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

Timothy B. Armstrong* Yale University

Michal Kolesár† Princeton University

May 22, 2017

These supplemental materials provide appendices not included in the main text. 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 D.1.

Appendix E Unknown Error Distribution

The Gaussian regression model (1) 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:

$$y_i = f(x_i) + u_i,\ \{u_i\}_{i=1}^n \sim Q,\ f \in \mathcal{F},\ Q \in \mathcal{Q}_n \tag{S1}$$

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

*email: timothy.armstrong@yale.edu

†email: mkolesar@princeton.edu

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

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

$$\hat{L} = a_n + \sum_{i=1}^n w_{i,n} y_i \tag{S2}$$

where a_n and $w_{i,n}$ are a sequence and triangular array respectively. For now, we assume that a_n and $w_{i,n}$ are nonrandom, (which, in the case of the estimator $\hat{L}_{\delta,\tilde{\sigma}(\cdot)}$, requires that $\tilde{\sigma}(\cdot)$ and δ be nonrandom). We provide conditions that allow for random a_n and $w_{i,n}$ after stating our result for nonrandom weights. For a class $\mathcal{G}$, the maximum and minimum bias are

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

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

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

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

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

$$\frac{\sum_{i=1}^{n} w_{i,n} u_i}{s_{n,Q}} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{d}{\to}} \mathcal{N}(0,1) \tag{S3}$$

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

$$\frac{\widehat{\text{se}}_n}{s_{n,Q}} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 1. \tag{S4}$$

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

Theorem E.1. *Let $\hat{L}$ be an estimator of the form (S2), and suppose that (S3) and (S4) hold. Let $\hat{c} = \hat{L} - \overline{bias}_{\mathcal{F}}(\hat{L}) - \widehat{se}_n z_{1-\alpha}$, and let $b = \max\{|\overline{bias}_{\mathcal{F}}(\hat{L})|, |\underline{bias}_{\mathcal{F}}(\hat{L})|\}$. Then*

$$\liminf_{n\to\infty} \inf_{f\in\mathcal{F},Q\in\mathcal{Q}_n} P_{f,Q}\left(Lf \in [\hat{c},\infty)\right) \geq 1-\alpha \tag{S5}$$

and

$$\liminf_{n\to\infty} \inf_{f\in\mathcal{F},Q\in\mathcal{Q}_n} P_{f,Q}\left(Lf \in \left\{\hat{L} \pm \widehat{se}_n \operatorname{cv}_\alpha\left(b/\widehat{se}_n\right)\right\}\right) \geq 1-\alpha. \tag{S6}$$

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

$$\limsup_{n\to\infty} \sup_{Q\in\mathcal{Q}_n} \frac{\sup_{g\in\mathcal{G}} q_{g,Q,\beta}(Lg-\hat{c})}{\overline{bias}_{\mathcal{F}}(\hat{L}) - \underline{bias}_{\mathcal{G}}(\hat{L}) + s_{n,Q}(z_{1-\alpha}+z_\beta)} \leq 1 \tag{S7}$$

and the length of the two-sided CI will satisfy

$$\frac{\operatorname{cv}_\alpha\left(b/\widehat{se}_n\right)\widehat{se}_n}{\operatorname{cv}_\alpha\left(b/s_{n,Q}\right)s_{n,Q}} \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 1.$$

Suppose, in addition, that $\hat{L} = \hat{L}_{\delta,\mathcal{F},\mathcal{G},\tilde{\sigma}(\cdot)}$ with $\tilde{\sigma}(\cdot) = \sigma_Q(\cdot)$ where $\sigma_Q^2(x_i) = var_Q(u_i)$ and, for each n, there exists a $Q_n \in \mathcal{Q}_n$ such that $\{u_i\}_{i=1}^n$ are independent and normal under Q_n. Then then no one-sided CI satisfying (S5) can satisfy (S7) with the constant 1 replaced by a strictly smaller constant on the right hand side.

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

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

using the fact that $\overline{\text{bias}}_{\mathcal{F}}(\hat{L}) + \sum_{i=1}^{n} w_{i,n}u_i \geq \hat{L} - Lf$ for all $f \in \mathcal{F}$ by the definition of $\overline{\text{bias}}_{\mathcal{F}}$. The right hand side converges to $1-\alpha$ uniformly over $f \in \mathcal{F}$ and $Q \in \mathcal{Q}_n$ by (S3) and (S4). For the two-sided CI, first note that

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

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

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

where $c_n = \frac{\text{cv}_\alpha(b/\widehat{\text{se}}_n)\widehat{\text{se}}_n}{\text{cv}_\alpha\left(b/s_{n,Q}\right)s_{n,Q}}$, $\tilde{Z}_n = \sum_{i=1}^{n} w_{i,n}u_i/s_{n,Q}$ and $r = \left(a_n + \sum_{i=1}^{n} w_{i,n}f(x_i) - Lf\right)/s_{n,Q}$. By (S3) and the fact that $c_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 1$, this is equal to $P_{f,Q}\left(|Z+r| \leq \text{cv}_\alpha\left(b/s_{n,Q}\right)\right)$ (where $Z \sim \mathcal{N}(0,1)$) plus a term that converges to zero uniformly over f, Q (this can be seen by using the fact that convergence in distribution to a continuous distribution implies uniform convergence of the cdfs; see Lemma 2.11 in van der Vaart 1998). Since $|r| \leq b/s_{n,Q}$, (S6) follows.

To show (S7), note that,

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

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

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

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

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

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

where $\tilde{\delta}_n = z_{1-\alpha_n} + z_\beta$ and $1-\alpha_n$ is the coverage of $[\tilde{c}, \infty)$ over $\mathcal{F}$, $\mathcal{Q}_n$. When $\hat{L} = \hat{L}_{\delta,\mathcal{F},\mathcal{G},\sigma_Q(\cdot)}$, the denominator in (S7) for $Q = Q_n$ is equal to $\omega_{\sigma_{Q_n}(\cdot),n}(z_{1-\alpha} + z_\beta)$, which gives

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

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

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

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

Theorem E.2. *Suppose that* $\hat{L}$ *and* $\widehat{se}_n$ *satisfy* (S3) *and* (S4). *Let* $\tilde{L}$ *and* $\widetilde{se}_n$ *be another estimator and standard error, and let* $\widetilde{\overline{bias}}_n$ *and* $\widetilde{\underline{bias}}_n$ *be (possibly data dependent) worst-case bias formulas for* $\tilde{L}$ *under* $\mathcal{F}$. *Suppose that*

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

Let $\tilde{c} = \tilde{L} - \widetilde{\overline{bias}}_n - \widetilde{se}_n z_{1-\alpha}$, *and let* $\tilde{b} = \max\{|\widetilde{\overline{bias}}_n|, |\widetilde{\underline{bias}}_n|\}$. *Then (S5) and (S6) hold with* $\hat{c}$ *replaced by* $\tilde{c}$, $\hat{L}$ *replaced by* $\tilde{L}$, b *replaced by* $\tilde{b}$ *and* $\widehat{se}_n$ *replaced by* $\widetilde{se}_n$. *Furthermore, the performance of the CIs is asymptotically equivalent in the sense that*

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

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

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

Thus,

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

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

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

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

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

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

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

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

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

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

We consider the estimate

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

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

$$\sup_n \sup_{x,x' :\, d_X(x,x') \leq \eta} \max\left\{\left|f^*_{\tilde{\sigma}(\cdot),\delta}(x) - f^*_{\tilde{\sigma}(\cdot),\delta}(x')\right|, \left|g^*_{\tilde{\sigma}(\cdot),\delta}(x) - g^*_{\tilde{\sigma}(\cdot),\delta}(x')\right|\right\} \leq \overline{g}(\eta), \tag{S8}$$

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

$$\min_{1\leq i\leq n}\sum_{j=1}^{n} I\left(d_X(x_j,x_i)\leq \eta\right)\to\infty. \tag{S9}$$

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

$$\max_{1\leq i\leq n}|\hat{f}(x_i)-f(x_i)| \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} 0. \tag{S10}$$

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

$$1/K\leq\sigma_Q^2(x_i)\leq K \text{ and } E_Q|u_i|^{2+\eta}\leq K. \tag{S11}$$

In cases where function class $\mathcal{F}$ imposes smoothness on f, (S8) will often follow directly from the definition of $\mathcal{F}$. For example, it holds for the Lipschitz class $\{f\colon |f(x)-f(x')|\leq Cd_X(x,x')\}$. The condition (S9) 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 (S10) will hold under regularity conditions for a variety of choices of $\hat{f}$. It is worth noting that smoothness assumptions on $\mathcal{F}$ needed for this assumption are typically weaker than those needed for asymptotic equivalence with Gaussian white noise. For example, if $\mathcal{X}=\mathbb{R}^k$ with the Euclidean norm, (S8) will hold automatically for Hölder classes with exponent less than or equal to 1, while equivalence with Gaussian white noise requires that the exponent be greater than $k/2$ (see Brown and Zhang, 1998). Furthermore, we do not require any explicit characterization of the limiting form of the optimal CI. In particular, we do not require that the weights for the optimal estimator converge to a limiting optimal kernel or efficient influence function.

The condition (S11) is used to verify a Lindeberg condition for the central limit theorem used to obtain (S3), which we do in the next lemma.

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

such that, for all i,

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

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

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

Then

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

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

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

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

Theorem E.3. *Let* $\hat{L}_{\delta,\tilde{\sigma}(\cdot)}$ *and* $\widehat{se}_{\delta,\tilde{\sigma}(\cdot),n}^2$ *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 (S8) and (S9) hold. Suppose the errors satisfy (S11) and are independent over* i *for all* n *and* $Q \in \mathcal{Q}_n$. *Then (S3) holds. If, in addition, the estimator* $\hat{f}$ *satisfies (S10), then (S4) holds with* $\widehat{se}_n$ *given by* $\widehat{se}_{\delta,\tilde{\sigma}(\cdot),n}$.

Proof. Condition (S3) 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 $\max_{1\le i\le n}(g^*_{\tilde\sigma(\cdot),\delta}(x_i) - f^*_{\tilde\sigma(\cdot),\delta}(x_i))^2 \to 0$. To this end, suppose, to the contrary, that there exists some $c > 0$ such that $\max_{1\le i\le n}(g^*_{\tilde\sigma(\cdot),\delta}(x_i) - f^*_{\tilde\sigma(\cdot),\delta}(x_i))^2 > c^2$ infinitely often. Let η be small enough so that $\overline{g}(\eta) \le c/4$. Then, for n such that this holds and k_n achieving this maximum,

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

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

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

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

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

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

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

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

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

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

Corollary E.1. *If, in addition to the conditions in Theorem E.3, $\sigma_Q^2(x) = \sigma^2 \cdot \tilde{\sigma}^2(x)$ for all n and $Q \in \mathcal{Q}_n$, then, letting $\beta = \Phi(\delta/\sigma - z_{1-\alpha})$, no CI satisfying (S5) can satisfy (S7) with the constant 1 replaced by a strictly smaller constant on the right hand side.*

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 D.1 in Supplemental Appendix G.

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

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

The solution is achieved by shrinking Y toward 0, namely $\delta(Y) = c_\rho(\tau)Y$, with $c_\rho(\tau) = \tau^2/(1+\tau^2)$, which gives $\rho_A(\tau) = \tau^2/(1+\tau^2)$. The length of the smallest fixed-length affine $100 \cdot (1-\alpha)\%$ confidence interval is

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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, $0\le\omega_n(\delta)\le\omega_n(1)\max\{\delta,1\}\le(\omega_n(1)+1)\max\{\delta,1\}$ 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 D.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 D.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 D.4, 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* $\frac{1}{nh_n}\sum_{i=1}^n m(x_i/h_n)1(x_i > 0) \to p_{X,+}(0)\int_0^\infty m(u)\,du$ *and* $\frac{1}{nh_n}\sum_{i=1}^n m(x_i/h_n)1(x_i < 0) \to p_{X,-}(0)\int_{-\infty}^0 m(u)\,du$ *for any bounded function* m *with bounded support and any* h_n *with* $0 < \liminf_n h_n n^{1/(2p+1)} \leq \limsup_n h_n n^{1/(2p+1)} < \infty$*.*

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

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

Assumption 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 D.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

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

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

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

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

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

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

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

Let

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

and

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

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

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

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

and

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

Lemma 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.3 or Assumption G.4 with* $\tilde{\sigma}_+(0) = \sigma_+(0)$ *and* $\tilde{\sigma}_-(0) = \sigma_-(0)$*, then* $n^{2p/(2p+1)}\tilde{v}_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} v_\infty$ *under Assumption G.3 and* $\lim_{n\to\infty} n^{2p/(2p+1)}\tilde{v}_n = v_\infty$ *under Assumption G.4.*

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

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

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

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

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

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

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

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

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

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

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

Proof. We verify the conditions of Theorem E.1. Condition (S4) follows from Lemma G.1. To verify (S3), note that $\hat{L}$ takes the general form in Theorem E.1 with $w_{n,i}$ given by $w_{n,i} = k_n^+(x_i/h_n)/\sum_{j=1}^n k_n^+(x_j/h_n)1(x_j > 0)$ for $x_i > 0$ and $w_{n,i} = k_n^-(x_i/h_n)/\sum_{j=1}^n k_n^-(x_j/h_n) \cdot 1(x_j < 0)$ for $x_i < 0$. The uniform central limit theorem in (S3) with $w_{n,i}$ taking this form follows from Lemma E.1. This gives the asymptotic coverage statements.

For the asymptotic formulas for excess length of the one-sided CI and length of the two-sided CI, we apply Theorem E.2 with $n^{-p/(2p+1)}\overline{\text{bias}}_\infty$ playing the role of $\widetilde{\text{bias}}_n$ and

$n^{-p/(2p+1)}v_\infty$ playing the role of $\widetilde{\text{se}}_n$. Finally, the last statement of the theorem is immediate from Theorem E.2. □

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

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

where $e_1 = (1, 0, \dots, 0)'$ and $p(x) = (1, x, x^2, \dots, x^{p-1})'$. 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 $h_n = h_{n,+}$,

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

and

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

Let M^+ be the $(p-1)\times(p-1)$ matrix with $\int_0^\infty u^{j+k-2} k_+^*(u)$ as the i,jth entry, and let M^- be the $(p-1)\times(p-1)$ matrix with $\int_{-\infty}^0 u^{j+k-2} k_-^*(u)$ as the i,jth entry. Under Assumption G.1, for k_+^* and k_-^* bounded with bounded support, $\frac{1}{nh_n}\sum_{i=1}^n p(x_i/h_{+,n}) p(x_i/h_{+,n})' k_+^*(x_i/h_{+,n}) \cdot 1(x_i > 0) \to M^+ p_{X,+}(0)$ and similarly $\frac{1}{nh_n}\sum_{i=1}^n p(x_i/h_{-,n}) p(x_i/h_{-,n})' k_+^*(x_i/h_{-,n}) \cdot 1(x_i < 0) \to M^- p_{X,-}(0)$. Furthermore, Assumption 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 D, 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 (38) in Section D.2 with $\tilde{\sigma}(\cdot)$ in place of $\sigma(\cdot)$. We use $\omega_{\tilde{\sigma}(\cdot),n}(\delta)$ to denote the modulus, or $\omega_n(\delta)$ when the context is clear. The corresponding estimator $\hat{L}_{\delta,\tilde{\sigma}(\cdot)}$ is then given by Equation (45) in Section D.2 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 (44) in Section D.2, with $\tilde{\sigma}^2(x_i)$ in place of $\sigma^2(x_i)$, and the solution takes the form given in that section. Let $h_n = n^{-1/(2p+1)}$. We will consider a sequence $b = b_n$, and will define $\tilde{b}_n = n^{p/(2p+1)}b_n = h_n^{-p}b_n$. Under Assumption 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 $j = 1, \ldots, p-1$, 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.2. These values of $\tilde{b}_{+,n}$, $\tilde{b}_{-,n}$, $\tilde{d}_{+,n}$ and $\tilde{d}_{-,n}$ minimize $G_n(b_+, b_-, d_+, d_-)$ subject to $b_+ + b_- = \tilde{b}_n$ where, letting $\mathcal{A}(x_i, b, d) = h_n^p b + \sum_{j=1}^{p-1} h_n^{p-j} d_j x_i^j$,

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

with

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

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

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

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

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

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

We proceed by verifying the conditions of Theorem G.1 for the case where $\tilde{\sigma}(\cdot)$ is nonrandom and satisfies Assumption G.4, and then using Theorem E.2 for the case where $\tilde{\sigma}(\cdot)$ satisfies Assumption G.3. The limiting kernel k^+ and k^- in Assumption G.5 will correspond to an asymptotic version of the modulus problem, which we now describe. Let

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

Consider the limiting inverse modulus problem

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

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

To derive the form of the limiting modulus of continuity, we argue as in Donoho and Low (1992). Let $k_1^+(\cdot; \tilde b_{+,\infty,1}, \tilde d_{+,\infty,1})$ and $k_1^+(\cdot; \tilde b_{+,\infty,1}, \tilde d_{+,\infty,1})$ solve the inverse modulus problem $\omega_\infty^{-1}(2\tilde b_\infty)$ for $\tilde b_\infty = 1$. The feasible set for a given $\tilde b_\infty$ consists of all b_+, b_-, d_+, d_- such that $b_+ + b_- \geq \tilde b_\infty$, and a given b_+, b_-, d_+, d_- in this set achieves the value

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

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 $\omega_\infty(\delta) = \omega_{\tilde{\sigma}_+(0), \tilde{\sigma}_-(0), \infty}(\delta) = \delta^{2p/(2p+1)} \omega_{\tilde{\sigma}_+(0), \tilde{\sigma}_-(0), \infty}(1)$.

If $\tilde{b}_\infty = \omega_\infty(\delta_\infty)$ for some δ_∞, then it can be checked that the limiting variance and worst-case bias defined in Section G.1 correspond to the limiting modulus problem:

$$\overline{\text{bias}}_\infty = \frac{1}{2}\left(\omega_\infty(\delta_\infty) - \delta_\infty \omega_\infty'(\delta_\infty)\right), \quad \sqrt{v_\infty} = \omega_\infty'(\delta_\infty). \tag{S12}$$

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

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

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

The CIs $[\hat{c}_{\alpha, \delta_n}, \infty)$ *and* $[\hat{L} - \hat{\chi}, \hat{L} + \hat{\chi}]$ *have uniform asymptotic coverage at least* $1 - \alpha$. *In addition,* $n^{p/(2p+1)} \hat{\chi} \xrightarrow[\mathcal{F}, \mathcal{Q}_n]{p} \chi_\infty$ *where* $\chi_\infty = \text{cv}_\alpha(\overline{bias}_\infty / \sqrt{v_\infty})\sqrt{v_\infty}$ *with* $\overline{bias}_\infty$ *and* $\sqrt{v_\infty}$ *given in* (S12) *and* $\delta_\infty = z_\beta + z_{1-\alpha}$ *if excess length is the criterion,* $\delta_\infty = \arg\min_\delta \text{cv}_\alpha(\frac{\omega_\infty(\delta)}{2\omega'_\infty(\delta)} - \frac{\delta}{2})\omega'_\infty(\delta)$ *if FLCI length is the criterion, and* $\delta_\infty = \arg\min_\delta [\frac{1}{4}\left(\omega_\infty(\delta_\infty) - \delta_\infty \omega'_\infty(\delta_\infty)\right)^2 + \omega'_\infty(\delta)^2]$ *if RMSE is the criterion.*

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

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

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

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

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

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

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

must satisfy

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

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

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

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

Lemma 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,

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

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 $(\bar{b}_+,\bar{d}_+)$,

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

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

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

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 $d = (d_1, \ldots, d_{p-1})$. 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 $B > 0$

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

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 $\max\{|b_+|, |d_{+,1}|, \ldots, |d_{+,p-1}|\} = 1$). 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 $\eta > 0$, $\inf_{\max\{|b_+|,|d_{+,1}|,\ldots,|d_{+,p-1}|\}\ge B} G_n(b_+, b_-, d_+, d_-) \ge (B^2\eta - \eta^{-1})/\sup_{|x|\le h_n} \tilde{\sigma}^2(x)$ 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 $\max\{|b_+|, |d_{+,1}|, \ldots, |d_{+,p-1}|, |b_-|, |d_{-,1}|, \ldots, |d_{-,p-1}|\} \ge B$. 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 $\tilde{b}_n \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p} \tilde{b}_\infty > 0$, $(\tilde{b}_{+,n}, \tilde{b}_{-,n}, \tilde{d}_{+,n}, \tilde{d}_{-,n}) \xrightarrow[\mathcal{F},\mathcal{Q}_n]{p}$*

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

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

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

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

Lemma 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*

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

Proof. Note that

$$\begin{aligned}|k^+_{\tilde{\sigma}(\cdot)}(x;\tilde{b}_{+,n},\tilde{d}_{+,n}) - k^+_{\tilde{\sigma}_+(0)}(x;\tilde{b}_{+,\infty},\tilde{d}_{+,\infty})| &\le |k^+_{\tilde{\sigma}(\cdot)}(x;\tilde{b}_{+,n},\tilde{d}_{+,n}) - k^+_{\tilde{\sigma}_+(0)}(x;\tilde{b}_{+,n},\tilde{d}_{+,n})| \\ &+ |k^+_{\tilde{\sigma}_+(0)}(x;\tilde{b}_{+,n},\tilde{d}_{+,n}) - k^+_{\tilde{\sigma}_+(0)}(x;\tilde{b}_{+,\infty},\tilde{d}_{+,\infty})|.\end{aligned}$$

Under Assumption G.4, the first term is, for large enough n, bounded by a constant times $\sup_{0<x<h_nK} |\tilde{\sigma}^{-2}(x) - \tilde{\sigma}_+^{-2}(0)|$, where K is bound on the support of $k_1^+(\cdot;b_+,d_+)$ over b_+, d_+ in a neighborhood of $\tilde{b}_{+,\infty}, \tilde{d}_{+,\infty}$. This converges to zero by Assumption 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. □

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

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

Proof. We have

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

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

We have

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

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

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

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

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

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

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

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

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

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

Thus,

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

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

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

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

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

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

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

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

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

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

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

which can be made arbitrarily small by making η small.

For the bias formulas, we have

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

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

For the variance formulas, we have

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

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

We are now ready to prove Theorem G.3. First, consider the case with $\tilde{\sigma}(\cdot)$ is deterministic and Assumption G.4 holding. By Lemma G.7, $\delta_n \to \delta_\infty$. By Lemma G.6, it then follows that, under Assumption G.4, $n^{p/(2p+1)}w_n(\delta_n) \to \omega_\infty(\delta_\infty)$ so that Lemma G.8 applies to show that

Assumption G.5 holds with $k^+(x) = k^+_{\tilde{\sigma}_+(0)}(x; \tilde{b}_{+,\infty}, \tilde{d}_{+,\infty})$ and $k^-(x) = k^-_{\tilde{\sigma}_-(0)}(x; \tilde{b}_{-,\infty}, \tilde{d}_{-,\infty})$, where $(\tilde{b}_{+,\infty}, \tilde{d}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{-,\infty})$ minimize $G_\infty(\tilde{b}_{+,\infty}, \tilde{d}_{+,\infty}, \tilde{b}_{-,\infty}, \tilde{d}_{-,\infty})$ subject to $\tilde{b}_{+,\infty} + \tilde{b}_{-,\infty} = \omega_\infty(\delta_\infty)/2$. The coverage statements and convergence of $n^{p/(2p+1)}\hat{\chi}$ then follow from Theorem G.1 and by calculating $\overline{\text{bias}}_\infty$ and v_∞ in terms of the limiting modulus.

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

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

we must have $\chi \geq \sup_{\delta > 0} \frac{n^{p/(2p+1)}\omega_{\sigma(\cdot),n}(\delta)}{\delta} \chi_{A,\alpha+\eta}(\delta/2)$ by the results of Donoho (1994) (using the characterization of optimal half-length at the beginning of Supplemental Appendix F). This converges to $\sup_{\delta>0} \frac{\omega_\infty(\delta)}{\delta} \chi_{A,\alpha+\eta}(\delta/2)$ 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 $\chi \geq \lim_{\eta \downarrow 0} \sup_{\delta > 0} \frac{\omega_\infty(\delta)}{\delta} \chi_{A,\alpha+\eta}(\delta/2) = \sup_{\delta>0} \frac{\omega_\infty(\delta)}{\delta} \chi_{A,\alpha}(\delta/2)$ (the limit with respect to η follows since there exist $0 < \underline{\delta} < \overline{\delta} < \infty$ such that the supremum over δ is taken $[\underline{\delta}, \overline{\delta}]$ for η in a neighborhood of zero, and since $\chi_{A,\alpha}(\delta/2)$ is continuous with respect to α uniformly over δ in compact sets).

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

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

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

In the case where Assumption G.3 holds rather than Assumption G.4, it follows from Lemma G.7 that $\delta_n \underset{\mathcal{F},\mathcal{Q}_n}{\overset{p}{\to}} \delta_\infty$. Then, by Lemma G.9, the conditions of Theorem E.2 hold with $\hat{L}_{\delta_n,\tilde{\sigma}(\cdot)}$ playing the role of $\hat{L}^*$ and $\hat{L}_{\delta_n,\sigma(\cdot)}$ playing the role of $\hat{L}$. The results then follow from Theorem E.2 and the arguments above applied to the CIs based on $\hat{L}_{\delta_n,\sigma(\cdot)}$.

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, 279–287.

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

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

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

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

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

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*, Cambridge, UK ; New York, NY, USA: Cambridge University Press.

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