{b}_\infty$, *then* $(\tilde{b}_{+,n}, \tilde{b}_{-,n}, \tilde{d}_{+,n}, \tilde{d}_{-,n}) \to (\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty})$. *If Assumption H.3 holds and* $\tilde{b}_n \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} \tilde{b}_\infty > 0$, $(\tilde{b}_{+,n}, \tilde{b}_{-,n}, \tilde{d}_{+,n}, \tilde{d}_{-,n}) \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} (\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty})$.

Proof. By Lemma H.3, B can be chosen so that $\|(\tilde{b}_{+,n}, \tilde{b}_{-,n}, \tilde{d}_{+,n}, \tilde{d}_{-,n})\| \leq B$ for large enough n under Assumption H.4 and $\|(\tilde{b}_{+,n}, \tilde{b}_{-,n}, \tilde{d}_{+,n}, \tilde{d}_{-,n}) \leq B\|$ with probability one uniformly over $\mathcal{F}, \mathcal{Q}_n$ under Assumption H.3. The result follows from Lemma H.2, continuity of G_∞ and the fact that G_∞ has a unique minimizer. □

Lemma H.5. *If Assumption H.4 holds and* $\tilde{b}_n \to \tilde{b}_\infty > 0$, *then* $\omega_n^{-1}(n^{p/(2p+1)}\tilde{b}_n) \to \omega_\infty^{-1}(\tilde{b}_\infty)$. *If Assumption H.3 holds and* $\tilde{b}_n \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} b_\infty > 0$, *then* $\omega_n^{-1}(n^{p/(2p+1)}\tilde{b}_n) \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} \omega_\infty^{-1}(\tilde{b}_\infty)$.

Proof. The result is immediate from Lemmas H.2 and H.4. □

Lemma H.6. *Under Assumption H.4, we have, for any* $\overline{\delta} > 0$,

$$\sup_{0<\delta\leq\overline{\delta}} \left| n^{p/(2p+1)}\omega_n(\delta) - \omega_\infty(\delta) \right| \to 0.$$

Under Assumption H.3, we have, for any $\overline{\delta} > 0$,

$$\sup_{0<\delta\leq\overline{\delta}} \left| n^{p/(2p+1)}\omega_n(\delta) - \omega_\infty(\delta) \right| \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0.$$

Proof. The first statement is immediate from Lemma H.5 and Lemma G.1 (with $n^{p/(2p+1)}\omega_n$ playing the role of ω_n in that lemma). For the second claim, note that, if $|\hat{\sigma}_+ - \sigma_+(0)| \leq \eta$ and $|\hat{\sigma}_- - \sigma_-(0)| \leq \eta$, $\omega_{n,\underline{\sigma}(\cdot)}(\delta) \leq \omega_{\tilde{\sigma}(\cdot),n}(\delta) \leq \omega_{n,\overline{\sigma}(\cdot)}(\delta)$, where $\underline{\sigma}(x) = (\sigma_+(0) - \eta)1(x >$

$0) + (\sigma_-(0) - \eta)1(x < 0)$ and $\overline{\sigma}(x)$ is defined similarly. Applying the first statement in the lemma and the fact that $|\hat{\sigma}_+ - \sigma_+(0)| \leq \eta$ and $|\hat{\sigma}_- - \sigma_-(0)| \leq \eta$ with probability approaching one uniformly over $\mathcal{F}, \mathcal{Q}_n$, it follows that, for any $\varepsilon > 0$, we will have

$$\omega_{\underline{\sigma}_+(0),\underline{\sigma}_-(0),\infty}(\delta) - \varepsilon \leq n^{p/(2p+1)}\omega_n(\delta) \leq \omega_{\overline{\sigma}_+(0),\overline{\sigma}_-(0),\infty}(\delta) + \varepsilon$$

for all $0 < \delta < \overline{\delta}$ with probability approaching one uniformly over $\mathcal{F}, \mathcal{Q}_n$. By making η and ε small, both sides can be made arbitrarily close to $\omega_\infty(\delta) = \omega_{\infty,\sigma_+(0),\sigma_-(0)}(\delta)$. □

Lemma H.7. *Let r denote $\sqrt{\rho_A}$ or $\chi_{A,\alpha}$. Under Assumption H.4,*

$$\sup_{\delta>0} n^{p/(2p+1)}\omega_n(\delta)r(\delta/2)/\delta \to \sup_{\delta>0} \omega_\infty(\delta)r(\delta/2)/\delta.$$

Let δ_n minimize the left-hand side of the above display, and let δ^* minimize the right-hand side. Then $\delta_n \to \delta^*$ under Assumption H.4 and $\delta_n \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} \delta^*$ under Assumption H.3. In addition, for any $0 < \alpha < 1$ and Z a standard normal variable,

$$\lim_{n\to\infty} (1-\alpha)E[n^{p/(2p+1)}\omega_n(2(z_{1-\alpha} - Z))|Z \leq z_{1-\alpha}] = (1-\alpha)E[\omega_\infty(2(z_{1-\alpha} - Z))|Z \leq z_{1-\alpha}].$$

Proof. All the statements are immediate from Lemmas H.6 and G.2 except for the statement that $\delta_n \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} \delta^*$ under Assumption H.3. The statement that $\delta_n \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} \delta^*$ under Assumption H.3 follows by using Lemma H.6 and analogous arguments to those in Lemma G.2 to show that there exist $0 < \underline{\delta} < \overline{\delta}$ such that $\delta_n \in [\underline{\delta}, \overline{\delta}]$ with probability approaching on uniformly in $\mathcal{F}, \mathcal{Q}_n$, and that $\sup_{\delta\in[\underline{\delta},\overline{\delta}]} \left|n^{p/(2p+1)}\omega_n(\delta)r(\delta/2)/\delta - \omega(\delta)r(\delta/2)/\delta\right| \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0$. □

Lemma H.8. *Under Assumptions H.1 and H.2, the following hold. If Assumption H.4 holds and $\tilde{b}_n$ is a deterministic sequence with $\tilde{b}_n \to \tilde{b}_\infty > 0$, then*

$$\sup_x |k^+_{\tilde{\sigma}(\cdot)}(x; \tilde{b}_{+,n}, \tilde{d}_{+,n}) - k^+_{\tilde{\sigma}_+(0)}(x; \tilde{b}_{+,\infty}, \tilde{d}_{+,\infty})| \to 0,$$
$$\sup_x |k^-_{\tilde{\sigma}(\cdot)}(x; \tilde{b}_{-,n}, \tilde{d}_{-,n}) - k^-_{\tilde{\sigma}_-(0)}(x; \tilde{b}_{-,\infty}, \tilde{d}_{-,\infty})| \to 0.$$

If Assumption H.3 holds and $\tilde{b}_n$ is a random sequence with $\tilde{b}_n \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} \tilde{b}_\infty > 0$, then

$$\sup_x |k^+_{\tilde{\sigma}(\cdot)}(x; \tilde{b}_{+,n}, \tilde{d}_{+,n}) - k^+_{\tilde{\sigma}_+(0)}(x; \tilde{b}_{+,\infty}, \tilde{d}_{+,\infty})| \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0,$$
$$\sup_x |k^-_{\tilde{\sigma}(\cdot)}(x; \tilde{b}_{-,n}, \tilde{d}_{-,n}) - k^-_{\tilde{\sigma}_-(0)}(x; \tilde{b}_{-,\infty}, \tilde{d}_{-,\infty})| \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0$$

Proof. Note that

$$|k^+_{\tilde\sigma(\cdot)}(x;\tilde b_{+,n},\tilde d_{+,n}) - k^+_{\tilde\sigma_+(0)}(x;\tilde b_{+,\infty},\tilde d_{+,\infty})| \le |k^+_{\tilde\sigma(\cdot)}(x;\tilde b_{+,n},\tilde d_{+,n}) - k^+_{\tilde\sigma_+(0)}(x;\tilde b_{+,n},\tilde d_{+,n})| + |k^+_{\tilde\sigma_+(0)}(x;\tilde b_{+,n},\tilde d_{+,n}) - k^+_{\tilde\sigma_+(0)}(x;\tilde b_{+,\infty},\tilde d_{+,\infty})|.$$

Under Assumption H.4, the first term is, for large enough n, bounded by a constant times $\sup_{0<x<h_nK}|\tilde\sigma^{-2}(x)-\tilde\sigma_+^{-2}(0)|$, where K is bound on the support of $k_1^+(\cdot;b_+,d_+)$ over b_+,d_+ in a neighborhood of $\tilde b_{+,\infty},\tilde d_{+,\infty}$. This converges to zero by Assumption H.4. The second term converges to zero by Lipschitz continuity of $k^+_{\tilde\sigma_+(0)}$. Under Assumption H.3, the first term is bounded by a constant times $|\hat\sigma_+^{-2}-\tilde\sigma_+(0)|$, which converges in probability to zero uniformly over $\mathcal{F},\mathcal{Q}_n$ by assumption. The second term converges in probability to zero uniformly over $\mathcal{F},\mathcal{Q}_n$ by Lipschitz continuity of $k^+_{\tilde\sigma_+(0)}$. Similar arguments apply to $k^-_{\tilde\sigma(\cdot)}$ in both cases. □

Lemma H.9. *Under Assumptions H.1 and H.2, the following holds. Let $\hat L$ denote the estimator $\hat L_{\delta_n,\tilde\sigma(\cdot)}$ where $\tilde\sigma(\cdot)$ satisfies Assumption H.4 and $\delta_n=\omega^{-1}_{\tilde\sigma(\cdot),n}(2n^{-p/(2p+1)}\tilde b_n)$ where $\tilde b_n$ is a deterministic sequence with $\tilde b_n\to\tilde b_\infty$. Let $\overline{bias}_n$ and $\tilde v_n$ denote the corresponding worst-case bias and variance formulas. Let $\hat L^*$ denote the estimator $\hat L_{\delta_n^*,\tilde\sigma(\cdot)}$ where $\tilde\sigma^*(\cdot)=\hat\sigma_+1(x>0)+\hat\sigma_-1(x<0)$ satisfies Assumption H.3 with the same value of $\tilde\sigma_+(0)$ and $\tilde\sigma_-(0)$ and $\delta_n^*=\omega^{-1}_{\tilde\sigma(\cdot),n}(2n^{-p/(2p+1)}\tilde b_n^*)$ where $\tilde b_n^*\underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}}\tilde b_\infty$. Let $\overline{bias}_n^*$ and $\tilde v_n^*$ denote the corresponding worst-case bias and variance formulas. Then*

$$n^{p/(2p+1)}\left(\hat L-\hat L^*\right)\underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}}0,\qquad n^{p/(2p+1)}\left(\overline{bias}_n-\overline{bias}_n^*\right)\underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}}0,\qquad \frac{\tilde v_n}{\tilde v_n^*}\underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}}1.$$

Proof. We have

$$\hat L=\frac{1}{nh_n}\sum_{i=1}^n w_n(x_i/h_n)y_i=\frac{1}{nh_n}\sum_{i=1}^n w_n(x_i/h_n)f(x_i)+\frac{1}{nh_n}\sum_{i=1}^n w_n(x_i/h_n)u_i$$

where $w_n(u)=\frac{k^+_{\tilde\sigma(\cdot)}(u;\tilde b_{+,n},\tilde d_{+,n})}{\frac{1}{nh_n}\sum_{j=1}^n k^+_{\tilde\sigma(\cdot)}(x_j/h_n;\tilde b_{+,n},\tilde d_{+,n})}$ for $u>0$ and similarly with $k^+_{\tilde\sigma(\cdot)}$ replaced by $k^-_{\tilde\sigma(\cdot)}$ for $u<0$ (here, $\tilde d_{+,n}$, $\tilde d_{-,n}$, $\tilde b_{+,n}$ and $\tilde b_{-,n}$ are the coefficients in the solution to the inverse modulus problem defined above). Similarly, $\hat L^*$ takes the same form with w_n replaced by $w_n^*(u)=\frac{k^+_{\tilde\sigma^*(\cdot)}(u;\tilde b_n^*,\tilde d_n^*)}{\frac{1}{nh_n}\sum_{j=1}^n k^+_{\tilde\sigma^*(\cdot)}(x_j/h_n;\tilde b_n^*,\tilde d_n^*)}$ for $u>0$ and similarly for $u<0$ (with $\tilde d^*_{+,n}$, $\tilde d^*_{-,n}$, $\tilde b^*_{+,n}$ and $\tilde b^*_{-,n}$ the coefficients in the solution to the corresponding inverse modulus problem). Let $w_\infty(u)=\frac{k^+_{\tilde\sigma(\cdot)}(u;\tilde b_n^*,\tilde d_n^*)}{p_{X,+}(0)\int k^+_{\tilde\sigma(\cdot)}(u;\tilde b_\infty,\tilde d_\infty)\,du}$ Note that, by Lemma H.8, $\sup_u|w_n(u)-w_\infty(u)|\to 0$ and

$\sup_u |w_n^*(u) - w_\infty(u)| \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0$.

We have

$$\hat{L} - \hat{L}^* = \frac{1}{nh_n}\sum_{i=1}^n [w_n(x_i/h_n) - w_n^*(x_i/h_n)]r(x_i) + \frac{1}{nh_n}\sum_{i=1}^n [w_n(x_i/h_n) - w_n^*(x_i/h_n)]u_i$$

where $f(x) = \sum_{j=0}^{p-1} f_+^{(j)}(0)x^j 1(x>0)/j! + \sum_{j=0}^{p-1} f_-^{(j)}(0)x^j 1(x<0)/j! + r(x)$ and we use the fact that $\sum_{i=1}^n w_n(x_i/h_n)x_i^j = \sum_{i=1}^n w_n^*(x_i/h_n)x_i^j$ for $j = 0, \ldots, p-1$. Let B be such that, with probability approaching one, $w_n(x) = w_n^*(x) = 0$ for all x with $|x| \geq B$. The first term is bounded by

$$\frac{C}{nh_n}\sum_{i=1}^n |w_n(x_i/h_n) - w_n^*(x_i/h_n)| \cdot |x_i|^p \leq \sup_x |w_n(x) - w_n^*(x)|\, Bh_n^p \frac{C}{nh_n}\sum_{i=1}^n 1(|x_i/h_n| \leq B).$$

It follows from Lemma H.8 that $\sup_x |w_n(x) - w_n^*(x)| \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0$. Also, $\frac{1}{nh_n}\sum_{i=1}^n 1(|x_i/h_n| \leq B)$ converges to a finite constant by Assumption H.1. Thus, the above display converges uniformly in probability to zero when scaled by $n^{p/(2p+1)} = h_n^{-p}$.

For the last term in $\hat{L} - \hat{L}^*$, scaling by $n^{p/(2p+1)}$ gives

$$\frac{1}{\sqrt{nh_n}}\sum_{i=1}^n [w_n(x_i/h_n) - w_\infty(x_i/h_n)]u_i - \frac{1}{\sqrt{nh_n}}\sum_{i=1}^n [w_n^*(x_i/h_n) - w_\infty(x_i/h_n)]u_i.$$

The first term has mean zero and variance $\frac{1}{nh}\sum_{i=1}^n [w_n(x_i/h_n) - w_\infty(x_i/h_n)]^2\sigma^2(x_i)$ which is bounded by $\{\sup_u [w_n(u) - w_\infty(u)]^2\}\left[\sup_{|x|\leq Bh_n}\sigma^2(x)\right]\frac{1}{nh}\sum_{i=1}^n 1(|x_i/h_n| \leq B) \to 0$. Let $c_{n,+} = \frac{\hat{\sigma}_+^2}{nh_n}\sum_{i=1}^n k_{\tilde{\sigma}^*(\cdot)}(x_i/h_n; \tilde{b}_{+,n}^*, \tilde{d}_{+,n}^*)$ and $c_{\infty,+} = \tilde{\sigma}_+^2(0)p_{X,+}(0)\int k_{\tilde{\sigma}^*(\cdot)}(u; \tilde{b}_\infty, \tilde{d}_\infty)$ so that $c_{n,+} \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} c_{\infty,+}$, and define $c_{n,-}$ and $c_{\infty,-}$ analogously. With this notation, we have, for $x_i > 0$,

$$w_n^*(x_i/h_n) = c_{n,+}^{-1}\hat{\sigma}_+^2 k_{\tilde{\sigma}^*(\cdot)}(x_i/h_n; \tilde{b}_{+,n}^*, \tilde{d}_{+,n}^*) = c_{n,+}^{-1}h_+(x_i/h_n; \tilde{b}_{+,n}^*, \tilde{d}_{+,n}^*)$$

and $w_\infty(u) = c_{\infty,+}^{-1}h_+(x_i/h_n; \tilde{b}_{+,\infty}, \tilde{d}_{+,\infty})$ where

$$h_+(u; b_+, d_+) = \left(b_+ + \sum_{j=1}^{p-1} d_{+,j}u^j - C|u|^p\right)_+ - \left(b_+ + \sum_{j=1}^{p-1} d_{+,j}u^j + C|u|^p\right)_-.$$

Thus,

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}[w_n^*(x_i/h_n)-w_\infty(x_i/h_n)]1(x_i>0)u_i$$
$$=\frac{c_{n,+}^{-1}}{\sqrt{nh}}\sum_{i=1}^{n}[h_+(u;\tilde{b}_{+,n},\tilde{d}_{+,n})-h_+(u;\tilde{b}_{+,\infty},\tilde{d}_{+,\infty})]1(x_i>0)u_i$$
$$+\frac{(c_{n,+}^{-1}-c_{n,\infty}^{-1})}{\sqrt{nh}}\sum_{i=1}^{n}h_+(u;\tilde{b}_{+,\infty},\tilde{d}_{+,\infty})1(x_i>0)u_i.$$

The last term converges to zero uniformly in probability by Slutsky's Theorem. The first term can be written as $c_{n,+}^{-1}$ times the sum of

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\left[\left(\tilde{b}_{+,n}^*+\sum_{j=1}^{p-1}\tilde{d}_{+,n,j}^*\left(\frac{x_i}{h_n}\right)^j-C\left|\frac{x_i}{h_n}\right|^p\right)_+\right.$$
$$\left.-\left(\tilde{b}_{+,\infty}+\sum_{j=1}^{p-1}\tilde{d}_{+,\infty,j}\left(\frac{x_i}{h_n}\right)^j-C\left|\frac{x_i}{h_n}\right|^p\right)_+\right]u_i$$

and a corresponding term with $(\cdot)_+$ replaced by $(\cdot)_-$, which can be dealt with using similar arguments. Letting $A(b_+,d_+)=\{u\colon b_++\sum_{j=1}^{p-1}d_{+,j}u^j-C|u|^p\geq 0\}$, the above display is equal to

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\left(\tilde{b}_{+,n}^*-\tilde{b}_{+,\infty}+\sum_{j=1}^{p-1}(\tilde{d}_{+,n,j}^*-\tilde{d}_{+,\infty,j})\left(\frac{x_i}{h_n}\right)^j\right)1(x_i/h_n\in A(\tilde{b}_{+,\infty},\tilde{d}_{+,\infty}))u_i$$
$$+\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\left(\tilde{b}_{+,n}^*+\sum_{j=1}^{p-1}d_{+,n,j}^*\left(\frac{x_i}{h_n}\right)^j-C\left|\frac{x_i}{h_n}\right|^p\right)$$
$$\cdot\left[1(x_i/h_n\in A(\tilde{b}_{+,n}^*,\tilde{d}_{+,n}^*))-1(x_i/h_n\in A(\tilde{b}_{+,\infty},\tilde{d}_{+,\infty}))\right]u_i.$$

The first term converges to zero uniformly in probability by Slutsky's Theorem. The second term can be written as a sum of terms of the form

$$\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}(x_i/h_n)^j\left[1(x_i/h_n\in A(\tilde{b}_{+,n}^*,\tilde{d}_{+,n}^*))-1(x_i/h_n\in A(\tilde{b}_{+,\infty},\tilde{d}_{+,\infty}))\right]u_i$$

times sequences that converge uniformly in probability to finite constants. To show that this converges in probability to zero uniformly over $\mathcal{F}, \mathcal{Q}_n$, note that, letting $u_1^*, \ldots, u_k^*$ be the positive zeros of the polynomial $\tilde{b}_{+,\infty} + \sum_{j=1}^{p-1} \tilde{d}_{+,j,\infty} u^j + Cu^p$, the following statement will hold with probability approaching one uniformly over $\mathcal{F}, \mathcal{Q}_n$ for any $\eta > 0$: for all u with $1(u \in A(\tilde{b}_{+,n}^*, \tilde{d}_{+,n}^*)) - 1(u \in A(\tilde{b}_{+,\infty}, \tilde{d}_{+,\infty})) \neq 0$, there exists ℓ such that $|u - u_\ell^*| \leq \eta$. It follows that the above display is, with probability approaching one uniformly over $\mathcal{F}, \mathcal{Q}_n$, bounded by a constant times the sum over $j = 0, \ldots, p$ and $\ell = 1, \ldots, k$ of

$$\max_{-1 \leq t \leq 1} \left| \frac{1}{\sqrt{nh_n}} \sum_{i:\, u_\ell - \eta \leq x_i/h_n \leq u_\ell + t\eta} (x_i/h_n)^j u_i \right|.$$

By Kolmogorov's inequality (see pp. 62-63 in Durrett, 1996), the probability of this quantity being greater than a given $\delta > 0$ under a given f, Q is bounded by

$$\frac{1}{\delta^2} \frac{1}{nh_n} \sum_{i:\, u_\ell - \eta \leq x_i/h_n \leq u_\ell + \eta} var_Q \left[(x_i/h_n)^j u_i \right] = \frac{1}{\delta^2} \frac{1}{nh_n} \sum_{i:\, u_\ell - \eta \leq x_i/h_n \leq u_\ell + \eta} (x_i/h_n)^{2j} \sigma^2(x_i)$$
$$\to \frac{p_{X,+}(0) \sigma_+^2(0)}{\delta^2} \int_{u_\ell^* - \eta}^{u_\ell^* + \eta} u^{2j}\, du,$$

which can be made arbitrarily small by making η small.

For the bias formulas, we have

$$\left| \overline{\text{bias}}_n - \overline{\text{bias}}_n^* \right| = \frac{C}{nh_n} \left| \sum_{i=1}^n |w_n(x_i/h_n) x_i^p| - \sum_{i=1}^n |w_n^*(x_i/h_n) x_i^p| \right|$$
$$\leq \frac{C}{nh_n} \sum_{i=1}^n |w_n(x_i/h_n) - w_n^*(x_i/h_n)| \cdot |x_i|^p.$$

This converges to zero when scaled by $n^{p/(2p+1)}$ by arguments given above.

For the variance formulas, we have

$$|\tilde{v}_n - \tilde{v}_n^*| = \frac{1}{(nh_n)^2} \left| \sum_{i=1}^n w_n(x_i/h_n)^2 \tilde{\sigma}^2(x_i) - \sum_{i=1}^n w_n^*(x_i/h_n)^2 \tilde{\sigma}^{*2}(x_i) \right|$$
$$\leq \frac{1}{(nh_n)^2} \sum_{i=1}^n \left| w_n(x_i/h_n)^2 \tilde{\sigma}^2(x_i) - w_n^*(x_i/h_n)^2 \tilde{\sigma}^{*2}(x_i) \right|$$
$$\leq \frac{1}{nh_n} \max_{|x| \leq B} \left| w_n(x)^2 \tilde{\sigma}^2(x) - w_n^*(x)^2 \tilde{\sigma}^{*2}(x) \right| \cdot \frac{1}{nh_n} \sum_{i=1}^n 1(|x_i/h_n| \leq B)$$

with probability approaching one where B is a bound on the support of $w_n(x)$ and $w_n^*(x)$ that holds with probability approaching one. Since $\frac{1}{nh_n}\sum_{i=1}^n 1(|x_i/h_n| \leq B)$ converges to a constant by Assumption H.1 and $\tilde{v}_n = n^{-2p/(2p+1)}v_\infty(1+o(1)) = (nh_n)^{-1}v_\infty(1+o(1))$, dividing the above display by $\tilde{v}_n$ gives an expression that is bounded by a constant times $\max_{|x|\leq Bh_n} |w_n(x)^2\tilde{\sigma}^2(x) - w_n^*(x)^2\tilde{\sigma}^{*2}(x)|$, which converges uniformly in probability to zero. □

We are now ready to prove Theorem H.3. First, consider the case with $\tilde{\sigma}(\cdot)$ is deterministic and Assumption H.4 holding. By Lemma H.7, $\delta_n \to \delta_\infty$. By Lemma H.6, it then follows that, under Assumption H.4, $n^{p/(2p+1)}w_n(\delta_n) \to \omega_\infty(\delta_\infty)$ so that Lemma H.8 applies to show that Assumption H.5 holds with $k^+(x) = k^+_{\tilde{\sigma}_+(0)}(x;\tilde{b}_{+,\infty},\tilde{d}_{+,\infty})$ and $k^-(x) = k^-_{\tilde{\sigma}_-(0)}(x;\tilde{b}_{-,\infty},\tilde{d}_{-,\infty})$, where $(\tilde{b}_{+,\infty},\tilde{d}_{+,\infty},\tilde{b}_{-,\infty},\tilde{d}_{-,\infty})$ minimize $G_\infty(\tilde{b}_{+,\infty},\tilde{d}_{+,\infty},\tilde{b}_{-,\infty},\tilde{d}_{-,\infty})$ subject to $\tilde{b}_{+,\infty}+\tilde{b}_{-,\infty} = \omega_\infty(\delta_\infty)/2$. The coverage statements and convergence of $n^{p/(2p+1)}\hat{\chi}$ then follow from Theorem H.1 and by calculating $\overline{\text{bias}}_\infty$ and v_∞ in terms of the limiting modulus.

We now prove the optimality statements (under which the assumption was made that, for each n, there exists a $Q \in \mathcal{Q}_n$ such that the errors are normally distributed). In this case, for any $\eta > 0$, if a linear estimator $\tilde{L}$ and constant χ satisfy

$$\inf_{f\in\mathcal{F},Q\in\mathcal{Q}_n} P\Big(Lf \in \{\tilde{L} \pm n^{-p/(2p+1)}\chi\}\Big) \geq 1-\alpha-\eta,$$

we must have $\chi \geq \sup_{\delta>0}\frac{n^{p/(2p+1)}\omega_{\sigma(\cdot),n}(\delta)}{\delta}\chi_{A,\alpha+\eta}(\delta/2)$ by the results of Donoho (1994) (using the characterization of optimal half-length at the beginning of Supplemental Appendix G). This converges to $\sup_{\delta>0}\frac{\omega_\infty(\delta)}{\delta}\chi_{A,\alpha+\eta}(\delta/2)$ by Lemma H.7. If $\liminf_n \inf_{f\in\mathcal{F},Q\in\mathcal{Q}_n} P(Lf \in \{\tilde{L} \pm n^{-p/(2p+1)}\chi\}) \geq 1-\alpha$, then, for any $\eta > 0$, the above display must hold for large enough n, so that $\chi \geq \lim_{\eta\downarrow 0}\sup_{\delta>0}\frac{\omega_\infty(\delta)}{\delta}\chi_{A,\alpha+\eta}(\delta/2) = \sup_{\delta>0}\frac{\omega_\infty(\delta)}{\delta}\chi_{A,\alpha}(\delta/2)$ (the limit with respect to η follows since there exist $0 < \underline{\delta} < \overline{\delta} < \infty$ such that the supremum over δ is taken on $[\underline{\delta},\overline{\delta}]$ for η in a neighborhood of zero, and since $\chi_{A,\alpha}(\delta/2)$ is continuous with respect to α uniformly over δ in compact sets).

For the asymptotic efficiency bound regarding expected length among all confidence intervals, note that, for any $\eta > 0$, any CI satisfying the asymptotic coverage requirement must be a $1-\alpha-\eta$ CI for large enough n, which means that, since the CI is valid under the $Q_n \in \mathcal{Q}_n$ where the errors are normal, the expected length of the CI at $f = 0$ and this Q_n scaled by $n^{p/(2p+1)}$ is at least

$$(1-\alpha-\eta)E\left[n^{p/(2p+1)}\omega_{\sigma(\cdot),n}(2(z_{1-\alpha-\eta}-Z))|Z \leq z_{1-\alpha-\eta}\right]$$

by Corollary 3.3. This converges to $(1-\alpha-\eta)E\left[\omega_\infty(2(z_{1-\alpha-\eta}-Z)) \mid Z \leq z_{1-\alpha-\eta}\right]$ by Lemma H.7. The result follows from taking $\eta \to 0$ and using the dominated convergence theorem, and using the fact that $\omega_\infty(\delta) = \omega_\infty(1)\delta^{2p/(2p+1)}$. The asymptotic efficiency bounds for the feasible one-sided CI follow from similar arguments, using Theorem 3.1 and Corollary 3.2 along with Theorem H.1 and Lemma G.3.

In the case where Assumption H.3 holds rather than Assumption H.4, it follows from Lemma H.7 that $\delta_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \delta_\infty$. Then, by Lemma H.9, the conditions in the last display of Theorem H.1 hold with $\hat{L}_{\delta_n,\tilde{\sigma}(\cdot)}$ playing the role of $\hat{L}^*$ and $\hat{L}_{\delta_n,\sigma(\cdot)}$ playing the role of $\hat{L}$. The results then follow from Theorem H.1 and the arguments above applied to the CIs based on $\hat{L}_{\delta_n,\sigma(\cdot)}$.

References

Armstrong, T. B. and M. Kolesár (2016a): "Optimal inference in a class of regression models," ArXiv:1511.06028v2.

——— (2016b): "Simple and honest confidence intervals in nonparametric regression," ArXiv: 1606.01200.

Brown, L. D. and M. G. Low (1996): "Asymptotic equivalence of nonparametric regression and white noise," *Annals of Statistics*, 24, 2384–2398.

Brown, L. D. and C.-H. Zhang (1998): "Asymptotic Nonequivalence of Nonparametric Experiments When the Smoothness Index is 1/2," *The Annals of Statistics*, 26, 279–287.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

Cheng, M.-Y., J. Fan, and J. S. Marron (1997): "On automatic boundary corrections," *The Annals of Statistics*, 25, 1691–1708.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81, 667–737.

Donoho, D. L. (1994): "Statistical Estimation and Optimal Recovery," *The Annals of Statistics*, 22, 238–270.

Donoho, D. L. and M. G. Low (1992): "Renormalization Exponents and Optimal Pointwise Rates of Convergence," *The Annals of Statistics*, 20, 944–970.

Drees, H. (1999): "On fixed-length confidence intervals for a bounded normal mean," *Statistics & Probability Letters*, 44, 399–404.

Durrett, R. (1996): *Probability: Theory and Examples*, Belmont, California: Duxbury Press, 2nd ed.

Fan, J. (1993): "Local Linear Regression Smoothers and Their Minimax Efficiencies," *The Annals of Statistics*, 21, 196–216.

Imbens, G. W. and K. Kalyanaraman (2012): "Optimal bandwidth choice for the regression discontinuity estimator," *The Review of Economic Studies*, 79, 933–959.

Lee, D. S. (2008): "Randomized experiments from non-random selection in U.S. House elections," *Journal of Econometrics*, 142, 675–697.

Nussbaum, M. (1996): "Asymptotic equivalence of density estimation and Gaussian white noise," *The Annals of Statistics*, 24, 2399–2430.

Reiss, M. (2008): "Asymptotic Equivalence for Nonparametric Regression with Multivariate and Random Design," *Annals of Statistics*, 36, 1957–1982.

Sacks, J. and D. Ylvisaker (1978): "Linear Estimation for Approximately Linear Models," *The Annals of Statistics*, 6, 1122–1137.

Stock, J. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identifcation and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews and J. Stock, Cambgridge, UK: Cambridge University Press, 80–108.

van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge, UK ; New York, NY, USA: Cambridge University Press.

von Bahr, B. and C.-G. Esseen (1965): "Inequalities for the rth Absolute Moment of a Sum of Random Variables, $1 \leq r \leq 2$," *The Annals of Mathematical Statistics*, 36, 299–303.

	$\sigma^2 = 0.1295$			$\sigma^2 = 4 \cdot 0.1295$		
CI method	Cov. (%)	Bias	RL	Cov. (%)	Bias	RL
Design 1, $(b_1, b_2) = (0.45, 0.75)$						
Conventional, $\hat{h}_{IK}$	10.1	-0.098	0.54	81.7	-0.099	0.72
RBC, $\hat{h}_{IK}, \rho = 1$	64.4	-0.049	0.80	93.9	-0.050	1.06
Conventional, $\hat{h}_{CCT}$	91.2	-0.010	1.01	92.7	-0.010	1.26
RBC, $\hat{h}_{CCT}$	93.7	0.003	1.18	93.6	0.007	1.48
FLCI, $C = 1$	94.6	-0.024	1	94.9	-0.069	1
FLCI, $C = 3$	96.7	-0.009	1.25	96.5	-0.028	1.25
Design 2, $(b_1, b_2) = (0.4, 0.9)$						
Conventional, $\hat{h}_{IK}$	54.2	-0.063	0.68	89.6	-0.085	0.77
RBC, $\hat{h}_{IK}, \rho = 1$	94.8	-0.006	1.00	95.9	-0.043	1.13
Conventional, $\hat{h}_{CCT}$	91.4	-0.009	1.02	92.7	-0.009	1.26
RBC, $\hat{h}_{CCT}$	93.6	0.003	1.19	93.6	0.007	1.49
FLCI, $C = 1$	94.5	-0.024	1	95.0	-0.065	1
FLCI, $C = 3$	96.8	-0.009	1.25	96.5	-0.028	1.25
Design 3, $(b_1, b_2) = (0.25, 0.65)$						
Conventional, $\hat{h}_{IK}$	87.8	-0.030	0.74	91.4	-0.009	0.76
RBC, $\hat{h}_{IK}, \rho = 1$	94.8	-0.014	1.09	95.0	-0.044	1.12
Conventional, $\hat{h}_{CCT}$	90.9	-0.014	0.97	92.8	-0.013	1.25
RBC, $\hat{h}_{CCT}$	92.2	-0.009	1.14	93.5	-0.007	1.48
FLCI, $C = 1$	94.7	-0.022	1	96.7	-0.028	1
FLCI, $C = 3$	96.8	-0.009	1.25	96.6	-0.025	1.25
Design 4, $f(x) = 0$						
Conventional, $\hat{h}_{IK}$	93.2	0.000	0.54	93.2	-0.001	0.72
RBC, $\hat{h}_{IK}, \rho = 1$	95.2	0.000	0.80	95.2	0.001	1.06
Conventional, $\hat{h}_{CCT}$	93.1	0.001	0.94	93.1	0.003	1.25
RBC, $\hat{h}_{CCT}$	93.5	0.001	1.12	93.5	0.004	1.48
FLCI, $C = 1$	96.8	0.001	1	96.9	0.000	1
FLCI, $C = 3$	96.8	0.001	1.25	96.8	0.002	1.25

Table S1: Monte Carlo simulation, $C = 1$. Coverage ("Cov") and relative length relative to optimal fixed-length CI for $\mathcal{F}_{RDH,2}(1)$ ("RL"). "Bias" refers to bias of estimator around which CI is centered. 11,000 simulation draws.

	$\sigma^2 = 0.1295$			$\sigma^2 = 4 \cdot 0.1295$		
CI method	Cov. (%)	Bias	RL	Cov. (%)	Bias	RL
Design 1, $(b_1, b_2) = (0.45, 0.75)$						
Conventional, $\hat{h}_{IK}$	0.1	-0.292	0.44	22.4	-0.296	0.58
RBC, $\hat{h}_{IK}, \rho = 1$	27.1	-0.127	0.65	77.8	-0.149	0.85
Conventional, $\hat{h}_{CCT}$	89.3	-0.019	0.94	91.6	-0.031	1.05
RBC, $\hat{h}_{CCT}$	93.7	0.004	1.06	93.7	0.012	1.22
FLCI, $C = 1$	67.3	-8.078	0.80	73.1	-0.209	0.80
FLCI, $C = 3$	94.5	-0.032	1	94.6	-0.089	1
Design 2, $(b_1, b_2) = (0.4, 0.9)$						
Conventional, $\hat{h}_{IK}$	60.0	-0.071	0.71	71.4	-0.193	0.72
RBC, $\hat{h}_{IK}, \rho = 1$	93.5	0.000	1.04	95.1	-0.020	1.05
Conventional, $\hat{h}_{CCT}$	89.7	-0.018	0.95	91.7	-0.029	1.05
RBC, $\hat{h}_{CCT}$	93.6	0.004	1.09	93.6	0.012	1.24
FLCI, $C = 1$	70.3	-0.073	0.80	76.3	-0.197	0.80
FLCI, $C = 3$	94.3	-0.030	1	94.6	-0.089	1
Design 3, $(b_1, b_2) = (0.25, 0.65)$						
Conventional, $\hat{h}_{IK}$	79.9	-0.052	0.76	89.2	-0.085	0.73
RBC, $\hat{h}_{IK}, \rho = 1$	93.3	0.001	1.13	94.6	-0.072	1.07
Conventional, $\hat{h}_{CCT}$	80.7	-0.032	0.87	91.8	-0.042	1.01
RBC, $\hat{h}_{CCT}$	86.2	-0.017	1.00	92.7	-0.027	1.20
FLCI, $C = 1$	73.5	-0.069	0.8	93.8	-0.084	0.80
FLCI, $C = 3$	94.4	-0.030	1	95.1	-0.078	1
Design 5, $f(x) = 0$						
Conventional, $\hat{h}_{IK}$	93.2	0.000	0.43	93.2	-0.001	0.57
RBC, $\hat{h}_{IK}, \rho = 1$	95.2	0.000	0.64	95.2	0.001	0.85
Conventional, $\hat{h}_{CCT}$	93.1	0.001	0.75	93.1	0.003	1.00
RBC, $\hat{h}_{CCT}$	93.5	0.001	0.89	93.5	0.004	1.18
FLCI, $C = 1$	96.8	0.001	0.80	96.9	0.000	0.80
FLCI, $C = 3$	96.8	0.001	1	96.7	0.002	1

Table S2: Monte Carlo simulation, $C = 3$. Coverage ("Cov") and relative length relative to optimal fixed-length CI for $\mathcal{F}_{RDH,2}(1)$ ("RL"). "Bias" refers to bias of estimator around which CI is centered. 11,000 simulation draws.

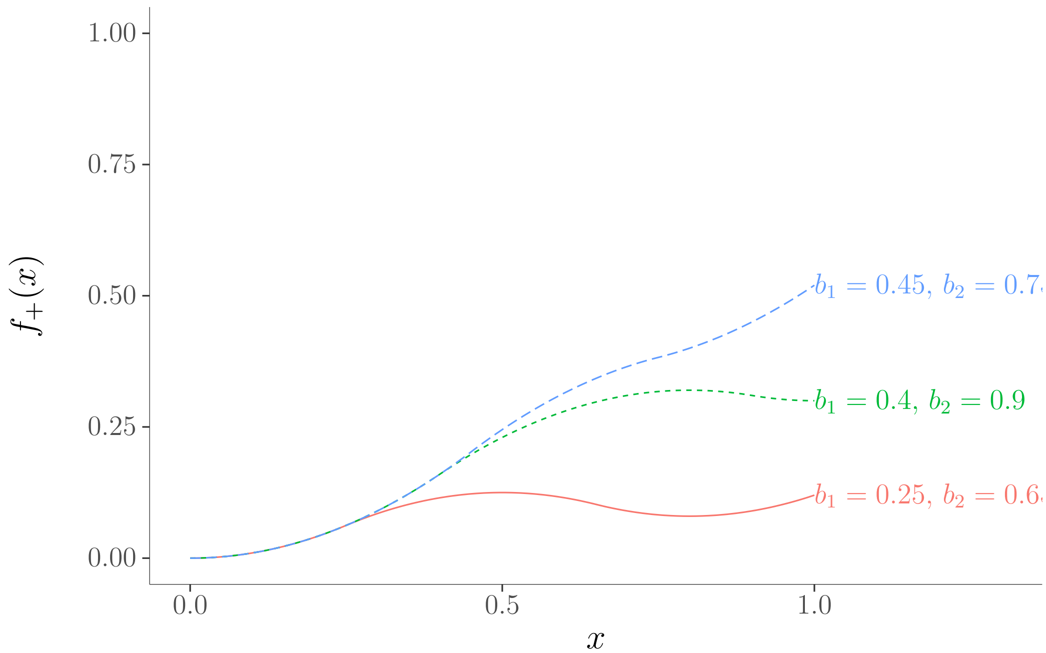

Figure S1: Regression function for Monte Carlo simulation, Designs 1–3, and $C = 1$. Knots $b_1 = 0.45, b_2 = 0.75$ correspond to Design 1, $b_1 = 0.4, b_2 = 0.9$ to Design 2, and $b_1 = 0.25, b_2 = 0.65$ to Design 3.