Supplemental Materials for "Optimal inference in a class of regression models"

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

May 26, 2016

These supplemental materials provide appendices not included in the main text. Supplemental Appendix D gives derivations for the examples given in Sections 4 and 5. Supplemental Appendix E considers feasible versions of the procedures in Section 3 in the case with unknown error distribution and derives their asymptotic efficiency. Supplemental Appendix F gives some auxiliary results used for relative asymptotic asymptotic efficiency comparisons. Supplemental Appendix G gives the proof of Theorem 5.1. Supplemental Appendix H contains additional figures for the Lee (2008) empirical example in Section 5.

Appendix D Details for calculations in Sections 4 and 5

This section gives details for the solutions to the modulus problem described in the examples in the main text.

D.1 Unconstrained Linear Regression

For the unconstrained linear regression model of Section 4.1.1, the modulus problem (24) reduces to $2\max_\theta \ell'\theta$ s.t $\|X\theta\| \leq \delta/2$. This leads to the Lagrangian

$$2\ell'\theta + \lambda\left(\delta^2/4 - \theta' X' X\theta\right),$$

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

which gives the first order condition

$$2\ell - 2\lambda X'X\theta_\delta^* = 0 \Longrightarrow X'X\theta_\delta^* = \ell/\lambda \Longrightarrow \theta_\delta^* = (X'X)^{-1}\ell/\lambda$$

where the last implication holds assuming $X'X$ is invertible. Using the formula for $\hat{L}_\delta$ under translation invariance and centrosymmetry, we obtain

$$\hat{L}_\delta = \frac{\theta_\delta^{*\prime} X'Y}{\theta_\delta^{*\prime} X'X\iota_L} = \frac{\ell'(X'X)^{-1}X'Y}{\ell'(X'X)^{-1}X'X\iota_L} = \ell'(X'X)^{-1}X'Y.$$

D.2 Linear Regression with Elliptical Constraints

For the linear regression model with elliptical constraints considered in Section 4.1.2, the Lagrangian becomes

$$2\ell'\theta + \lambda_1\left(\delta^2/4 - \theta'X'X\theta\right) + \lambda_2\left(C^2 - \theta'M'M\theta\right)$$

and the first order conditions give

$$\ell - \lambda_1 X'X\theta_\delta^* - \lambda_2 M'M\theta_\delta^* = 0 \Longrightarrow \theta_\delta^* = (\lambda_1 X'X + \lambda_2 M'M)^{-1}\ell$$

assuming $(\lambda_1 X'X + \lambda_2 M'M)$ is invertible. By the envelope theorem, we have $\frac{\omega'(\delta)}{2\delta} = \frac{d}{ds}\omega(\sqrt{s})|_{s=\delta^2} = \lambda_1/4$. Using the formula for $\hat{L}_\delta$ under centrosymmetry, this gives

$$\hat{L}_\delta = \frac{2\omega'(\delta)}{\delta}\theta_\delta^{*\prime}X'Y = \lambda_1\theta_\delta^{*\prime}X'Y = \ell'(X'X + (\lambda_2/\lambda_1)M'M)^{-1}X'Y$$

where λ_1 and λ_2 depend on δ and are given by the solution to the equations

$$\ell'(\lambda_1 X'X + \lambda_2 M'M)^{-1}M'M(\lambda_1 X'X + \lambda_2 M'M)^{-1}\ell = \theta_\delta^{*\prime}M'M\theta_\delta^* = C^2$$
$$\ell'(\lambda_1 X'X + \lambda_2 M'M)^{-1}X'X(\lambda_1 X'X + \lambda_2 M'M)^{-1}\ell = \theta_\delta^{*\prime}X'X\theta_\delta^* = \delta^2/4.$$

D.3 Regression Discontinuity

This section gives the details for the solution to the modulus problem $\omega(\delta; \mathcal{F}_{RDT,p})$ for the regression discontinuity parameter given in Section 5.1, which is a simple generalization of the results of Sacks and Ylvisaker (1978). It also gives details for implementing the optimal inference procedures in Section 5.3.

To describe the solution to the modulus problem, define

$$g_{b,C}(x) = g_{+,b,C}(x) - g_{-,b,C}(x),$$

where

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

and 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 $d_+ = (d_{+,1}, \ldots, d_{-,p-1})$, $d_- = (d_{-,1}, \ldots, d_{-,p-1})$, 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 the class $\mathcal{F}_{T,p}$ can be written as

$$f_+(x) = b_+ + \sum_{j=1}^{p-1} d_{+,j}x^j + r_+(x)$$

where $|r_+(x)| \leq C|x|^p$. Given b_+, d_+, the function $r_+(x)$ given by $-C|x|^p$ when $b_+ + \sum_{j=1}^{p=1} d_{+,j}x^j \geq C|x|^p$, $-\left(b_+ + \sum_{j=1}^{p=1} d_{+,j}x^j\right)$ when $\left|b_+ + \sum_{j=1}^{p=1} d_{+,j}x^j\right| < C|x|^p$ and $C|x|^p$ when $b_+ + \sum_{j=1}^{p=1} d_{+,j}x^j \leq -C|x|^p$ minimizes $|f_+(x_i)|$ simultaneously for all i. If $r_+(x)$ did not take this form, 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 off $2(f_+(0) + f_-(0))$ by increasing b_+ and leaving d_+ and r_+ the same. Thus, $r_+(x)$ takes the form given above, and plugging this in to the above display 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

$$0 = \sum_{i=1}^{n} \frac{g_{-,b,C}(x_i)}{\sigma^2(x_i)} \left(x_i, \ldots, x_i^{p-1}\right)', \tag{S1}$$

$$0 = \sum_{i=1}^{n} \frac{g_{+,b,C}(x_i)}{\sigma^2(x_i)} \left(x_i, \ldots, x_i^{p-1}\right)', \tag{S2}$$

and

$$\sum_{i=1}^{n} \frac{g_{+,b,C}(x_i)}{\sigma^2(x_i)} = \sum_{i=1}^{n} \frac{g_{-,b,C}(x_i)}{\sigma^2(x_i)}. \tag{S3}$$

The constraint in (25) must be binding at the optimum, which gives the additional equation

$$\delta^2/4 = \sum_{i=1}^{n} \frac{g_{b,C}(x_i)^2}{\sigma^2(x_i)} = b\sum_{i=1}^{n} \frac{g_{+,b,C}(x_i)}{\sigma^2(x_i)} - C\sum_{i=1}^{n} \frac{|g_{b,C}(x_i)||x_i|^p}{\sigma^2(x_i)}, \tag{S4}$$

where the equality follows from (S1)–(S2).

To get some intuition for these conditions, note that Conditions (S1) and (S2) ensure that $\hat{L}_\delta$ is unbiased for piecewise polynomial functions of the form $(b_+ + \sum_{j=1}^{p-1} d_+ x^j)1(x > 0) - (b_- + \sum_{j=1}^{p-1} d_- x^j)1(x < 0)$—otherwise the bias of $\hat{L}_\delta$ would be unbounded. Condition (S3) ensures that the weight given to observations on either side of the cutoff is the same, so that $\hat{L}_\delta$ is unbiased for constant functions. 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

$$\frac{\omega^{-1}(2b; \mathcal{F}_{RDT,p})^2}{4} = \inf_{f_+ - f_- \in \mathcal{F}_{RDT,p}(C)} \sum_{i=1}^{n} \left(\frac{f_+^2(x_i)}{\sigma^2(x_i)} + \frac{f_-^2(x_i)}{\sigma^2(x_i)} \right) \text{ s.t. } 2(f_+(0) + f_-(0)) \geq 2b \tag{S5}$$

for $b = b(\delta)$. Since the objective for the inverse modulus is strictly convex, this also shows that the solution is unique up to the values at the x_is. Using the fact that $\mathcal{F}_{RDT,p}(C)$ is translation invariant and symmetric then yields the expression (26).

The worst-case bias and standard deviation of $\hat{L}_{\delta,\mathcal{F}_{RDT,p}(C)}$ is given by

$$\overline{\text{bias}}(\hat{L}_{\delta,\mathcal{F}_{RDT,p}(C)}) = -\underline{\text{bias}}(\hat{L}_{\delta,\mathcal{F}_{RDT,p}(C)}) = b(\delta) - \frac{\sum_{i=1}^{n} g^2_{b(\delta),C}(x_i)/\sigma^2(x_i)}{\sum_{i=1}^{n} g_{+,b(\delta),C}(x_i)/\sigma^2(x_i)},$$
$$\text{sd}(\hat{L}_\delta) = \frac{\sqrt{\sum_{i=1}^{n} g^2_{b(\delta),C}(x_i)/\sigma^2(x_i)}}{\sum_{i=1}^{n} g_{+,b(\delta),C}(x_i)/\sigma^2(x_i)}.$$

To construct two-sided fixed-length CIs, we need to find δ_χ given in Theorem 3.2. It follows from Equation (S4) that finding δ_χ is equivalent to finding a solution b_χ to

$$b \cdot c_\chi \left(\sqrt{\sum_{i=1}^{n} \frac{g_{b,C}(x_i)^2}{\sigma^2(x_i)}} \right) \sum_{i=1}^{n} \frac{g_{+,b,C}(x_i)}{\sigma^2(x_i)} = \sum_{i=1}^{n} \frac{g_{b,C}(x_i)^2}{\sigma^2(x_i)},$$

with the least favorable function given by $g_{b_\chi,C}$. By similar arguments, the least favorable function for MSE estimators is given by $g_{b_\rho,C}$, where b_ρ solves

$$1 = C \sum_{i=1}^{n} \frac{|g_{b,C}(x_i)||x_i|^p}{\sigma^2(x_i)}.$$

Appendix E Unknown Error Distribution

The Gaussian regression model (8) makes the assumption of normal iid 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:

$$y_i = f(x_i) + u_i,\ \{u_i\}_{i=1}^n \sim Q,\ f \in \mathcal{F},\ Q \in \mathcal{Q}_n \tag{S6}$$

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. 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 (possibly data dependent) candidate variance function $\tilde{\sigma}^2(\cdot)$, let $K_{\tilde{\sigma}(\cdot),n} f = (f(x_1)/\tilde{\sigma}(x_1), \ldots, f(x_n)/\tilde{\sigma}(x_n))'$, 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 (12) with this choice of K and $Y = (y_1/\tilde{\sigma}(x_1), \ldots, y_n/\tilde{\sigma}(x_n))'$, and let $f^*_{\tilde{\sigma}(\cdot),\delta}$ and $g^*_{\tilde{\sigma}(\cdot),\delta}$ denote the least favorable functions used in forming this estimate. We allow δ to be data dependent as well unless stated otherwise, but leave this implicit in the notation. We assume throughout this section that $\mathcal{G} \subseteq \mathcal{F}$. More generally, we will consider estimators of the form

$$\hat{L} = a_n + \sum_{i=1}^{n} w_{i,n} y_i \tag{S7}$$

where a_n and $w_{i,n}$ are a (potentially data dependent) sequence and triangular array respectively. For a class $\mathcal{G}$, define

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

In the case where a_n and $w_{i,n}$ are deterministic, $\overline{\text{bias}}_{\mathcal{G}}(\hat{L})$ and $\underline{\text{bias}}_{\mathcal{G}}(\hat{L})$ give the worst-case bias on each side over the class $\mathcal{G}$, which matches the definition in Section 3. However, with random a_n and $w_{i,n}$, $\overline{\text{bias}}_{\mathcal{G}}(\hat{L})$ and $\underline{\text{bias}}_{\mathcal{G}}(\hat{L})$ are random and no longer give the worst-case bias. The arguments used to derive the formula (13) extend to this definition of $\overline{\text{bias}}_{\mathcal{G}}$, so that $\overline{\text{bias}}_{\mathcal{F}}(\hat{L}_{\delta,\mathcal{F},\mathcal{G},\tilde{\sigma}(\cdot)}) = -\underline{\text{bias}}_{\mathcal{G}}(\hat{L}_{\delta,\mathcal{F},\mathcal{G},\tilde{\sigma}(\cdot)}) = \frac{1}{2}(\omega_{n,\tilde{\sigma}(\cdot)}(\delta;\mathcal{F},\mathcal{G}) - \delta\omega'_{n,\tilde{\sigma}(\cdot)}(\delta;\mathcal{F},\mathcal{G}))$.

Suppose that an estimate $\widehat{\text{se}}_n$ of the standard deviation of $\hat{L}$ is available such that the uniform central limit theorem

$$\frac{\sum_{i=1}^{n} w_{i,n} u_i}{\widehat{\text{se}}_n} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{d}{\to}} N(0,1) \tag{S8}$$

holds. Using these quantities, one can form analogues of the CIs treated in Theorem 3.4, where the worst-case bias and standard deviation are replaced by $\overline{\text{bias}}_{\mathcal{F}}(\hat{L})$ and $\widehat{\text{se}}_n$. The following theorem shows that this gives asymptotically valid CIs.

Theorem E.1. *Let $\hat{L}$ be an estimator of the form (S7), and suppose that (S8) holds. 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*

$$\liminf_{n\to\infty} \inf_{f\in\mathcal{F},Q\in\mathcal{Q}_n} P_{f,Q}\left(Lf \in [\hat{c},\infty)\right) \geq 1-\alpha \tag{S9}$$

and

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

If, in addition,

$$\frac{\overline{bias}_{\mathcal{F}}(\hat{L}) - \overline{bias}_{n,Q}}{s_{n,Q}} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0, \qquad \frac{\underline{bias}_{\mathcal{G}}(\hat{L}) - \underline{bias}_{n,Q}}{s_{n,Q}} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0, \qquad \frac{\widehat{se}_n}{s_{n,Q}} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 1 \tag{S10}$$

for some nonrandom sequences $\overline{bias}_{n,Q}$, $\underline{bias}_{n,Q}$ and $s_{n,Q}$, then the same statements hold with $\overline{bias}_{n,Q}$, $\underline{bias}_{n,Q}$ and $s_{n,Q}$ replacing $\overline{bias}_{\mathcal{F}}(\hat{L})$, $\underline{bias}_{\mathcal{G}}(\hat{L})$ and $\widehat{se}_n$. In addition, the worst-case βth quantile excess length of the one-sided CI over $\mathcal{G}$ will satisfy

$$\limsup_{n\to\infty} \sup_{Q\in\mathcal{Q}_n} \frac{\sup_{g\in\mathcal{G}} q_{g,Q,\beta}(\hat{c}-Lg)}{\overline{bias}_{n,Q}-\underline{bias}_{n,Q}+s_{n,Q}(z_{1-\alpha}+z_\beta)} \le 1 \tag{S11}$$

and the length of the two-sided CI will satisfy

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

where $b_{n,Q}=\max\{|\overline{bias}_{n,Q}|,|\underline{bias}_{n,Q}|\}$. If (S10) holds with

$$\begin{aligned}
&\overline{bias}_{n,Q}=-\underline{bias}_{n,Q}=\frac{1}{2}\left(\omega_{\sigma_Q(\cdot),n}(z_{1-\alpha}+z_\beta)-(z_{1-\alpha}+z_\beta)\omega'_{\sigma_Q(\cdot),n}(z_{1-\alpha}+z_\beta)\right),\\
&s_{n,Q}=\omega'_{\sigma_Q(\cdot),n}(z_{1-\alpha}+z_\beta)
\end{aligned} \tag{S12}$$

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 (S9) can satisfy (S11) with the constant 1 replaced by a strictly smaller constant on the right hand side.

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

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

using the fact that $\overline{\mathrm{bias}}_{\mathcal{F}}(\hat{L})+\sum_{i=1}^n w_{i,n}u_i\ge \hat{L}-Lf$ for all $f\in\mathcal{F}$ by the definition of $\overline{\mathrm{bias}}_{\mathcal{F}}$. The right hand side converges to $1-\alpha$ uniformly over $f\in\mathcal{F}$ and $Q\in\mathcal{Q}_n$ by (S8). For coverage of the two-sided CI, note that $(\hat{L}-Lf)/\widehat{\mathrm{se}}_n=Z_n+s$ where $s=\left(a_n+\sum_{i=1}^n w_{i,n}f(x_i)-Lf\right)/\widehat{\mathrm{se}}_n$, and $|s|\le b/\widehat{\mathrm{se}}_n$ by definition. Hence,

$$\begin{aligned}
P_{f,Q}\left(Lf\in\left\{\hat{L}\pm\mathrm{cv}_\alpha\left(b/\widehat{\mathrm{se}}_n\right)\widehat{\mathrm{se}}_n\right\}\right)&=P_{f,Q}\left(\left|\hat{L}-Lf\right|/\widehat{\mathrm{se}}_n\le \mathrm{cv}_\alpha\left(b/\widehat{\mathrm{se}}_n\right)\right)\\
&\ge \inf_{|s|\le t} P_{f,Q}\left(|Z_n+s|\le \mathrm{cv}_\alpha(t)\right).
\end{aligned}$$

By definition of cv_α, the difference between the last probability and $1-\alpha$ is bounded by

$$\left|\inf_{|s|\leq t} P_{f,Q}\left(|Z_n+s|\leq \mathrm{cv}_\alpha(t)\right) - \inf_{|s|\leq t} P\left(|Z+s|\leq \mathrm{cv}_\alpha(t)\right)\right| \\ \leq \sup_{c,d}\left|P_{f,Q}\left(c\leq Z_n\leq d\right) - P\left(c\leq Z\leq d\right)\right|,$$

which converges to zero uniformly over $f\in\mathcal{F}$ and $Q\in\mathcal{Q}_n$ (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). Similar arguments apply under (S10) with $\overline{\mathrm{bias}}_{n,Q}$, $\underline{\mathrm{bias}}_{n,Q}$ and $s_{n,Q}$ replacing $\overline{\mathrm{bias}}_{\mathcal{F}}(\hat{L})$, $\underline{\mathrm{bias}}_{\mathcal{G}}(\hat{L})$ and $\widehat{\mathrm{se}}_n$.

To show (S11), note that,

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

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

$$\frac{Lg-\hat{c}}{\overline{\mathrm{bias}}_{n,Q} - \underline{\mathrm{bias}}_{n,Q} + s_{n,Q}(z_{1-\alpha}+z_\beta)} - 1 \\ \leq \frac{\frac{\overline{\mathrm{bias}}_{\mathcal{F}}(\hat{L})-\overline{\mathrm{bias}}_{n,Q}}{s_{n,Q}} - \frac{\underline{\mathrm{bias}}_{\mathcal{G}}(\hat{L})-\underline{\mathrm{bias}}_{n,Q}}{s_{n,Q}} + \frac{\widehat{\mathrm{se}}_n}{s_{n,Q}}(z_{1-\alpha}-Z_n) - (z_{1-\alpha}+z_\beta)}{\left(\overline{\mathrm{bias}}_{n,Q}-\underline{\mathrm{bias}}_{n,Q}\right)/s_{n,Q} + (z_{1-\alpha}+z_\beta)}.$$

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 convergence of the relative length of the two-sided CIs, note that

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

which converges to zero uniformly over $f\in\mathcal{F}, Q\in\mathcal{Q}_n$ since $\mathrm{cv}_\alpha(t)$ is bounded from below and uniformly continuous with respect to t.

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 $\mathcal{F}, \mathcal{Q}_n$. Under (S12), the denominator in (S11) for $Q = Q_n$ is equal to $\omega_{\sigma_{Q_n}(\cdot),n}(z_{1-\alpha}+z_\beta)$, which gives

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

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 E.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 E.1 gives a high level result that allows for different estimators or guesses of the variance function $\tilde{\sigma}(\cdot)$ used when forming $\hat{L}_{\delta,\tilde{\sigma}(\cdot),n}$, as well as different choices for the estimate $\widehat{se}_n^2$ of the variance of the estimator. The best choices will depend on the particulars of the situation. If the variance function is difficult to estimate, one can compromise by choosing $\tilde{\sigma}(\cdot)$ to be a constant function such as an estimate of the average of the variance function or an estimate of the variance function at a particular point. In forming the standard error $\widehat{\text{se}}_n$, one can use the formula $\omega'_{\tilde{\sigma}(\cdot),n}(\delta)$ from Theorem 3.1 if $\tilde{\sigma}(\cdot)$ is a good enough estimate of the variance function. If not, one can use a heteroskedasticity consistent variance estimator. This mirrors the situation in the special case where $\mathcal{F}$ imposes a (otherwise unconstrained) linear model: one can use OLS or feasible weighted least squares. In the former case, one must use a sandwich variance estimator if there is heteroskedasticity. In the latter case, one need not use the sandwich variance estimator if the conditional variance can be estimated well enough, although one may want to use it anyway to guard against misspecification in the conditional variance estimator. This approach has been proposed, for instance, in Wooldridge (2010).[1]

[1] See also Romano and Wolf (2015) for a recent discussion of these issues in the context of the unconstrained linear model.

With this in mind, we now introduce an estimator and standard error formula that give asymptotic coverage for essentially arbitrary statistics L under generic 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

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

We consider the estimate

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

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

$$\sup_n \sup_{x,x' :\, d_X(x,x') \leq \eta} \max\left\{\left|f^*_{\tilde{\sigma}(\cdot),\delta}(x) - f^*_{\tilde{\sigma}(\cdot),\delta}(x')\right|, \left|g^*_{\tilde{\sigma}(\cdot),\delta}(x) - g^*_{\tilde{\sigma}(\cdot),\delta}(x')\right|\right\} \leq \overline{g}(\eta), \tag{S13}$$

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

$$\min_{1\leq i\leq n} \sum_{j=1}^n I\left(d_X(x_j, x_i) \leq \eta\right) \to \infty. \tag{S14}$$

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

$$\sup_{f\in\mathcal{F}} \max_{1\leq i\leq n} |\hat{f}(x_i) - f(x_i)| \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0. \tag{S15}$$

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,

$$1/K \le \sigma_Q^2(x_i) \le K \text{ and } E_Q|u_i|^{2+\eta} \le K. \tag{S16}$$

In cases where function class $\mathcal{F}$ imposes smoothness on f, (S13) will often follow directly from the definition of $\mathcal{F}$. For example, it holds for the Lipschitz class $\{f \colon |f(x) - f(x')| \le Cd_X(x,x')\}$. The condition (S14) 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 (S15) will hold under regularity conditions for a variety of choices of $\hat{f}$. It is worth noting that smoothness assumptions $\mathcal{F}$ 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, (S13) 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).

The condition (S16) is used to verify a Lindeberg condition for the central limit theorem used to obtain (S8), which we do in the next Lemma.

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

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

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

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

Then

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

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

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

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

Theorem E.2. *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 (S13) and (S14) hold. Suppose the errors satisfy (S16) and are independent over i for all n and $Q \in \mathcal{Q}_n$. Then (S8) holds with $\widehat{se}_n$ given by $var_Q(\hat{L}_{\delta,\tilde{\sigma}(\cdot)})$. If, in addition, the estimator $\hat{f}$ satisfies (S15), then $\widehat{se}^2_{\delta,\tilde{\sigma}(\cdot),n}/var_Q(\hat{L}_{\delta,\tilde{\sigma}(\cdot)}) \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 1$ and (S8) holds with $\widehat{se}_n$ given by $\widehat{se}^2_{\delta,\tilde{\sigma}(\cdot),n}$.*

Proof. The first part of the theorem will follow by applying Lemma E.1 to show convergence under arbitrary sequences $Q_n \in \mathcal{Q}_n$ so long as

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

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

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

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 $\max_{1\leq i\leq n}(f(x_i)-\hat{f}(x_i))^2\sum_{i=1}^n a_{n,i}\tilde{\sigma}^2(x_i) = \max_{1\leq i\leq n}(f(x_i)-\hat{f}(x_i))^2\delta^2$, which converges in probability to zero uniformly over f and Q by assumption. Similarly, the last term is bounded by $\max_{1\leq i\leq n}|f(x_i)-\hat{f}(x_i)|$ 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

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

which converges to zero since $\max_{1\leq i\leq n} a_{n,i} \to 0$ 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 (E.2) 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 E.1. *If, in addition to the conditions in Theorem E.2, $\sigma_Q^2(x) = \sigma^2\cdot\tilde{\sigma}^2(x)$ for all n and $Q \in \mathcal{Q}_n$, then (S10) and (S12) hold with $\beta = \Phi(\delta/\sigma - z_{1-\alpha})$.*

Appendix F 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 5.1 in Supplemental Appendix G.

Before stating these results, we recall the characterization of minimax affine performance given in Donoho (1994), which we restated in Theorem 3.2: the optimal affine minimax root MSE is given by

$$\frac{\omega(\delta_\rho)}{\delta_\rho}\sqrt{\rho_A\left(\frac{\delta_\rho}{2\sigma}\right)}\sigma,$$

and is achieved by $\hat{L}_{\delta_\rho}$, where δ_ρ solves $c_\rho(\delta/(2\sigma)) = \delta\omega'(\delta)/\omega(\delta)$, assuming a solution exists. Donoho (1994) shows that, when this holds, δ_ρ maximizes the above display over all values of $\delta > 0$, so that the minimax root MSE is given by

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

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. 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 F.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 $(iii) \implies (i)$. To show $(i) \implies (iii)$, 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 F.2. *Suppose that the conditions of Lemma F.1 hold with* $\lim_{\delta\to0}\omega_\infty(\delta)=0$ *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 F.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, $0 \leq \omega_n(\delta) \leq \omega_n(1)\max\{\delta, 1\} \leq (\omega_n(1) + 1)\max\{\delta, 1\}$ for all $\delta > 0$. The result then follows from the dominated convergence theorem. □

Lemma F.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 G Asymptotics for Regression Discontinuity

This section proves Theorem 5.1. We first give a general result for linear estimators under high-level conditions in Section G.1. We then consider local polynomial estimators in Section G.2 and optimal estimators with a plug-in variance estimate in Section G.3. Theorem 5.1 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 Section 5.5, using the conditions in that section. We repeat these conditions here for convenience.

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

Assumption G.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_i*s 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 5.1 considers affine estimators that are optimal under the assumption that the variance function is given by $\hat{\sigma}_+ I(x > 0) + \hat{\sigma}_- I(x < 0)$, 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 G.3. $\tilde{\sigma}(x) = \hat{\sigma}_+ I(x > 0) + \hat{\sigma}_- I(x < 0)$ *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 G.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 G.3 corresponds to the estimate of the variance function used in the application. It generalizes Assumption 5.3 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 G.4 is used in deriving bounds based on infeasible estimates that use the true variance function.

Note that, under Assumption G.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 G.4 and G.3.

G.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

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

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

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

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

Under Assumption G.5, the sequences of kernels k_n^+ and k_n^- are nonrandom: they can depend on the x_i's but not on the y_i's. We will also consider estimators where k_n^+ and k_n^- depend on the Y_i's, usually through a preliminary estimate of the variance (as with the optimal affine estimator computed with an estimated variance function). For these, we use the following assumption as an alternative to Assumption G.5 (i).

Assumption G.6. *For some constant* B, k_n^+ *has support contained in* $[0, B]$ *with probability approaching one uniformly over* $\mathcal{F}, \mathcal{Q}_n$. *In addition, for some* k^+ *with* $\int k^+(u)\,du > 0$, $\sup_x |k_n^+(x) - k^+(x)| \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0$, *and* $\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}[k_n^+(x_i/h_n) - k^+(x_i/h_n)]u_i \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} 0$. *The same holds for* k_n^- *for some* k^-.

Let

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

and

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

Note that, when k_n^+ and k_n^- are nonrandom, v_n is the (constant over $Q \in \mathcal{Q}_n$) variance of $\hat{L}$, and that $\overline{\text{bias}}_n = \sup_{f\in\mathcal{F}}(E_{f,Q}\hat{L} - Lf) = -\inf_{f\in\mathcal{F}}(E_{f,Q}\hat{L} - Lf)$ for any $Q \in \mathcal{Q}_n$ under Assumption G.5 (ii), since, for any $f \in \mathcal{F}$, we can write $f_+(x) = f_+(0) + \sum_{j=1}^{p-1}\frac{f_+^{(j)}(0)}{j!}x^j + r_+(x)$ and $f_-(x) = f_-(0) + \sum_{j=1}^{p-1}\frac{f_-^{(j)}(0)}{j!}x^j + r_-(x)$ with $|r_+(x)| \le C|x|^p$ and $|r_-(x)| \le C|x|^p$, which gives

$$E_{f,Q}\hat{L} - Lf =$$
$$\frac{\sum_{i=1}^n k_n^+(x_i/h_n)I(x_i>0)(f_+(x_i)-f_+(0))}{\sum_{i=1}^n k_n^+(x_i/h_n)I(x_i>0)} + \frac{\sum_{i=1}^n k_n^-(x_i/h_n)I(x_i<0)(f_-(x_i)-f_-(0))}{\sum_{i=1}^n k_n^-(x_i/h_n)I(x_i<0)}$$
$$= \frac{\sum_{i=1}^n k_n^+(x_i/h_n)I(x_i>0)Cr_+(x)}{\sum_{i=1}^n k_n^+(x_i/h_n)I(x_i>0)} + \frac{\sum_{i=1}^n k_n^-(x_i/h_n)I(x_i<0)Cr_-(x)}{\sum_{i=1}^n k_n^-(x_i/h_n)I(x_i<0)},$$

which is maximized over $f \in \mathcal{F}$ by taking any function with $r_+(x) = C|x|^p I(x>0)$ $\text{sign}(k_n^+(x/h_n))$ and $r_-(x) = C|x|^p I(x<0)\text{sign}(k_n^-(x/h_n))$ and minimized by taking the negative of these functions. In the case where k_n^+ and k_n^- are random, we keep the definitions of v_n and $\overline{\text{bias}}_n$ given above, but we note that, in this case, v_n and $\overline{\text{bias}}_n$ are random variables and do not give the actual variance and maximum bias of $\hat{L}$.

To form a feasible CI, we need an estimate of v_n. 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 G.1. *Suppose that Assumption G.1 holds. If Assumption G.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 G.4 with* $\tilde\sigma_+(0) = \sigma_+(0)$ *and* $\tilde\sigma_-(0) = \sigma_-(0)$*, then* $\lim_{n\to\infty} n^{2p/(2p+1)}\tilde v_n = v_\infty$*. If Assumption G.5 holds with part (i) replaced with Assumption G.6, then* $n^{p/(2p+1)}\overline{bias}_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \overline{bias}_\infty$*. If, in addition,* $\tilde\sigma(\cdot)$ *satisfies Assumption G.3 or G.4 with* $\tilde\sigma_+(0) = \sigma_+(0)$ *and* $\tilde\sigma_-(0) = \sigma_-(0)$*, then* $\tilde v_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} v_\infty$*.*

Proof. The results are immediate from the calculations preceding the lemma. □

Theorem G.1. *Suppose that Assumptions G.1 and G.2 hold, Assumption G.3 or G.4 holds with* $\tilde\sigma_+(0) = \sigma_+(0)$ *and* $\tilde\sigma_-(0) = \sigma_-(0)$*, and that Assumption G.5 holds, with part (i) possibly replaced by Assumption G.6. Then*

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

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

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

In addition, $n^{p/(2p+1)}\,\mathrm{cv}_\alpha\left(\overline{bias}_n/\tilde v_n\right)\tilde v_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \mathrm{cv}_\alpha\left(\overline{bias}_\infty/v_\infty\right)v_\infty$ *under Assumption G.3 and* $n^{p/(2p+1)}\,\mathrm{cv}_\alpha\left(\overline{bias}_n/\tilde v_n\right)\tilde v_n \to \mathrm{cv}_\alpha\left(\overline{bias}_\infty/v_\infty\right)v_\infty$ *under Assumption G.4. The minimax* β *quantile of the one-sided CI is given asymptotically by*

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

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

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

Proof. For the first three displays, we verify the conditions of Theorem E.1 with $\overline{\mathrm{bias}}_{n,Q} =$

$n^{-p/(2p+1)}\overline{\text{bias}}_\infty$, $\underline{\text{bias}}_{n,Q} = -n^{-p/(2p+1)}\overline{\text{bias}}_\infty$ and $s_{n,q} = n^{-p/(2p+1)}\sqrt{v_\infty}$. This gives the first three displays with $=$ replaced by $\geq$ in the first two displays and $\leq$ in the third display. We then give sequences f_n under which each of the statements hold with equality with the supremum or infimum over $\mathcal{F}$ replaced by $\{f_n\}$.

Let

$$\text{bias}_n(f) = \\ \frac{\sum_{i=1}^n k_n^+(x_i/h_n)I(x_i>0)(f(x_i)-f_+(0))}{\sum_{i=1}^n k_n^+(x_i/h_n)I(x_i>0)} + \frac{\sum_{i=1}^n k_n^-(x_i/h_n)I(x_i<0)(f_-(x_i)-f_-(0))}{\sum_{i=1}^n k_n^-(x_i/h_n)I(x_i<0)}.$$

Then $\overline{\text{bias}}_n = \sup_{f\in\mathcal{F}} \text{bias}_n(f) = -\inf_{f\in\mathcal{F}} \text{bias}_n(f)$ (note that, as with $\overline{\text{bias}}_n$, $\text{bias}_n(f)$ is a random quantity in the case where k_n^+ and k_n^- are random). Note that

$$\frac{\hat{L} - Lf - \text{bias}_n(f)}{\sqrt{v_n}} = \frac{\sum_{i=1}^n w_{i,n}u_i}{\sqrt{\sum_{i=1}^n w_{n,i}^2\sigma^2(x_i)}}.$$

where $w_{i,n}$ is given by $w_{i,n} = k_n^+(x_i/h_n)/\sum_{j=1}^n k_n^+(x_j/h_n)I(x_j>0)$ for $x_i > 0$ and $w_{i,n} = k_n^-(x_i/h_n)/\sum_{j=1}^n k_n^-(x_j/h_n)I(x_j<0)$ for $x_i < 0$. Under Assumption G.4, this converges in distribution to a $N(0,1)$ law uniformly over $\mathcal{F}, \mathcal{Q}_n$ by Lemma E.1. Under Assumption G.3, the same convergence can be seen to hold by noting that the sequence is asymptotically equivalent to the same sequence with k_n^+ replaced by k^+ and k_n^- replaced by k^-. The same central limit theorem holds with v_n replaced by $\tilde{v}_n$ by Lemma G.1. This verifies condition (S8). Condition (S10) holds with $\overline{\text{bias}}_{n,Q} = n^{-p/(2p+1)}\overline{\text{bias}}_\infty$, $\underline{\text{bias}}_{n,Q} = -n^{-p/(2p+1)}\overline{\text{bias}}_\infty$ and $s_{n,q} = n^{-p/(2p+1)}\sqrt{v_\infty}$ by Lemma G.1.

To show that these statements hold with equality, we consider the sequences f_n and $-f_n$ with $f_n(x) = C|x|^p I(x>0)\text{sign}(k^+(x/h_n)) - C|x|^p I(x<0)\text{sign}(k^-(x/h_n))$. Note that $n^{p/(2p+1)}(\overline{\text{bias}}_n - \text{bias}_n(f_n))$ is equal to the sum of

$$n^{p/(2p+1)}h_n^p \frac{\frac{1}{nh_n}\sum_{i=1}^n k_n^+(x_i/h_n)I(x_i>0)C|x_i/h_n|^p\left[\text{sign}(k_n^+(x_i/h_n)) - \text{sign}(k^+(x_i/h_n))\right]}{\frac{1}{nh_n}\sum_{i=1}^n k_n^+(x_i/h_n)I(x_i>0)}$$

and the corresponding term for the negative observations. For any i such that

$$\text{sign}(k_n^+(x_i/h_n)) - \text{sign}(k^+(x_i/h_n)) \neq 0,$$

we must have $|k_n^+(x_i/h_n)| \leq \sup_x |k_n^+(x) - k^+(x)|$. It follows from this and the bound on the support of k_n^+ that the above display converges in probability to zero uniformly over $\mathcal{F}, \mathcal{Q}_n$. From this and a similar argument for the negative observations, it follows that $n^{p/(2p+1)}\left(\overline{\text{bias}}_n - \text{bias}_n(f_n)\right) \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0$. Similarly, $n^{p/(2p+1)}\left(-\overline{\text{bias}}_n - \text{bias}_n(-f_n)\right) \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0$. From this and Lemma G.1, it follows that $(\hat{L} - Lf_n)/\sqrt{\tilde{v}_n} \underset{\{f_n\},\mathcal{Q}_n}{\overset{d}{\to}} N(\overline{\text{bias}}_\infty/\sqrt{v_\infty}, 1)$ and $(\hat{L} - L(-f_n))/\sqrt{\tilde{v}_n} \underset{\{-f_n\},\mathcal{Q}_n}{\overset{d}{\to}} N(-\overline{\text{bias}}_\infty/\sqrt{v_\infty}, 1)$. Replacing the infimum in the first two displays with f_n, Q_n for any $Q_n \in \mathcal{Q}_n$ and replacing the supremum in the third display with $-f_n, Q_n$, it follows that these limits hold with equality.

For the last display, note that $\text{bias}(f) = 0$ for any $f \in \mathcal{F}_{RDT,p}(0)$, so

$$(\hat{L} - Lf)/\sqrt{\tilde{v}_n} \underset{\mathcal{F}_{RDT,p}(0),\mathcal{Q}_n}{\overset{d}{\to}} N(0,1).$$

Thus, letting $Z_n = -(\hat{L} - Lf)/\sqrt{\tilde{v}_n}$, we have

$$\begin{aligned} n^{p/(2p+1)} & \sup_{f\in\mathcal{F}_{RDT,p}(0),Q\in\mathcal{Q}_n} q_{f,Q,\beta}(Lf - \hat{c}) \\ &= n^{p/(2p+1)} \sup_{f\in\mathcal{F}_{RDT,p}(0),Q\in\mathcal{Q}_n} q_{f,Q,\beta}\left(\sqrt{\tilde{v}_n}Z_n + \overline{\text{bias}}_n + z_{1-\alpha}\sqrt{\tilde{v}_n}\right) \\ &\to \overline{\text{bias}}_\infty + (z_\beta + z_{1-\alpha})\sqrt{v_\infty}. \end{aligned}$$

□

G.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

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

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 Section G.1, with $h_n = h_{n,+}$,

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

and

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

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 G.1, for k_+^* and k_-^* bounded with bounded support, $\frac{1}{nh_n} \sum_{i=1}^n p(x_i/h_{+,n}) p(x_i/h_{+,n})' k_+^*(x_i/h_{+,n}) \cdot I(x_i > 0) \to M^+ p_{X,+}(0)$ and similarly $\frac{1}{nh_n} \sum_{i=1}^n p(x_i/h_{-,n}) p(x_i/h_{-,n})' k_+^*(x_i/h_{-,n}) \cdot I(x_i < 0) \to M^- p_{X,-}(0)$. Furthermore, Assumption G.5 (ii) follows immediately from the normal equations for the local polynomial estimator. This gives the following result.

Theorem G.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 G.5 holds for the local polynomial estimator so long as Assumption G.1 holds.*

G.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 G.3 or Assumption G.4. We use the same notation as in Section 5, 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 (25) in Section 5.1 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 (26) in Section 5.1 with $\tilde{\sigma}(\cdot)$ in place of $\sigma(\cdot)$.

We will deal with the inverse modulus, and use Lemma F.1 to obtain results for the modulus itself. The inverse modulus $\omega^{-1}_{\tilde{\sigma}(\cdot),n}(2b)$ is given by Equation (S5) in Section D.3, 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 G.4, we will assume that $\tilde{b}_n \to \tilde{b}_\infty$ for some $\tilde{b}_\infty > 0$. Under Assumption G.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 G.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 Section D.3. 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 $\mathcal{A}(x_i,b,d) = h_n^p b + \sum_{j=1}^{p-1} h_n^{p-j} d_j x_i^j$,

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

with

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

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

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 $\tilde{\sigma}(x) = c$. The estimator $\hat{L}_{\delta,\tilde{\sigma}(\cdot)}$ with $\omega_{\tilde{\sigma}(\cdot),n}(\delta) = 2b_n$ takes the general kernel form in Section G.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 Section G.1, $\overline{\text{bias}}_n$ (which is a random quantity under Assumption G.3) 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 Theorem 3.4). If δ is chosen to minimize the length of the fixed-length CI, the half-length will be given by

$$\mathrm{cv}_\alpha(\overline{\mathrm{bias}}_n/\sqrt{\tilde{v}_n})\sqrt{\tilde{v}_n} = \sup_{\delta>0} \frac{\omega_{\tilde{\sigma}(\cdot),n}(\delta)}{\delta}\chi_{A,\alpha}\left(\frac{\delta}{2}\right),$$

and δ will achieve the maximum (see Theorem 3.2 and the discussion at the beginning of Supplemental Appendix F).

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 $f_+(u) = \tilde{\sigma}_+^2(0)k^+_{\tilde{\sigma}_+(0)}(u; b_+, d_+) = k_1^+(u; b_+, d_+)$ 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 Section D.3). 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_∞.

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 $\bar{d}_+ = (d_{+,1}/\tilde{b}_\infty^{(p-1)/p}, \ldots, d_{+,p-1}/\tilde{b}_\infty^{1/p})'$ and similarly for $\bar{d}_-$. This uses the fact that, for any $h > 0$, $h^p k_1^+(u/h; b_+, d_+) = k_1^+(u; b_+ h^p, d_{+,1} h^{p-1}, d_{+,2} h^{p-2}, \ldots, d_{+,p-1} h)$ 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 $\omega_\infty(\delta) = \omega_{\tilde{\sigma}_+(0), \tilde{\sigma}_-(0), \infty}(\delta) = \delta^{2p/(2p+1)} \omega_{\tilde{\sigma}_+(0), \tilde{\sigma}_-(0), \infty}(1)$.

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. We recall that, by results in Donoho (1994), choosing δ to optimize the fixed-length CI criterion leads to maximizing $\frac{\omega_{\tilde{\sigma}(\cdot),n}(\delta)}{\delta} \chi_{A,\alpha}\left(\frac{\delta}{2}\right)$, and the half-length of the resulting CI is given by the maximized value of this formula. Similarly, choosing δ to optimize minimax MSE leads to maximizing $\frac{\omega_{\tilde{\sigma}(\cdot),n}(\delta)}{\delta} \sqrt{\rho_A\left(\frac{\delta}{2}\right)}$. (See Theorem 3.2 and the discussion at the beginning of Supplemental Appendix F.)

Theorem G.3. *Suppose Assumptions G.1 and G.2 hold. Let* $k^+(x) = k^+_{\tilde{\sigma}_+(0)}(x; \tilde{b}_{+,\infty}, \tilde{d}_{+,\infty})$ *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* $\tilde{b}_{+,\infty} + \tilde{b}_{-,\infty} = \omega_\infty(\delta^*)/2$ *for some* $\delta^* > 0$*. Then, under Assumption G.4, Assumption G.5 holds for* $\hat{L}_{\delta, \tilde{\sigma}(\cdot)}$ *with this definition of* k^+ *and* k^- *so long as* $\delta_n \to \delta^*$*. Under Assumption G.3, Assumption G.6 holds for* $\hat{L}_{\delta, \tilde{\sigma}(\cdot)}$ *with this definition of* k^+ *and* k^- *so long as* $\delta_n \xrightarrow[\mathcal{F}, \mathcal{Q}_n]{p} \delta^*$*. If* δ_n *is chosen to maximize* $\frac{\omega_{\tilde{\sigma}(\cdot),n}(\delta)}{\delta} \sqrt{\rho_A\left(\frac{\delta}{2}\right)}$*, then this holds with* δ^* *maximizing* $\frac{\omega_{\tilde{\sigma}_+(0), \tilde{\sigma}_-(0), \infty}(\delta)}{\delta} \sqrt{\rho_A\left(\frac{\delta}{2}\right)}$*, and similarly for the case where* $\sqrt{\rho_A}$ *is replaced by* $\chi_{A,\alpha}$*.*

If these assumptions hold with $\tilde{\sigma}_+(0) = \sigma_+(0)$ *and* $\tilde{\sigma}_-(0) = \sigma_-(0)$*, then, letting* $\hat{\chi}$ *be the half-length of the fixed-length CI based on* $\hat{L}_\delta$ *and the modulus* $\omega_{\tilde{\sigma}(\cdot),n}$ *(using the formula (23) from Theorem 3.4), we have* $n^{p/(2p+1)} \hat{\chi} \xrightarrow[\mathcal{F}, \mathcal{Q}_n]{p} \mathrm{cv}_\alpha(\overline{bias}_\infty / v_\infty^{1/2}) v_\infty^{1/2}$*, where* $\overline{bias}_\infty$ *and* v_∞ *are* *given in Theorem G.1 for k^+ and k^- given above. If, in addition, for each n, there exists a $Q \in \mathcal{Q}_n$ such that the errors are normally distributed, then we have the following. If δ_n is chosen to maximize $\frac{\omega_{\tilde{\sigma}(\cdot),n}}{\delta}\chi_{A,\alpha}\left(\frac{\delta}{2}\right)$ and either Assumption G.4 or Assumption G.3 holds, 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 $\chi < \text{cv}(\overline{bias}_\infty/v_\infty^{1/2})v_\infty^{1/2}$ (where $\overline{bias}_\infty$ and v_∞ are given in Theorem G.1 for k^+ and k^- given above). In addition, for any sequence of confidence sets $\mathcal{C}$ with $\liminf_{n\to\infty}\inf_{f\in\mathcal{F},Q\in\mathcal{Q}_n} P_{f,Q}\left(Lf\in\mathcal{C}\right) \geq 1-\alpha$, we have the following bound on the asymptotic efficiency improvement at any $f \in \mathcal{F}_{RDT,p}(0)$:

$$\liminf_{n\to\infty}\sup_{Q\in\mathcal{Q}_n} \frac{n^{p/(2p+1)}E_{f,Q}\lambda(\mathcal{C})}{v_\infty^{1/2}\,\text{cv}(\overline{\text{bias}}_\infty/v_\infty^{1/2})} \geq \frac{(1-\alpha)E[(z_{1-\alpha}-Z)^{2p/(2p+1)} \mid Z \leq z_{1-\alpha}]}{(c_\chi^{-1}(2p/(2p+1)))^{2p/(2p+1)-1}\chi_{A,\alpha}(c_\chi^{-1}(2p/(2p+1)))}$$

where $Z \sim N(0,1)$. If $\delta_n = z_\beta + z_{1-\alpha}$, then, letting $[\hat{c}_{\alpha,\delta},\infty)$ be the CI formed with $\hat{L}_\delta$ based on the modulus $\omega_{\tilde{\sigma}(\cdot),n}$ (using the formula (23) from Theorem 3.4), 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

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

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

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

To prove this theorem, we first prove a series of lemmas.

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

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

Under Assumption G.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 G.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 $(\bar{b}_+,\bar{d}_+)$,

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

Since the term in brackets converges to a finite constant by Assumption G.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 G.4,

$$\begin{aligned} &\left|G_n(b_+,b_-,d_+,d_-) - \left[\tilde{G}_n(b_+,d_+)\tilde{\sigma}_+^2(0) + \tilde{G}_n(b_-,d_-)\tilde{\sigma}_-^2(0)\right]\right| \le \\ &\bar{k}\cdot\left[\frac{1}{nh_n}\sum_{i=1}^n I(|x_i/h_n| \le K_1)\right]\left[\sup_{0<x\le K_1h_n}\left|\tilde{\sigma}_+^2(0) - \tilde{\sigma}_+^2(x)\right| + \sup_{-K_1h_n\le x<0}\left|\tilde{\sigma}_-^2(0) - \tilde{\sigma}_-^2(x)\right|\right] \end{aligned}$$

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 G.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 G.3. *Under Assumption G.4,* $\|(\tilde{b}_{+,n},\tilde{b}_{-,n},\tilde{d}_{+,n},\tilde{d}_{-,n})\| \le B$ *for some constant* B *and* n *large enough. Under Assumption G.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 $d = (d_1, \dots, d_{p-1})$. Note $G_n(b_+, b_-, d_+, d_-)$ is bounded from below by $1/\sup_{|x|\le h_n} \tilde{\sigma}^2(x)$ times

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

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

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

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 $\max\{|b_+|,|d_{+,1}|,\dots,|d_{+,p-1}|\} = 1$). Combining this with the previous display and the fact that $\frac{1}{nh}\sum_{i:|x_i|\le h_n} C^2$ converges to a finite constant, it follows that, for some $\eta > 0$, $\inf_{\max\{|b_+|,|d_{+,1}|,\dots,|d_{+,p-1}|\}\ge B} G_n(b_+,b_-,d_+,d_-) \ge (B^2\eta - \eta^{-1})/\sup_{|x|\le h_n}\tilde{\sigma}^2(x)$ for large enough n. Let K be such that $G_\infty(\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty}) \le K/2$ and $\max\{\tilde{\sigma}_+^2(0), \tilde{\sigma}_-^2(0)\} \le K/2$. Under Assumption G.4, $G_n(\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty}) < K$ and $\sup_{|x|\le h_n}\tilde{\sigma}^2(x) \le K$ for large enough n. Under Assumption G.3, $G_n(\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty}) < K$ and $\sup_{|x|\le h_n}\tilde{\sigma}^2(x) \le K$ 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 $G_n(\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty}) \le K$ and $\sup_{|x|\le h_n}\tilde{\sigma}^2(x) \le K$, $(\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty})$ will give a lower value of G_n than any (b_+,b_-,d_+,d_-) with $\max\{|b_+|,|d_{+,1}|,\dots,|d_{+,p-1}|,|b_-|,|d_{-,1}|,\dots,|d_{-,p-1}|\} \ge B$. 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 G.4. *If Assumption G.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 G.3 holds and $\tilde{b}_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \tilde{b}_\infty > 0$, $(\tilde{b}_{+,n}, \tilde{b}_{-,n}, \tilde{d}_{+,n}, \tilde{d}_{-,n}) \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}}$*

$(\tilde{b}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{+,\infty}, \tilde{d}_{-,\infty})$.

Proof. By Lemma G.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 G.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 G.3. The result follows from Lemma G.2, continuity of G_∞ and the fact that G_∞ has a unique minimizer. □

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

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

Lemma G.6. *Under Assumption G.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 G.3, we have, for any $\overline{\delta} > 0$,

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

Proof. The first statement is immediate from Lemma G.5 and Lemma F.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 $\underline{\sigma}(x) = (\sigma_+(0) - \eta)I(x > 0) + (\sigma_-(0) - \eta)I(x < 0)$ 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 G.7. *Let r denote $\sqrt{\rho_A}$ or $\chi_{A,\alpha}$. Under Assumption G.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 G.4 and $\delta_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \delta^*$ under Assumption G.3. In *addition, for any* $0 < \alpha < 1$ *and* Z *a standard normal variable,*

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

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

Lemma G.8. *Under Assumptions G.1 and G.2, the following hold. If Assumption G.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 G.3 holds and $\tilde{b}_n$ *is a random sequence with* $\tilde{b}_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \tilde{b}_\infty > 0$*, then*

$$\sup_x |k^+_{\tilde{\sigma}(\cdot)}(x;\tilde{b}_{+,n},\tilde{d}_{+,n}) - k^+_{\tilde{\sigma}_+(0)}(x;\tilde{b}_{+,\infty},\tilde{d}_{+,\infty})| \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0,$$
$$\sup_x |k^-_{\tilde{\sigma}(\cdot)}(x;\tilde{b}_{-,n},\tilde{d}_{-,n}) - k^-_{\tilde{\sigma}_-(0)}(x;\tilde{b}_{-,\infty},\tilde{d}_{-,\infty})| \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0,$$
$$\frac{1}{\sqrt{nh_n}}\sum_{i=1}^n \left[k^+_{\tilde{\sigma}(\cdot)}(x_i/h_n;\tilde{b}_{+,n},\tilde{d}_{+,n}) - k^+_{\tilde{\sigma}_+(0)}(x_i/h_n;\tilde{b}_{+,\infty},\tilde{d}_{+,\infty})\right] u_i \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0$$

and

$$\frac{1}{\sqrt{nh_n}}\sum_{i=1}^n \left[k^-_{\tilde{\sigma}(\cdot)}(x_i/h_n;\tilde{b}_{-,n},\tilde{d}_{-,n}) - k^-_{\tilde{\sigma}_-(0)}(x_i/h_n;\tilde{b}_{-,\infty},\tilde{d}_{-,\infty})\right] u_i \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0.$$

Proof. For the first four statements, 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 G.4, the first term is, for large enough n, bounded by a constant times

$\sup_{0<x<h_n K} |\tilde{\sigma}^{-2}(x) - \tilde{\sigma}_+^{-2}(0)|$, 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 G.4. The second term converges to zero by Lipschitz continuity of $k^+_{\tilde{\sigma}_+(0)}$. Under Assumption G.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.

For the last two statements, note that, under Assumption G.3,

$$\frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left[k^+_{\tilde{\sigma}(\cdot)}(x_i/h_n; \tilde{b}_{+,n}, \tilde{d}_{+,n}) - k^+_{\tilde{\sigma}_+(0)}(x_i/h_n; \tilde{b}_{+,\infty}, \tilde{d}_{+,\infty}) \right] u_i$$
$$= \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left[k^+_{\tilde{\sigma}_+(0)}(x_i/h_n; \tilde{b}_{+,n}, \tilde{d}_{+,n}) - k^+_{\tilde{\sigma}_+(0)}(x_i/h_n; \tilde{b}_{+,\infty}, \tilde{d}_{+,\infty}) \right] u_i$$
$$+ \frac{\hat{\sigma}_+^{-2} - \tilde{\sigma}_+^{-2}(0)}{\tilde{\sigma}_+^{-2}(0)} \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n k^+_{\tilde{\sigma}_+(0)}(x_i/h_n; \tilde{b}_{+,n}, \tilde{d}_{+,n}) u_i. \quad \text{(S17)}$$

Under Assumptions G.1, G.2 and G.3, $\frac{1}{\sqrt{nh_n}} \sum_{i=1}^n k^+_{\tilde{\sigma}_+(0)}(x_i/h_n; \tilde{b}_{+,\infty}, \tilde{d}_{+,\infty}) u_i$ converges in distribution to a normal law uniformly over $\mathcal{F}, \mathcal{Q}_n$. Thus, assuming that the first term converges in probability to zero uniformly over $\mathcal{F}, \mathcal{Q}_n$, the same convergence to a normal law will hold for $\frac{1}{\sqrt{nh_n}} \sum_{i=1}^n k^+_{\tilde{\sigma}_+(0)}(x_i/h_n; \tilde{b}_{+,n}, \tilde{d}_{+,n}) u_i$ and, from this and the fact that $\hat{\sigma}_+^{-2} - \tilde{\sigma}_+^{-2}(0) \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0$, it will follow that the second term in the display converges in probability to zero in probability uniformly over $\mathcal{F}, \mathcal{Q}_n$. It therefore suffices to show that the first term converges in probability to zero in probability uniformly over $\mathcal{F}, \mathcal{Q}_n$.

To this end, note that,

$$\frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left[\left(\tilde{b}_{+,n} + \sum_{j=1}^{p-1} \tilde{d}_{+,n} (x_i/h_n)^j - C|x_i/h_n|^p \right)_+ - \left(\tilde{b}_{+,\infty} + \sum_{j=1}^{p-1} \tilde{d}_{+,\infty} (x_i/h_n)^j - C|x_i/h_n|^p \right)_+ \right]$$
$$= \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left(\tilde{b}_{+,n} - \tilde{b}_{+,\infty} + \sum_{j=1}^{p-1} \left(\tilde{d}_{+,j,n} - \tilde{d}_{+,j,\infty} \right) (x_i/h_n)^j \right) I\left(x_i/h_n \in A(\tilde{b}_{+,\infty}, \tilde{d}_{+,\infty}) \right) u_i$$
$$+ \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left(\tilde{b}_{+,n} + \sum_{j=1}^{p-1} \tilde{d}_{+,j,n} (x_i/h_n)^j - C|x_i/h_n|^p \right)$$
$$\cdot \left[I(x_i/h_n \in A(\tilde{b}_{+,n}, \tilde{d}_{+,n})) - I(x_i/h_n \in A(\tilde{b}_{+,\infty}, \tilde{d}_{+,\infty})) \right] u_i \quad \text{(S18)}$$

where $A(b_+, d_+) = \left\{u\colon b_+ + \sum_{j=1}^{p-1} d_{+,j} u^j - C|u|^p \geq 0\right\}$. The first term can be written as

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

which converges in probability to zero uniformly over $\mathcal{F}, \mathcal{Q}_n$ since $\tilde{b}_{+,n} - \tilde{b}_{+,\infty} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0$, $\tilde{d}_{+,j,n} - \tilde{d}_{+,j,\infty} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0$ and $\frac{1}{\sqrt{nh_n}} \sum_{i=1}^{n} (x_i/h_n)^j I\left(x_i/h_n \in A(\tilde{b}_{+,\infty}, \tilde{d}_{+,\infty})\right) u_i$ converges in distribution to a normal law uniformly over $\mathcal{F}, \mathcal{Q}_n$ for $j = 0, 1, \ldots, p-1$.

For the second term in (S18), it suffices to show that

$$\frac{1}{\sqrt{nh_n}} \sum_{i=1}^{n} (x_i/h_n)^j \left[I(x_i/h_n \in A(\tilde{b}_{+,n}, \tilde{d}_{+,n})) - I(x_i/h_n \in A(\tilde{b}_{+,\infty}, \tilde{d}_{+,\infty}))\right] u_i \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0,$$

since the second term in (S18) can be written as a sum of terms of this form times terms that converge in probability to finite constants uniformly over $\mathcal{F}, \mathcal{Q}_n$. To this end, note that, letting $u_1^*, \ldots, u_k^*$ be the positive zeros of the polynomial $\tilde{b}_{+,\infty} + \sum_{j=1}^{p-1} \tilde{d}_{+,j,\infty} u^j + C u^p$, the following statement will hold with probability approaching one uniformly over $\mathcal{F}, \mathcal{Q}_n$ for any $\eta > 0$: for all u with $I(u \in A(\tilde{b}_{+,n}, \tilde{d}_{+,n})) - I(u \in A(\tilde{b}_{+,\infty}, \tilde{d}_{+,\infty})) \neq 0$, 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

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

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

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

which can be made arbitrarily small by making η small.

This shows that (S18) converges in probability to zero uniformly over $\mathcal{F}, \mathcal{Q}_n$. A similar argument applies to the rest of the terms that make up the first term in (S17), which, along with similar arguments for $k^-_{\tilde{\sigma}(\cdot)}$, gives the result. $\square$

We are now ready to prove Theorem G.3. If $\delta_n \to \delta^*$, then $n^{p/(2p+1)}\omega_n(\delta) \to \omega_\infty(\delta^*)$ by Lemma G.6, so, applying Lemma G.8 with $2\tilde{b}_n = n^{p/(2p+1)}\omega_n(\delta)$, it follows that Assumption G.5 holds with the given choices of k^+ and k^- as required. Similarly, if δ_n is random and $\delta_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \delta^*$, we have $n^{p/(2p+1)}\omega_n(\delta) \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \omega_\infty(\delta^*)$ by Lemma G.6, so the version of Lemma G.8 with random $\tilde{b}_n$ applies to verify Assumption G.6. The statement that $\delta_n \to \delta^*$ under Assumption G.4 and $\delta_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \delta^*$ under Assumption G.3 when δ_n maximizes $(\omega_{\tilde{\sigma}(\cdot),n}(\delta)/\delta)\sqrt{\rho_A(\delta/2)}$ or $(\omega_{\tilde{\sigma}(\cdot),n}(\delta)/\delta)\chi_{A,\alpha}(\delta/2)$ and δ^* maximizes $(\omega_\infty(\delta)/\delta)\sqrt{\rho_A(\delta/2)}$ or $(\omega_\infty(\delta)/\delta)\chi_{A,\alpha}(\delta/2)$ follows from Lemma G.7. Next, the statement that $n^{p/(2p+1)}\hat{\chi} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \text{cv}(\overline{\text{bias}}_\infty/v_\infty^{1/2})v_\infty^{1/2}$ where $\hat{\chi}$ is the half-length of the two-sided CI formed using $\omega_{\tilde{\sigma}(\cdot),n}$ follows from Theorem G.1, since $\hat{\chi} = \text{cv}(\overline{\text{bias}}_n/\tilde{v}_n^{1/2})\tilde{v}_n^{1/2}$ where $\overline{\text{bias}}_n$ and $\tilde{v}_n^{1/2}$ are defined as in Section G.1 for $k_n^+(u) = k^+_{\tilde{\sigma}(\cdot)}(u;\tilde{b}_n,\tilde{d}_n)$ and $k_n^-(u) = k^-_{\tilde{\sigma}(\cdot)}(u;\tilde{b}_n,\tilde{d}_n)$ with $2\tilde{b}_n = \omega_{\tilde{\sigma}(\cdot),n}(\delta)$.

We now prove the optimality statements (under which the assumption was made that $\tilde{\sigma}_+(0) = \sigma_+(0)$ and $\tilde{\sigma}_-(0) = \sigma_-(0)$ and, 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

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

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 Theorem 3.2 (using the characterization of optimal half-length at the beginning of Supplemental Appendix F). This converges to $\sup_{\delta>0} \frac{\omega_{\infty,\sigma_+(0),\sigma_-(0)}(\delta)}{\delta}\chi_{A,\alpha+\eta}(\delta/2)$ by Lemma G.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 $\chi \geq \lim_{\eta\downarrow 0}\sup_{\delta>0} \frac{\omega_{\infty,\sigma_+(0),\sigma_-(0)}(\delta)}{\delta}\chi_{A,\alpha+\eta}(\delta/2) = \sup_{\delta>0} \frac{\omega_{\infty,\sigma_+(0),\sigma_-(0)}(\delta)}{\delta}\chi_{A,\alpha}(\delta/2)$ (the limit with respect to η follows since there exist $0 < \underline{\delta} < \overline{\delta} < \infty$ such that the supremum over δ is taken $[\underline{\delta},\overline{\delta}]$ 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). Since $\text{cv}_\alpha(\overline{\text{bias}}_n/\tilde{v}_n^{1/2})\tilde{v}_n^{1/2} = \sup_{\delta>0} \frac{n^{p/(2p+1)}\omega_{\tilde{\sigma}(\cdot),n}(\delta)}{\delta}\chi_{A,\alpha}(\delta)$ under Assumption G.4 with δ optimizing the fixed-length CI, it follows from Lemma G.7

and the first part of the theorem shown above that

$$\mathrm{cv}_\alpha(\overline{\mathrm{bias}}_\infty/v_\infty^{1/2})v_\infty^{1/2} = \lim_{n\to\infty}\mathrm{cv}_\alpha(\overline{\mathrm{bias}}_n/\tilde{v}_n^{1/2})\tilde{v}_n^{1/2} = \lim_{n\to\infty}\sup_{\delta>0}\frac{n^{p/(2p+1)}\omega_{\tilde{\sigma}(\cdot),n}(\delta)}{\delta}\chi_{A,\alpha}(\delta/2)$$
$$= \sup_{\delta>0}\frac{\omega_{\infty,\sigma_+(0),\sigma_-(0)}(\delta)}{\delta}\chi_{A,\alpha}(\delta/2)$$

where the first two equalities use the definition of $\tilde{v}_n$ under Assumption G.4. This is the lower bound for χ given above, which proves the asymptotic efficiency bound for fixed-length affine confidence intervals.

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

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

by Corollary 3.3. This converges to

$$(1-\alpha-\eta)E\left[\omega_{\infty,\sigma_+(0),\sigma_-(0)}(2(z_{1-\alpha-\eta}-Z))|Z \le z_{1-\alpha-\eta}\right]$$

by Lemma G.7. The result follows from taking $\eta \to 0$ and using the dominated convergence theorem, and using the fact that $\omega_{\infty,\sigma_+(0),\sigma_-(0)}(\delta) = \omega_{\infty,\sigma_+(0),\sigma_-(0)}(1)\delta^{2p/(2p+1)}$ along with calculations similar to those given after Corollary 3.3. 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 G.1 and Lemma F.3.

Appendix H Supplemental Figures

Figure S1 plots a point estimator and one- and two-sided CIs based on a local linear estimator and triangular kernel and bandwidths chosen to be minimax optimal for the given performance criterion. Figure S2 plots plots one- and two-sided CIs based on a local linear estimator with bandwidths that minimize the maximum MSE.

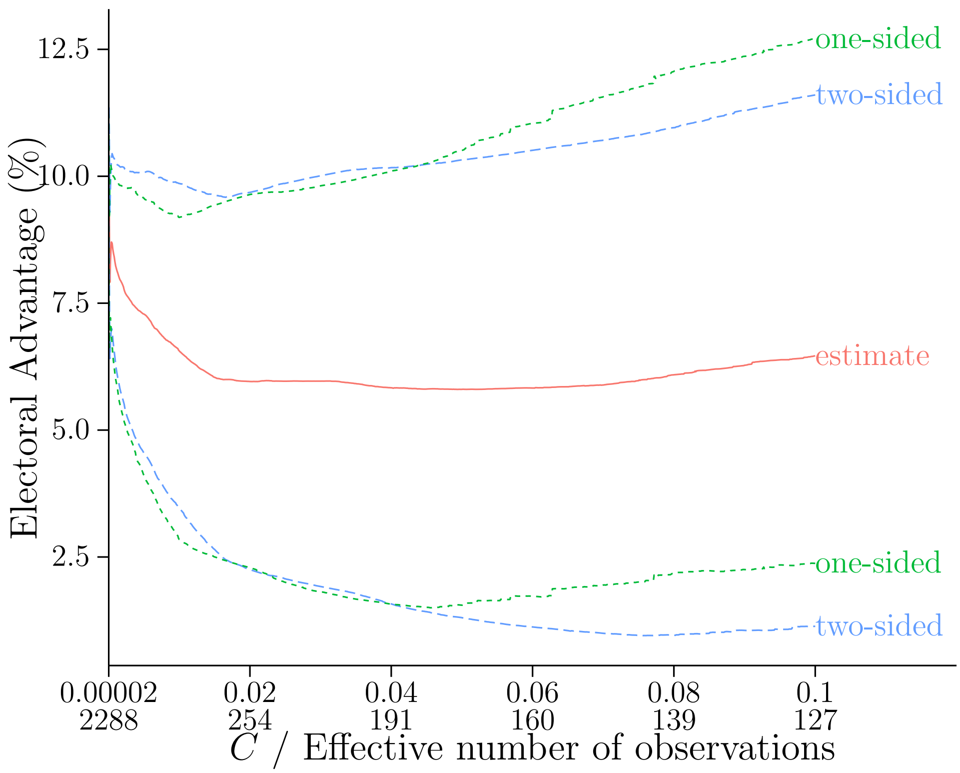

Figure S1: Lee (2008) RD example: local linear regression with triangular kernel. Estimator based on minimax MSE bandwidths (estimator), lower and upper limits of one-sided confidence intervals with bandwidths that are minimax for 0.8 quantile of excess length (one-sided), and shortest fixed-length CIs (two-sided) as function of smoothness C. Effective number of observations corresponds to n_e for the minimax MSE estimator as defined in Equation (27) in the text.

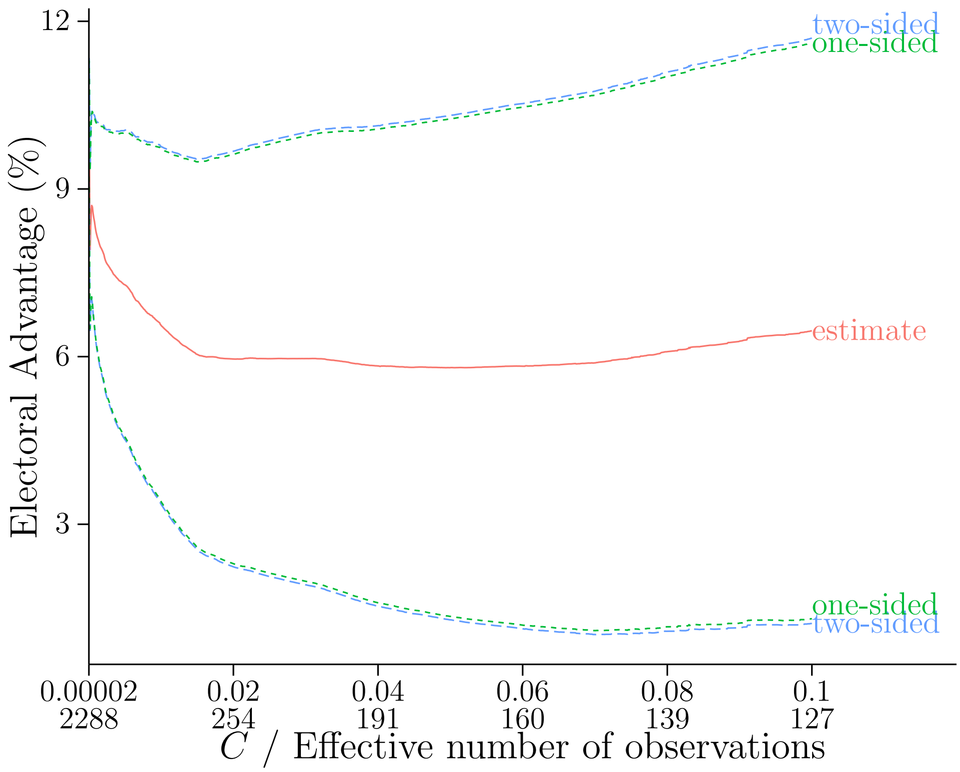

Figure S2: Lee (2008) RD example: local linear regression with triangular kernel and minimax MSE bandwidths (estimator) with two-sided CI (two-sided) as well as lower and upper limits of one-sided CIs around it as function of smoothness C. Effective number of observations corresponds to n_e for the minimax MSE estimator as defined in Equation (27) in the text.

References

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(1), 279–287.

Donoho, D. L. (1994): "Statistical Estimation and Optimal Recovery," *The Annals of Statistics*, 22(1), 238–270.

Donoho, D. L., and M. G. Low (1992): "Renormalization Exponents and Optimal Pointwise Rates of Convergence," *The Annals of Statistics*, 20(2), 944–970.

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

Fan, J. (1993): "Local Linear Regression Smoothers and Their Minimax Efficiencies," *The Annals of Statistics*, 21(1), 196–216.

Lee, D. S. (2008): "Randomized experiments from non-random selection in U.S. House elections," *Journal of Econometrics*, 142(2), 675–697.

Romano, J. P., and M. Wolf (2015): "Resurrecting Weighted Least Squares," SSRN Scholarly Paper ID 2491081, Social Science Research Network, Rochester, NY.

Sacks, J., and D. Ylvisaker (1978): "Linear Estimation for Approximately Linear Models," *The Annals of Statistics*, 6(5), 1122–1137.

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

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(1), 299–303.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data.* MIT Press, Cambridge, MA, second edn.