C Additional results for asymptotic confidence bands

C.1 Pointwise confidence bands based on central limit theory

In standard error-free nonparametric curve estimation problems, a common method for constructing confidence bands is through the limiting distribution of estimators (see e.g. Härdle, 1989a, Eubank and Speckman, 1993, and Xia, 1998). In theory, such procedures could be employed in the errors-in-variables setting too, as we show now.

First, as in the error-free case, confidence bands for g are easier to construct from the limiting distribution obtained when undersmoothing $\hat{g}$, i.e. when taking the bandwidth h smaller than the optimal size for estimating g. Without undersmoothing, constructing the bands in practice requires estimation of a complex bias term, which cannot be done without introducing a non negligible coverage error; see Section C.2 of the Supplementary Material for details. If we undersmooth $\hat{g}$, that is, take $h^2 = o\{(nh)^{-1} \int K_U^2\} = o(1)$, then it can be proved that

$$\hat{g}(x) - g(x) = Z_n(x) \sqrt{V_n(x)}, \tag{C.1}$$

where $Z_n(x)$ asymptotically has the $\mathrm{N}(0,1)$ distribution and $V_n(x) = (nh)^{-1}\nu^2(x) f_X^{-2}(x) \int K_U^2$, with $\nu^2(x) = \sigma^2 f_W(x) + \int_{-\infty}^{\infty} \{g(q) - g(x)\}^2 f_U(x-q)\, f_X(q)\, dq$ and σ^2 denoting the variance of V in the model at (2.1); see Theorem C.1 in Section C of the Supplementary Material. Therefore, provided we can construct consistent estimators $\hat{f}_X$ of f_X and $\hat{\nu}$ of ν, we can estimate $V_n(x)$ by $\hat{V}_n(x) = (nh)^{-1}\, \hat{\nu}^2(x)/\hat{f}_X^{-2}(x) \int K_U^2$, and an asymptotic pointwise confidence band with nominal coverage $1-\alpha$ can take the form:

$$\mathcal{B}_\alpha(\mathcal{I}) = \Big\{(x,y) \,:\, x \in \mathcal{I} \text{ and } \hat{g}(x) - \hat{V}_n^{1/2}(x)\, z_{1-\alpha/2} \le y \le \hat{g}(x) + \hat{V}_n^{1/2}(x)\, z_{1-\alpha/2}\Big\}, \tag{C.2}$$

where $\mathcal{I}$ denotes a compact interval, and $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-level quantile of the $\mathrm{N}(0,1)$ distribution.

In view of (C.1), under modest regularity conditions, including the assumption that f_X is bounded away from zero on $\mathcal{I}$, it can be proved that the pointwise band at (C.2) has approximately correct coverage. We can easily estimate f_X, using for example $\hat{f}_X$ at (2.5). However, unlike the error-free case or density deconvolution problems, there does not seem to be an attractive way of estimating ν consistently. See Section C.3 in the Supplementary Material, where we introduce two possible ways of estimating ν and discuss the difficulties with these approaches. Therefore, while they can be defined, confidence bands based on limiting distributions are not really practicable in the errors-in-variables context.

C.2 First-order properties of $\hat{g}$

Let $R(K_U) = \int K_U^2(v)\,dv$, $\kappa_2 = \int u^2 K(u)\,du$, $b(x) = g''(x) + 2\,f_X'(x)\,g'(x)\,f_X^{-1}(x)$, recall the definition of $s^2(h)$ in (5.3) and of $\nu^2(x)$ in Section C.1, and recall that $E(V) = 0$ and $\operatorname{var} V = \sigma^2$, where σ^2 denotes the variance of V in the model at (2.1). The following theorem is essentially due to Delaigle et al. (2009).

Theorem C.1. *Assume that* (2.2) *and Condition A hold, that f_U, f_X and g are continuous and uniformly bounded, that f_X and g have two continuous derivatives in a neighbourhood of x, that $f_X(x) > 0$, that $\operatorname{var} V < \infty$, that $h = h(n) \to 0$, and that $R(K_U) = o(nh)$. Then, $\hat{g}(x; h) - g(x)$ is asymptotically normally distributed with mean $\frac{1}{2}\,h^2\,\kappa_2\,b(x) + o(h^2)$ and variance $s^2(h)\,\nu^2(x)\,f_X^{-2}(x) + o\{s^2(h)\}$.*

Abusing terminology a little, we refer to the mean and variance of the asymptotic distribution as the asymptotic bias and variance of $\hat{g}$. An asymptotic confidence band with nominal coverage $1-\alpha$ could be defined via this result, replacing b and ν by estimators. However, estimating b is quite complex, and like the error-free case, the resulting asymptotic band would have coverage error that does not vanish asymptotically, and which would need to be corrected, for example by adding to the finite endpoints of the band, a correction factor (which would also need to be estimated). This issue can be avoided by undersmoothing $\hat{g}$ when constructing the band. In particular, an immediate consequence of Theorem C.1 is that, provided h is chosen such that $h^2 = o\{s^2(h)\}$ and $R(K_U) = o(nh)$, (C.1) holds. Consequently, the central limit theory-based confidence band at (C.2) has asymptotically correct coverage at each point.

C.3 Estimating ν

A consistent estimator of $\nu^2(x)$ can be defined by

$$\hat{\nu}^2(x) = \hat{\sigma}^2\,\hat{f}_W(x) + E\Big[\{\hat{g}(X^*) - \hat{g}(x)\}^2\,f_U(x - X^*)\;\Big|\;\mathcal{Z}\Big]\,, \tag{C.3}$$

where $\hat{\sigma}$ is defined at (3.1); $\hat{g}$ is given by (2.4); the resampled datum X^* is drawn from the population with distribution function $\widetilde{F}_X$, introduced in Section D.2; and $\hat{f}_W$ is a standard error-free kernel estimator:

$$\hat{f}_W(x) = \frac{1}{nh_W} \sum_{i=1}^{n} K_1\bigg(\frac{x - W_i}{h_W}\bigg)\,,$$

with K_1 a kernel of standard type for density estimation, and h_W a bandwidth of conventional size for error-free kernel density estimation. Assuming that, on a compact interval $\mathcal{J}$, $\hat{g}$ is bounded; and supposing in addition that the support of $\widetilde{F}_X$ is confined to $\mathcal{J}$; the value of $\hat{\nu}$ is guaranteed to be finite.

We could also define $\hat{\nu}$ by replacing the second term on the right-hand side of (C.3) by

$$\int_{-\infty}^{\infty} \{\hat{g}(q) - \hat{g}(x)\}^2 \, f_U(x-q) \, \hat{f}_X(q) \, dq \,,$$

but this quantity can be awkward to calculate since $\hat{f}_X$, as defined at (2.5), generally takes negative values on part of the real line.

The complexity of the problem of estimating ν is apparent not just in terms of the cumbersome nature of its formula, which depends on σ, f_W, f_X and g. Not only are these quantities estimated using different smoothing parameters, there is good reason to use different versions of those parameters for computing the estimator $\hat{\nu}$ than we would employ in other parts of the confidence-band problem. For example, to get best performance of $\hat{\nu}$ we would use, in the construction of $\hat{g}$ for (C.3), a bandwidth that was an order of magnitude smaller than would be employed if we were interested only in estimation of g, and also of different size than we would use if we were undersmoothing $\hat{g}$ for a confidence band.

D Additional concerning the bootstrap bands

D.1 Other bootstrap bands

An alternative equal-tailed band can be defined by

$$\mathrm{CB}_\alpha(\mathcal{I}) = \big\{(x,y) : \, x \in \mathcal{I} \text{ and } \hat{t}^{(2)}_{\alpha/2}(x) \leqslant y \leqslant \hat{t}^{(2)}_{1-\alpha/2}(x)\big\}, \tag{D.1}$$

where, for each x, $\hat{t}^{(2)}_{\alpha/2}(x)$ and $\hat{t}^{(2)}_{1-\alpha/2}(x)$ are such that

$$P\big\{\hat{g}^*(x;h_2) \leqslant \hat{t}^{(2)}_{\alpha/2}(x)\big|\mathcal{Z}\big\} = P\big\{\hat{g}^*(x;h_2) > \hat{t}^{(2)}_{1-\alpha/2}(x)\big|\mathcal{Z}\big\} = \tfrac{\alpha}{2}, \tag{D.2}$$

A symmetric band can be defined by

$$\mathrm{CB}_\alpha(\mathcal{I}) = \big\{(x,y) : \, x \in \mathcal{I} \text{ and } \hat{g}(x;h) - \hat{t}^{(3)}_{\alpha/2}(x) \leqslant y \leqslant \hat{g}(x;h) + \hat{t}^{(3)}_{\alpha/2}(x)\big\}, \tag{D.3}$$

where, for each x, $\hat{t}^{(3)}_{\alpha/2}(x)$ is such that

$$P\big\{\big|\hat{g}^*(x;h_2) - \hat{g}(x;h_0)\big| \leqslant \hat{t}^{(3)}_{\alpha/2}(x)\big|\mathcal{Z}\big\} = 1-\alpha. \tag{D.4}$$

These bands give approximately correct coverage at individual points x, i.e., as $n \to \infty$

$$P\big\{\hat{t}^{(2)}_{\alpha/2}(x) \leqslant g(x) \leqslant \hat{t}^{(2)}_{1-\alpha/2}(x)\big\} \to 1-\alpha \quad \text{for all} \quad x \in \mathcal{I}, \tag{D.5}$$

$$P\big\{\hat{g}(x;h) - \hat{t}^{(3)}_{\alpha/2}(x) \leqslant g(x) \leqslant \hat{g}(x;h) + \hat{t}^{(3)}_{\alpha/2}(x)\big\} \to 1-\alpha \quad \text{for all} \quad x \in \mathcal{I}. \tag{D.6}$$

Theoretical results similar to those in Section 5.1 can be established for such bands too. More specifically, let

$$P_2(w) = P\big[\delta(h)^{-1}|\hat{g}(x;h) - g_1(x)| \leqslant A_0\, w\big],$$

let $\alpha \in (0,1)$, and define $z_{2\alpha}$ and $w_{2\alpha}$ to be the solutions of $\Phi(z) = \frac{1}{2}(1+\alpha)$ and $P_2(w) = \alpha$, respectively. Theorem D.1 and Corollary D.1 give, respectively, Edgeworth and Cornish-Fisher expansion relating to the distribution of $\hat{g} - g_1$.

Theorem D.1. *If Conditions A, B and C hold then*

$$\Big|P_2(w) - \Big\{2\,\Phi(w) - 1 + 2\,\delta^2 p_2(w)\phi(w)\Big\}\Big| \leqslant B_1\big\{\delta^3 + (nh)^{-1} + \delta\,(nh)^{-1/2}\big\},$$

for all $n \geqslant n_0$, all $h \in (0, H_0]$ and all $0 \leqslant w \leqslant B_2$, where B_1, h and n_0 depend only on $\mathcal{I}$, the constants $C_4, \ldots, C_9$ and ϵ introduced in Condition B, on the length of the interval $\mathcal{I}_1$ there (the constant is larger for shorter intervals), on the distributions of U and V, on the kernel K and on B_2. If the distribution of V is either $\mathrm{N}(0,\sigma^2)$ *or gamma, as at* (3.2)*, then dependence on the distribution of V, above, can be replaced by dependence on the constants C_1, C_2 and C_3 in (B1).*

Corollary D.1. *If Conditions A, B and C hold then*

$$\big|w_{2\alpha} - \big\{z_{2\alpha} - \delta^2 p_2(z_{2\alpha})\big\}\big| \leqslant B_1\,\big\{\delta^3 + (nh)^{-1} + \delta\,(nh)^{-1/2}\big\},$$

for all $n \geqslant n_0$, all $h \in (0, H_0]$ and all $\alpha \in [B_2, 1-B_2]$, where B_2 denotes any number in the interval $(0,1)$, *and B_1, H_0 and n_0 have the dependence itemised in Theorem D.1.*

Simultaneous analogues of the pointwise confidence bands defined in (D.1) and (D.3) can be constructed by altering equations (D.2) and (D.4) to

$$P\big\{\hat{g}^*(x;h_2) \leqslant \hat{t}^{(2)}_{\alpha/2}(x)\ \forall x \in \mathcal{I}\big|\mathcal{Z}\big\} = P\big\{\hat{g}^*(x;h_2) > \hat{t}^{(2)}_{1-\alpha/2}(x)\big|\mathcal{Z}\ \forall x \in \mathcal{I}|\mathcal{Z}\big\} = \tfrac{\alpha}{2},$$

$$P\big\{\big|\hat{g}^*(x;h_2) - \hat{g}(x;h_0)\big| \leqslant \hat{t}^{(3)}_{\alpha/2}(x)\ \forall x \in \mathcal{I}\big|\mathcal{Z}\big\} = 1-\alpha.$$

In this case, and modulo appropriate assumptions, the following approximations would prevail in place of (D.5) and (D.6):

$$P\big\{\hat{t}^{(2)}_{\alpha/2}(x) \leqslant g(x) \leqslant \hat{t}^{(2)}_{1-\alpha/2}(x)\ \forall x \in \mathcal{I}\big\} \to 1-\alpha,$$

$$P\big\{\hat{g}(x;h) - \hat{t}^{(3)}_{\alpha/2}(x) \leqslant g(x) \leqslant \hat{g}(x;h) + \hat{t}^{(3)}_{\alpha/2}(x)\ \forall x \in \mathcal{I}\big\} \to 1-\alpha.$$

D.2 Estimating the distribution of X

In order to generate the bootstrap versions X_i^* of X_i, we need to estimate the distribution function F_X of X. We use the estimator of Hall and Lahiri (2008), which is defined from (2.5) by

$$\widehat{F}_X(x;h) = \int_{-\infty}^{x} \hat{f}_X(z;h)\,dz = n^{-1}\sum_{j=1}^{n} L_1(x - W_j;h),$$

where

$$L_1(u;h) = \tfrac{1}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\sin tu}{t}\,\frac{\phi_K(ht)}{\phi_U(t)}\,dt. \tag{D.7}$$

In particular, if ϕ_K is compactly supported, if $h > 0$ and if (2.2) holds, then the integral in (D.7) is well-defined and finite.

Generally, $\hat{F}_X$ is not monotone and can take values outside the interval $[0,1]$. We convert it to a proper distribution function in the same way as in Hall and Lahiri (2008), by taking

$$\widetilde{F}_X(v;h) = \min\Big[1, \sup\big\{\widehat{F}_X(z;h)\ :\ z \leqslant v\big\}\Big].$$

Bootstrap samples can then be constructed from $\widetilde{F}_X^{-1}$, the inverse of $\widetilde{F}_X$, which can be constructed by

$$\widetilde{F}_X^{-1}(u;h) = \sup\big\{v :\ \widetilde{F}_X(v;h) \leqslant u\big\},$$

where $0 < u < 1$. Our estimator $\widetilde{F}_X^{-1}$ of F_X^{-1} is such that if, for real numbers u, v_1 and v_2 with $v_1 < v_2$, we have $\widehat{F}_X(v_1;h) \leqslant u \leqslant \widehat{F}_X(v_2;h)$, then $v_1 \leqslant \widetilde{F}_X^{-1}(u;h) \leqslant v_2$.

D.3 Consistency of moment estimators of F_V

To justify the definitions of the estimators $\hat{\sigma}^2$ and $\hat{\zeta}$ at (3.1), observe from (2.1) that

$$\hat{\mu}_2 = A_{20} + 2\,A_{11} + A_{02}\,, \quad \hat{\mu}_3 = A_{30} + 3\,A_{21} + 3\,A_{12} + A_{03}\,,$$

where $A_{jk} = n^{-1}\sum_i g(X_i)^j\,V_i^k$. Assume that $E(V^6) + E\{g(X)^6\} < \infty$. Then, since V and X are independent and $E(V) = 0$, we have for $j = 2,3$: $A_{j0} = \xi_j + O_p(n^{-1/2})$, $A_{02} = \sigma^2 + O_p(n^{-1/2})$, $A_{03} = \zeta + O_p(n^{-1/2})$, $A_{12} = \xi_1\,\sigma^2 + O_p(n^{-1/2})$ and $|A_{11}| + |A_{21}| = O_p(n^{-1/2})$. Therefore, $\hat{\mu}_2 = \xi_2 + \sigma^2 + O_p(n^{-1/2})$ and $\hat{\mu}_3 = \xi_3 + 3\,\xi_1\,\sigma^2 + \zeta + O_p(n^{-1/2})$, whence $\sigma^2 = \hat{\mu}_2 - \xi_2 + O_p(n^{-1/2})$ and $\zeta = \hat{\mu}_3 - \xi_3 - 3\,\xi_1\,\sigma^2 + O_p(n^{-1/2})$. Therefore, provided $\hat{\xi}_1$ and $\hat{\xi}_3$ are consistent for ξ_1 and ξ_3, respectively, the estimators $\hat{\sigma}^2$ and $\hat{\zeta}$ at (3.1) are consistent for σ^2 and ζ. This can be proved using arguments similar to those in Delaigle and Hall (2011).

D.4 Theoretical arguments for the SIMEX bandwidth method

In this section, we give theoretical arguments justifying the SIMEX procedure suggested in Section 3.2. In particular, we prove that the bandwidth h_0 suggested there is an order of magnitude larger than the bandwidths h, h_1 and h_2, as required. We assume throughout that the errors U_i are ordinary smooth of order $\beta > 1$ (see Appendix A).

D.4.1 Theoretical bandwidth h_0

Our methodology for constructing confidence bands involves bandwidths h, h_0, h_1 and h_2. The bandwidths h and h_2 are of conventional size for estimating g in the context of errors-in-variables regression, and can in fact be taken to be identical (see Section 5.2). We take $h_2 = h$ and use the SIMEX procedure of Delaigle and Hall (2008) to choose h, resulting in h and h_2 being of conventional size, i.e. $n^{-1/(2\beta+5)}$. Our method for selecting h_1 is also conventional, although this time for deconvolution density estimation rather than regression, and in consequence, h_1 is also of size $n^{-1/(2\beta+5)}$.

The bandwidth h_0, however is chosen in a nonstandard way, in principle to minimise average coverage error,

$$\mathrm{ACE}(h_0) = \int_{\mathcal{J}} \big\{\mathrm{CP}(x; h_0) - (1-\alpha)\big\}^2 \, dx \,, \tag{D.8}$$

where $\mathrm{CP}(x; h_0)$ denotes the coverage probability of our confidence interval for $g(x)$ (see Section 3.2). Here and throughout this section, we replace the notation $\mathrm{CP}(x; c)$ used in Section 3.2 by $\mathrm{CP}(x; h_0)$, that is we use the notation $\mathrm{CP}(x; h_0) = P\{g(x) \in \mathrm{CB}_{h_0}(x)\}$, and give our arguments directly in terms of h_0 rather than in terms of c_0. This makes the theoretical arguments much simpler to present, without altering the key steps of the method.

We start by showing that choosing h_0 to minimise $\mathrm{ACE}(h_0)$ results in h_0 being an order of magnitude larger than h, h_1 or h_2, as requested. Recall that our confidence interval for $g(x)$ has the form

$$\mathrm{CI}(x) = \big(\hat{g}(x) - \hat{t}_{1-(\alpha/2)}(x) \, , \, \hat{g}(x) - \hat{t}_{\alpha/2}(x)\big) \,, \tag{D.9}$$

where $\hat{g}(x)$ is a deconvolution estimator of $g(x)$, computed using a conventional deconvolution bandwidth of size $n^{-1/(2\beta+5)}$; and the quantile estimator $\hat{t}_\gamma(x)$, for each $\gamma \in (0,1)$, is computed by bootstrapping from an empirical model where g is replaced by an estimator based on the bandwidth h_0:

$$Y_i^* = \hat{g}(X_i^*; h_0) + V_i^* \,, \tag{D.10}$$

in which the pairs (X_i^*, V_i^*) are resampled as suggested in section 2.3. Then,

$$\mathrm{CP}(x; h_0) = P\Big\{g(x) \in \Big(\hat{g}(x) - \hat{t}_{1-(\alpha/2)}(x) \, , \, \hat{g}(x) - \hat{t}_{\alpha/2}(x)\Big)\Big\} \,. \tag{D.11}$$

We can expand $\hat{t}_\gamma$ as follows:

$$\hat{t}_\gamma(x) = t_\gamma(x) + n^{-2/(2\beta+5)}\, W(x) + o_p\Big[n^{-2/(2\beta+5)}\,\Big\{\big(nh_0^{2\beta+5}\big)^{-1/2} + h_0^2\Big\}\Big]\,, \tag{D.12}$$

where

$$W(x) = \big(nh_0^{2\beta+9}\big)^{-1/2}\, a_1(x)\, Q_1(x) + h_0^2\, b_1(x)\,,$$

a_1 and b_1 are functions, $a_1 > 0$, the random variable $Q_1(x)$, depending on n, has the standard normal distribution, and $t_\gamma(x)$ solves the equation $P\{\hat{g}(x) - g(x) \le t_\gamma(x)\} = \gamma$. We can expand t_γ as follows:

$$t_\gamma(x) = n^{-2/(2\beta+5)}\, a_2(x)\, z_\gamma + n^{-2/(2\beta+5)}\, b_2(x) + o\big(n^{-2/(2\beta+5)}\big)\,, \tag{D.13}$$

where a_2 and b_2 are functions, b_2 is proportional to g'', and z_γ is the γ-level quantile of the standard normal distribution. Moreover, for the same functions a_2 and b_2,

$$\hat{g}(x) = g(x) + n^{-2/(2\beta+5)}\, a_2(x)\, Q_2(x) + n^{-2/(2\beta+5)}\, b_2(x) + o_p\big(n^{-2/(2\beta+5)}\big)\,, \tag{D.14}$$

where the random variable $Q_2(x)$ has the standard normal distribution.

Together, (D.11)–(D.14) imply that

$$\begin{aligned}
\mathrm{CP}(x; h_0) = P\Big[& n^{-2/(2\beta+5)}\, a_2(x)\, Q_2(x) + n^{-2/(2\beta+5)}\, b_2(x) \\
&\qquad - t_{1-(\alpha/2)}(x) - n^{-2/(2\beta+5)}\, W(x) + o_p\big(n^{-2/(2\beta+5)}\big)\,, \\
&\quad \le 0 \le n^{-2/(2\beta+5)}\, a_2(x)\, Q_2(x) + n^{-2/(2\beta+5)}\, b_2(x) \\
&\qquad - t_{\alpha/2}(x) - n^{-2/(2\beta+5)}\, W(x) + o_p\big(n^{-2/(2\beta+5)}\big)\Big] \\
= P\Big[& n^{-2/(2\beta+5)}\, a_2(x)\, Q_2(x) - n^{-2/(2\beta+5)}\, a_2(x)\, z_{1-(\alpha/2)} \\
&\qquad - n^{-2/(2\beta+5)}\, W(x) + o_p\big(n^{-2/(2\beta+5)}\big)\,, \\
&\quad \le 0 \le n^{-2/(2\beta+5)}\, a_2(x)\, Q_2(x) - n^{-2/(2\beta+5)}\, a_2(x)\, z_{\alpha/2} \\
&\qquad - n^{-2/(2\beta+5)}\, W(x) + o_p\big(n^{-2/(2\beta+5)}\big)\Big] \\
= P\Big[& z_{\alpha/2} + o_p(1) \le Q_2(x) - W(x)/a_2(x) \le z_{1-(\alpha/2)} + o_p(1)\Big]\,.
\end{aligned} \tag{D.15}$$

The pair $(Q_1(x), Q_2(x))$ is jointly normally distributed, although the coefficient of the correlation depends on the size of $nh_0^{2\beta+5}$; if $nh_0^{2\beta+5} \asymp 1$ then the absolute value of the coefficient is bounded away from 0 and 1, whereas if $nh_0^{2\beta+5} \to \infty$ then the coefficient converges to 0. Therefore,

$$\mathrm{ACE}(h_0) = \int_{\mathcal{J}} \Big\{p(x; h_0) - (1-\alpha)\Big\}^2\, dx + o(1)\,, \tag{D.16}$$

where

$$p(x;h_0) = P\Big\{z_{\alpha/2} \le Q_2(x) - W(x)/a_2(x) \le z_{1-(\alpha/2)}\Big\}\,. \tag{D.17}$$

If $h_0 \to 0$ then the probability in (D.17) converges to $1-\alpha$ if and only if $nh_0^{2\beta+5} \to \infty$. In particular, minimising $\mathrm{ACE}(h_0)$ necessarily produces a bandwidth h_0 that is larger, by an order of magnitude, than the one of size $n^{-1/(2\beta+5)}$ that is optimal for estimating g at a point.

D.4.2 Behaviour of the practical bandwidth h_0^+

As described in Section 3.2, since $\mathrm{ACE}(h_0)$ is not known in practice, we use instead:

$$\widehat{\mathrm{ACE}}^+(h_0) = \int_{\mathcal{J}} \Big\{\widehat{\mathrm{CP}^+}(x;h_0) - (1-\alpha)\Big\}^2\, dx\,, \tag{D.18}$$

where $\widehat{\mathrm{CP}^+}(x;h_0)$ is an estimator of the version $\mathrm{CP}^+(x;h_0)$ of $\mathrm{CP}(x;h_0)$ that arises in a related, but slightly different, deconvolution problem, in which $g(x)$ is replaced by $g^+(w) = E(Y \mid W=w)$, and the observed, noisy explanatory variables are not the W_is but the W_is contaminated by more noise having the distribution of U. (The superscript "+" signifies, here and below, that the variable W now plays the role that had been played previously by X.)

Next we shall prove that $\widehat{\mathrm{ACE}}^+(h_0)$, defined at (D.18), admits an asymptotic formula similar to (D.16):

$$\widehat{\mathrm{ACE}}^+(h_0) = \int_{\mathcal{J}} \Big\{p^+(x;h_0) - (1-\alpha)\Big\}^2\, dx + o_p(1)\,, \tag{D.19}$$

uniformly in h_0 such that $0 < h_0 \le n^{-C_1}$ and $nh_0^{2\beta+9} \ge C_2$, for any constants $C_1 \in (0,1)$ and $C_2 > 0$, where

$$p^+(x;h_0) = P\Big\{z_{\alpha/2} \le Q_2^+(x) - W^+(x)/a_2^+(x) \le z_{1-(\alpha/2)} \;\Big|\; \mathcal{Z}\Big\}$$

and the random variables $Q_2^+(x)$ and $W^+(x)$ have the same definitions as before except that the quantities a_j and b_j used in their definition are replaced by a_j^+ and b_j^+, say. Therefore the conclusion drawn in the last paragraph of Section D.4.1 applies to $\widehat{\mathrm{ACE}}^+(h_0)$ as well as to $\mathrm{ACE}(h_0)$. In particular, minimising $\widehat{\mathrm{ACE}}^+(h_0)$ necessarily produces a bandwidth h_0^+ that is larger, by an order of magnitude, than the one of size $n^{-1/(2\beta+5)}$ that is optimal for estimating g^+ at a point.

It remains to derive (D.19). Recall that $\widehat{\mathrm{ACE}}^+(h_0)$ is defined using a SIMEX argument, starting from an estimator of $g^+(w) = E(Y \mid W=w)$ rather than an estimator of $g(x) =$

$E(Y \mid X = x)$. Conditional on the data $\mathcal{Z}$, let $U_{1,1}, \ldots, U_{1,n}$ denote independent random variables sampled from the known error distribution and independent of the original data, and put

$$W_{1,i} = W_i + U_{1,i}\,.$$

Consider the problem of estimating g^+ using the data $(W_{1,1}, Y_1), \ldots, (W_{1,n}, Y_n)$ rather than the original data in $\mathcal{Z}$. We can solve this problem as before, by deconvolution, obtaining a point estimator $\widehat{g^+} = \widehat{g^+}(\cdot \mid h)$, where h is a conventional deconvolution bandwidth of size $n^{-1/(2\beta+5)}$, and constructing a confidence interval $\mathrm{CI}^+(x)$ analogous to the interval at (D.9). That involves computing a deconvolution estimator $\hat{f}_W$ of the density f_W of W, using the data $W_{1,1}, \ldots, W_{1,n}$ and again employing a bandwidth of size $n^{-1/(2\beta+5)}$; simulating, from this distribution, the bootstrap dataset $\mathcal{W}_1^+ = \{W_1^+, \ldots, W_n^+\}$, which here replaces $\mathcal{X}$; simulating $\mathcal{U}^* = \{U_1^*, \ldots, U_n^*\}$ from the distribution of U, and $\mathcal{V}^* = \{V_1^*, \ldots, V_n^*\}$ from the distribution of V, as before; putting $W_{1,i}^* = W_i^+ + U_i^*$ for $1 \le i \le n$, which replaces W_i^* at (2.9); and using the bootstrap dataset $\mathcal{Z}_1^* = \{(W_{1,1}^*, Y_1^*), \ldots, (W_{1,i}^*, Y_1^*)\}$ in place of $\mathcal{Z}^*$.

For each choice of $\mathcal{W}_1^+$, computed as above, we can construct a confidence interval $\mathrm{CI}^+(x; h_0)$ for g^+. Of course, we do not know g^+, and so in practice we cannot find $\mathrm{CI}^+(x; h_0)$ exactly, but we can estimate $g^+(w) = E(Y \mid W = w)$ relatively accurately by treating the data pairs (W_i, Y_i) as though they came from a conventional, error free regression problem, obtaining an estimator $\widehat{g^+}_{\mathrm{EF}}$ say, where the subscript EF denotes "error free." In this notation, the quantity $\widehat{\mathrm{CP}^+}(x; h_0)$ appearing in (D.18) is defined by

$$\widehat{\mathrm{CP}^+}(x; h_0) = P\Big\{\widehat{g^+}_{\mathrm{EF}}(x) \in \mathrm{CI}^+(x; h_0) \Bigm| \mathcal{Z}\Big\}\,,$$

where the expected value is taken over all choices of $\mathcal{W}_1^+$, conditional on the data. This is a slightly simplified version of the definition of $\widehat{\mathrm{CP}^+}$ in Section 3.3, which does not change the conclusions we shall draw, but makes the arguments briefer and more transparent.

The confidence interval $\mathrm{CI}^+(x)$ is, analogously to $\mathrm{CI}(x)$ at (D.9), given by

$$\mathrm{CI}^+(x) = \Big(\widehat{g^+}(x) - \hat{t}^+_{1-(\alpha/2)}(x)\,,\, \widehat{g^+}(x) - \hat{t}^+_{\alpha/2}(x)\Big)\,,$$

where the quantile estimator $\hat{t}^+_\gamma(x)$ is computed by resampling data from the empirical model

$$Y_i^* = \widehat{g^+}\big(W_i^{+*}; h_0\big) + V_i^*\,, \tag{D.20}$$

and W_i^{+*} has the distribution of W_i^+. The empirical model at (D.20) is a version of (D.10) in the present setting. Using this notation,

$$\widehat{\mathrm{CP}^+}(x; h_0) = P\Big\{\widehat{g^+}_{\mathrm{EF}}(x) \in \Big(\widehat{g^+}(x) - \hat{t}^+_{1-(\alpha/2)}(x)\,,\, \widehat{g^+}(x) - \hat{t}^+_{\alpha/2}(x)\Big) \Bigm| \mathcal{Z}\Big\}\,,$$

which formula is analogous to that at (D.11). The arguments leading to (D.12), (D.13), (D.14) and (D.15) give, on this occasion, the respective results:

$$\begin{aligned}
\hat{t}^+_\gamma(x) &= t^+_\gamma(x) + n^{-2/(2\beta+5)}\, W^+(x) + o_p\Big[n^{-2/(2\beta+5)} \Big\{ \big(nh_0^{2\beta+5}\big)^{-1/2} + h_0^2 \Big\}\Big]\,, \\
t^+_\gamma(x) &= n^{-2/(2\beta+5)}\, a_2^+(x)\, z_\gamma + n^{-2/(2\beta+5)}\, b_2^+(x) + o\big(n^{-2/(2\beta+5)}\big)\,, \\
\widehat{g^+}(x) &= g^+(x) + n^{-2/(2\beta+5)}\, a_2^+(x)\, Q_2^+(x) + n^{-2/(2\beta+5)}\, b_2^+(x) + o_p\big(n^{-2/(2\beta+5)}\big)\,, \\
\widehat{\mathrm{CP}^+}&(x; h_0) \\
&= P\Big[n^{-2/(2\beta+5)}\, a_2^+(x)\, Q_2^+(x) + n^{-2/(2\beta+5)}\, b_2^+(x) \\
&\qquad\qquad - t_{1-(\alpha/2)}(x) - n^{-2/(2\beta+5)}\, W^+(x) + o_p\big(n^{-2/(2\beta+5)}\big)\,, \\
&\qquad \le 0 \le n^{-2/(2\beta+5)}\, a_2^+(x)\, Q_2^+(x) + n^{-2/(2\beta+5)}\, b_2^+(x) \\
&\qquad\qquad - t_{\alpha/2}(x) - n^{-2/(2\beta+5)}\, W^+(x) + o_p\big(n^{-2/(2\beta+5)}\big) \;\Big|\; \mathcal{Z}\Big] \\
&= P\Big[n^{-2/(2\beta+5)}\, a_2^+(x)\, Q_2^+(x) - n^{-2/(2\beta+5)}\, a_2^+(x)\, z_{1-(\alpha/2)} \\
&\qquad\qquad - n^{-2/(2\beta+5)}\, W^+(x) + o_p\big(n^{-2/(2\beta+5)}\big)\,, \\
&\qquad \le 0 \le n^{-2/(2\beta+5)}\, a_2^+(x)\, Q_2^+(x) - n^{-2/(2\beta+5)}\, a_2^+(x)\, z_{\alpha/2} \\
&\qquad\qquad - n^{-2/(2\beta+5)}\, W^+(x) + o_p\big(n^{-2/(2\beta+5)}\big) \;\Big|\; \mathcal{Z}\Big] \\
&= P\Big[z_{\alpha/2} + o_p(1) \le Q_2^+(x) - W^+(x)/a_2^+(x) \le z_{1-(\alpha/2)} + o_p(1) \;\Big|\; \mathcal{Z}\Big]\,, \qquad \text{(D.21)}
\end{aligned}$$

In these results,

$$W^+(x) = \big(nh_0^{2\beta+9}\big)^{-1/2}\, a_1^+(x)\, Q_1^+(x) + h_0^2\, b_1^+(x)\,;$$

the random variables $Q_j^+(x)$ each have a standard normal distribution, conditional on $\mathcal{Z}$; the functions a_j^+ are nonnegative; and the remainders are of the stated orders uniformly in h_0 such that $0 < h_0 \le n^{-C_1}$ and $nh_0^{2\beta+9} \ge C_2$, for any constants $C_1 \in (0,1)$ and $C_2 > 0$. (See the following paragraph for a note on how this is interpreted conditional on $\mathcal{Z}$.) In deriving (D.21) we also used the fact that $\widehat{g^+}_{\mathrm{EF}}(x) = g^+(x) + O_p(n^{-2/5}) = g^+(x) + o_p(n^{-2/(2\beta+5)})$.

The two remainder terms, R_1 and R_2 say, that are denoted by $o_p(1)$ in (D.21), are of that size unconditionally, in the sense that $P(|R_j| > \eta) \to 0$, and so $P(|R_j| > \eta \,|\, \mathcal{Z}) \to 0$ in probability, for $j = 1, 2$. (The other remainder terms admit an analogous interpretation.) Therefore, in view of (D.21) and the definition of $\widehat{\mathrm{ACE}}^+$ at (D.18), (D.19) holds.

D.4.3 Conclusion

We conclude that the bandwidth $\hat{h}_0^+$ chosen to minimise $\widehat{\mathrm{ACE}}^+$ satisfies the conditions that the bandwidth h_0 should satisfy, i.e. it has the right order of magnitude to guarantee consistency of our band. However, its constant multiple can often be improved by the extrapolation step of SIMEX, which involves using also the version $^{++}$ of the problem. Here, the $^{++}$ quantities are defined in the obvious way from Section 3.3. In particular,

$$\widehat{\mathrm{ACE}}^{++}(h_0) = \int_{\mathcal{J}} \Big\{\widehat{\mathrm{CP}}^{++}(x;h_0) - (1-\alpha)\Big\}^2\, dx\,,$$

where $\widehat{\mathrm{CP}}^{++}(x;h_0)$ is an estimator of the version $\mathrm{CP}^{++}(x;h_0)$ of $\mathrm{CP}(x;h_0)$ that arises in the deconvolution problem in which $g(x)$ is replaced by $g^{++}(w) = E(Y \,|\, W+U=w)$, and the observed, noisy explanatory variables are not the W_is but the $W_{2,i}$s, defined in Section 3.3.

Arguments similar to those employed for the $^+$ version can be used to prove that minimising $\mathrm{ACE}^{++}(h_0)$ necessarily produces a bandwidth $\hat{h}_0^{++}$ that is larger, by an order of magnitude, than the one of size $n^{-1/(2\beta+5)}$ that is optimal for estimating g^{++} at a point. Moreover, since, in the $^+$ and $^{++}$ versions, the variables are contaminated with errors of the same distribution, $\hat{h}_0^+$ and $\hat{h}_0^{++}$ necessarily have the same magnitude, that is $\hat{h}_0^+ \asymp \hat{h}_0^{++}$ where this notation is used to indicate that $0 < \lim_{n\to\infty} \hat{h}_0^+/\hat{h}_0^{++} < \infty$, and the convergence is in probability. For the same reason, $\hat{h}^+ \asymp \hat{h}^{++}$, where these bandwidths denote the regression bandwidth $\hat{h}$ of Delaigle and Hall (2008), computed with data coming from the model with regression curves g^+ and g^{++}. In particular, they are of size $n^{-1/(2\beta+5)}$. Letting $\hat{c}_0^+ = \hat{h}_0^+/\hat{h}^+$ and $\hat{c}_0^{++} = \hat{h}_0^{++}/\hat{h}^{++}$ as in Section 3.2, we deduce that $\hat{c}_0^+ \asymp \hat{c}_0^{++}$. Therefore, $\hat{c}_0 = (\hat{c}_0^+)^2/\hat{c}_0^{++} \to \infty$ as $n \to \infty$, which proves that $\hat{h}_0 = \hat{c}_0\, \hat{h}$ is an order of magnitude larger than $\hat{h}$, as required (in fact, $\hat{h}_0 \asymp \hat{h}_0^+ \asymp h_0^{++}$).

As indicated above, the extra step involving the $^{++}$ quantities usually permits a better approximation of h_0 because the relation between the original quantities and the $^+$ quantities is mimicked well by the relation between the $^+$ quantities and the $^{++}$ quantities. In other words, $\hat{c}_0\, \hat{h}$ is usually a better approximation to the theoretical h_0 than is h_0^+. This can be proved formally using arguments employed by Delaigle and Hall (2008) in their regression bandwidth context, but that is beyond the scope of this paper.

E Proofs

E.1 Proof of Theorem 5.1

Preliminaries:

Recall that $a = fg$, where $f = f_X$ and is assumed to be evaluated at a particular $x \in \mathcal{I}$. Put $\hat{f} = \hat{f}_X$, $g_1 = (E\,\hat{a})/(E\,\hat{f})$, $\Delta_a = \hat{a} - E\,\hat{a}$ and $\Delta_f = \hat{f} - E\,\hat{f}$. Let z denote a real number, and set $T = T(z) = \Delta_a - (g_1+z)\Delta_f$ and $z' = z\,E\,\hat{f}$. In this notation, $\hat{g} - g_1 =$

$(\Delta_a - g_1\Delta_f)/(E\,\hat{f} + \Delta_f)$, from which it follows that,

$$P(\hat{g} - g_1 \leqslant z) = P(T \leqslant z') + \theta\, P(\Delta_f \leqslant -E\,\hat{f}), \tag{E.1}$$

where $\theta = 1 - 2P(T \leqslant z'|\Delta_f \leqslant -E\,\hat{f})$ which implies $|\theta| \leqslant 1$. (The added term takes care of the possibility that the denominator in the definition of $\hat{g}$ is not strictly positive.)

Observe that, for $x \in \mathcal{I}$,

$$|f(x) - E\{\hat{f}(x)\}| \leqslant \int |K(u)|\,|f(x) - f(x-hu)|du \leqslant C_6\, h^\epsilon \int |u|^\epsilon |K(u)|du,$$

where C_6 is as in (B2). A similar bound can be derived for $|a(x) - E\{\hat{a}(x)\}|$, using (B3) in place of (B2). Therefore, if $H_0 > 0$ is sufficiently small, its value depending only on $\mathcal{I}$, $C_4, \ldots, C_8$, ϵ and K, we have $|f(x) - E\{\hat{f}(x)\}| \leqslant H_0^{\epsilon/2}$ and $|g_1(x) - g(x)| \leqslant H_0^{\epsilon/2}$. We shall use these bounds at several places in the proof, without explicit mention.

Note too that, $T = (nh)^{-1}\{R(K_U)\}^{1/2} \sum_j \big(S_j - E\,S_j\big)$, where

$$S_j = \{R(K_U)\}^{-1/2}\big[g(X_j) + V_j - \{g_1(x) + z\}\big] K_U\{(x - W_j)/h\}.$$

Furthermore, defining $i = \sqrt{-1}$ for the moment, and writing $\phi_S(t) = E(e^{itS})$ for the characteristic function of $S = S_j$, we have:

$$\begin{aligned}\phi_S(t) =& E\bigg\{\exp\bigg(\frac{it}{R(K_U)^{1/2}}\big[g(X) + V - \{g_1(x) + z\}\big] K_U\{(x - W)/h\}\bigg)\bigg\}\\ =& h \iint \exp\Big(it\big[g(q) - \{g_1(x) + z\}\big] L_0(u)\Big)\phi_V\{tL_0(u)\}\\ & \times f_U(x - q - uh) f_X(q)\, du\, dq,\end{aligned} \tag{E.2}$$

where the integrals are over the whole real hyperplane.

Cramér continuity condition on the distribution of S:
It follows from Cramér's continuity condition on the distribution of V (see (B1)) that,

$$\text{for each } \epsilon_1 > 0, \text{ there exists } \epsilon_2 > 0 \text{ such that } |\phi_V(t)| \leqslant 1 - \epsilon_2 \text{ for all } |t| > \epsilon_1. \tag{E.3}$$

See e.g. Theorem 5, p. 59 of Gnedenko and Kolmogorov (1954). If the distribution of V is either $\mathrm{N}(0, \sigma^2)$ or gamma, as at (3.2), then ϵ_2 depends only on C_1, C_2 and C_3 in (B1).

In view of (C2) there exist constants $B_1, H_0 > 0$, depending only on K and the distribution of U, such that, for a nondegenerate interval $\mathcal{I}_2$, $|L_0(u)| \geqslant B_1$ for all $u \in \mathcal{I}_2$ and all $h \in (0, H_0]$. Given $\eta > 0$, and taking $\epsilon_1 = B_1\,\eta$ at (E.3), if $|t| \geqslant \eta$ we have

$$|tL_0(u)| \geqslant |t|\, B_1 \geqslant \eta\, B_1 = \epsilon_1, \tag{E.4}$$

for all $u \in \mathcal{I}_2$ and all $h \in (0, H_0]$. Then from (E.3) and (E.4) we deduce that there exists $\epsilon_2 > 0$, such that

$$|\phi_V\{t\,L_0(u)\}| \leqslant 1 - \epsilon_2, \text{ for all } |t| \geqslant \eta, \text{ all } u \in \mathcal{I}_2 \text{ and all } h \in (0, H_0]. \tag{E.5}$$

Hence, for all $h \in (0, H_0]$ and all $|t| \geqslant \eta$, it follows from (E.2), (E.5), (B2)–(B4) and (C3) that

$$\begin{aligned}
|\phi_S(t)| &\leqslant h \iint \big|\phi_V\{tL_0(u)\}\big| f_U(x-q-uh) f_X(q)\,du\,dq \\
&= 1 - \epsilon_2\, h \int_{-\infty}^{\infty} \int_{u \in \mathcal{I}_2} f_U(x-q-uh) f_X(q)\,du\,dq \\
&\leqslant 1 - \epsilon_2\, h \int_{q \in \mathcal{I}_1} f_X(q) \int_{u \in \mathcal{I}_2} f_U(x-q-uh)\,du\,dq \\
&\leqslant 1 - \epsilon_2\, h \int_{q \in \mathcal{I}_1} \int_{u \in \mathcal{I}_2} f_X(q) f_U(x-q) du\,dq + \epsilon_2\, h^2\, (\sup|f_U'|)(\sup f_X)|\mathcal{I}_1| \int_{\mathcal{I}_2} |u|\,du \\
&\leqslant 1 - B_2 h,
\end{aligned} \tag{E.6}$$

where $\mathcal{I}_1$ is as in (B4), $\mathcal{I}_2$ is as in (C2), $B_2 > 0$ and H_0 is chosen smaller if necessary (with both B_2 and H_0 depending only on $\mathcal{I}$, the constants $C_4, \ldots, C_9$ in (B), the kernel K, the length of the interval $\mathcal{I}_1$ (the constant is larger for shorter intervals), on $\mathcal{I}_2$ and the distributions of U and V (with dependence on the distribution of V replaced by dependence on C_1, C_2 and C_3 in (B1) if the distribution of V is known to be normal or gamma)). Result (E.6) implies that,

for each $\eta > 0$ there exist $H_0, \eta' > 0$, depending only on the quantities noted above, such that, for all $|t| > \eta$, all $x \in \mathcal{I}$, all real values of z, and all $h \in (0, H_0]$, we have $|\phi_S(t)| \leqslant 1 - \eta'\,h$. (E.7)

Moments of distribution of S_i:
For $r = 2, 3, 4$,

$$\begin{aligned}
E(S_i^r) &= E\Big(\big[g(X_i) + V_i - \{g_1(x) + z\}\big]^r L_0\{(x - W_i)/h\}^r\Big) \\
&= h \int \big[g(q) + v - \{g_1(x) + z\}\big]^r L_0(u)^r f_U(x-q-hu) dF_V(v) f_X(q)\,dq\,du,
\end{aligned} \tag{E.8}$$

and

$$R(K_U)^{1/2} h^{-1} E(S_i) = E\{\hat{a}(x)\} - \{g_1(x) + z\} E\{\hat{f}_X(x)\} = -z\,E\{\hat{f}_X(x)\}. \tag{E.9}$$

Hence, under the conditions of Theorem 5.1,

$|E(S_i^r)|/h \leqslant B_3 < \infty$, uniformly in integers $1 \leqslant r \leqslant 4$, in $h \in (0, H_0]$, in $x \in \mathcal{I}$ and in z such that $|z| \leqslant B_4$, for any $B_4 > 0$, where B_3 and H_0 depend only on $\mathcal{I}$, $C_4, \ldots, C_9$ in (B), on the length of the interval $\mathcal{I}_1$ (the constant is larger for shorter intervals), and on $\mathcal{I}_2$ and the distributions of U and V, with dependence on the distribution of V replaced by dependence on C_1, C_2 and C_3 in (B1) if the distribution of V is known to be normal or gamma. (E.10)

First step in derivation of expansion of distribution of $(nh)^{-1/2} \sum_i (S_i - ES_i)$*:*
Let $w \in \mathbb{R}$ and put $\nu_h^2 = h^{-1} \operatorname{var}(S_i)$ and

$$z = (nh)^{-1/2} R(K_U)^{1/2} \nu_h \, w / E(\hat{f}). \tag{E.11}$$

Note that ν_h depends on z, and that therefore, (E.11) defines z as a function of w; see (E.17) below. In view of (E.1),

$$\begin{aligned} P\Big[(nh)^{1/2} R(K_U)^{-1/2} E(\hat{f}) \{\hat{g}(x;h) - g_1(x)\} \leqslant \nu_h \, w\Big] &= P(T \leqslant z') + \theta \, P(\Delta_f \leqslant -E\,\hat{f}) \\ &= P_3(w) + \theta \, P(\Delta_f \leqslant -E\,\hat{f}), \quad \text{(E.12)} \end{aligned}$$

where

$$P_3(w) = P(T \leqslant z') = P\bigg\{\frac{1}{(nh)^{1/2}} \sum_j \big(S_j - E\,S_j\big) \leqslant \nu_h \, w\bigg\}. \tag{E.13}$$

Define $\gamma_h = E(S_i - E\,S_i)^3/(h\,\nu_h^3)$, a measure of the skewness of the distribution of S_i. The standard form of an Edgeworth expansion of the distribution of a sum of independent random variables, if it is valid for $\sum_i (S_i - E\,S_i)$, asserts that

$$\big|P_3(w) - \big\{\Phi(w) + (nh)^{-1/2} \tfrac{1}{6} \gamma_h \, (1 - w^2) \, \phi(w)\big\}\big| \leqslant B_5 (nh)^{-1}, \tag{E.14}$$

for a constant $B_5 > 0$. See, for example, Petrov (1975, pp. 134–139). In fact, (E.14) is valid:

Lemma E.1. *Under the conditions of Theorem 5.1,* (E.14) *holds uniformly in $x \in \mathcal{I}$ and in values of w satisfying $|w| \leqslant B_4$, for each $B_4 > 0$, where B_5 and H_0 depend only on $\mathcal{I}$, $C_4, \ldots, C_9$ in (B), on the length of the interval $\mathcal{I}_1$ (the constant is larger for shorter intervals), on $\mathcal{I}_2$, on the distributions of U and V, and on B_4. Dependence on the distribution of V can be replaced by dependence on C_1, C_2 and C_3, in (B1), if the distribution of V is known to be normal or gamma.*

Lemma E.1 follows using arguments almost identical to those in section 5.5 of Hall (1992a). For example, note that the Cramér-type continuity condition (E.6), i.e. $|\phi_S(t)| \leqslant 1 - B_2\,h$ for all H_0, t such that $|t| > \eta$ and $0 < H_0 \leqslant 1$, is the analogue of the Cramér's condition given in Lemma 5.6 of Hall (1992a). Also, taking $r = 4$ in (E.10), we obtain the fourth-moment bound $E|S_i|^4/h \leqslant B_3 < \infty$.

Second step in derivation of expansion of distribution of $(nh)^{-1/2} \sum_i (S_i - ES_i)$*:*
Taking $r = 2$ in (E.8), we have

$$\begin{aligned} E(S_i^2) &= h \int \Big[\{g(q) - g_1(x)\}^2 + \sigma^2 + z^2 - 2z \{g(q) - g_1(x)\} \Big] L_0(u)^2 \\ &\qquad\qquad\qquad\qquad \times f_U(x - q - hu) f_X(q)\, dq\, du \\ &= h\,(A_0^2 + A_1\, z^2 - 2\, a_2\, z), \end{aligned} \tag{E.15}$$

where $\sigma^2 = E(V^2)$ and A_0, A_1 and a_2 are as at (5.4), (5.5) and (5.7), respectively.

Results (E.9) and (E.15) imply that

$$\nu_h^2 = h^{-1} \operatorname{var}(S_i) = A_0^2 + 2\, A_3\, z^2 - 2\, a_2\, z, \tag{E.16}$$

where

$$2\, A_3 = A_1 - h\, R(K_U)^{-1} \{E\, \hat f_X\}^2 \sim \bigg(\int \ell^2 \bigg) \int f_U(x - q) f_X(q)\, dq > 0.$$

Recall that

$$\delta = (nh)^{-1/2}\, R(K_U)^{1/2} / E\{\hat f(x)\},$$

which converges to zero, uniformly in $x \in \mathcal{I}$ and $|w| \leqslant B_4$, and uniformly in the sense of Theorem 5.1, as $h \to 0$. The definition (E.11) entails $z = \delta\, \nu_h\, w$, whence by (E.16),

$$z^2/(\delta\, w)^2 = A_0^2 + 2\, A_3\, z^2 - 2\, a_2\, z,$$

whence, after solving a quadratic equation for z, we obtain:

$$z = \delta\, w\, \frac{[A_0^2\{1 - 2\, A_3\, (\delta\, w)^2\} + a_2^2\, (\delta\, w)^2]^{1/2} - a_2\, \delta\, w}{1 - 2\, A_3\, (\delta\, w)^2}\,.$$

By Taylor expansion, we obtain

$$z = A_0 \delta\, w \Big\{ 1 - A_0^{-1}\, a_2(\delta\, w) + (A_3 + \tfrac{1}{2} A_0^{-2}\, a_2^2)(\delta\, w)^2 + O(\delta^3) \Big\}, \tag{E.17}$$

uniformly in $H_0 \in (0, 1]$, in $|w| \leqslant B_4$, in x and in the sense of Theorem 5.1. In (E.17) and below, if $\eta > 0$ then "$O(\eta)$" denotes a quantity dominated by $A_3\, \eta$, uniformly in $h \in (0, H_0]$, $|w| \leqslant B_4$ and $x \in \mathcal{I}$, with the quantity η depending only on $\mathcal{I}$, $C_4, \ldots, C_9$ in (B), on the length of the interval $\mathcal{I}_1$, on $\mathcal{I}_2$ (the constant is larger for shorter intervals) and on the distributions of U and V (with dependence on the distribution of V replaced by dependence on C_1, C_2 and C_3, in (B1), if the distribution of V is known to be normal or gamma).

By Taylor expansion again, we deduce that

$$\nu_h^2 = A_0^2 \Big\{ 1 + 2\, A_3\, (\delta\, w)^2 - 2\, a_2\, A_0^{-1}\, \delta\, w\, (1 - A_0^{-1}\, a_2\, \delta\, w) + O(\delta^3) \Big\}.$$

and

$$\nu_h = A_0 \Big\{ 1 - A_0^{-1}\, a_2\, \delta w + (A_3 + \tfrac{1}{2} A_0^{-2} a_2^2)\, (\delta w)^2 + O(\delta^3) \Big\}. \tag{E.18}$$

Also, by (E.9) and (E.17), we have

$$E(S_i) = -R(K_U)^{-1/2} h\, z\, E(\hat{f}) = O\{R(K_U)^{-1/2} h\, \delta\} = O\{(h/n)^{1/2}\}, \tag{E.19}$$

uniformly in the above sense. From (E.19), (E.16) and (E.18), we deduce that, uniformly in the same sense,

$$\gamma_h = \frac{E(S_i^3) - 3\, E(S_i) \mathrm{var}(S_i) - E(S_i)^3}{h \nu_h^3} = \frac{E(S_i^3)}{h\, \nu_h^3} + O\{(h/n)^{1/2}\}. \tag{E.20}$$

Note too that, by (E.8),

$$\begin{aligned} E(S_i^3) &= h \iiint \Big[\{g(q) - g_1(x)\}^3 + 3\, (v - z) \{g(q) - g_1(x)\}^2 \\ &\qquad\qquad + 3\, (v - z)^2 \{g(q) - g_1(x)\} + (v - z)^3 \Big] \\ &\qquad\qquad \times L_0(u)^3\, f_U(x - q - hu) dF_V(v)\, f_X(q)\, dq\, du \\ &= h \iint \Big(\{g(q) - g_1(x)\}^3 + E(V^3) - 3 \left[\{g(q) - g_1(x)\}^2 + \sigma^3 \right] z \\ &\quad + \ 3\, (\sigma^2 + z^2)\, \{g(q) - g_1(x)\} - z^3 \Big) \\ &\quad \times L_0(u)^3\, f_U(x - q - hu)\, f_X(q)\, dq\, du \\ &= h \Big\{ A_2 - 3\, A_4\, z + 3\, (\sigma^2 + z^2)\, a_3 - A_5\, z^3 \Big\}, \end{aligned}$$

where A_2 is as at (5.6), a_3 is as at (5.7), and

$$\begin{aligned} A_4 &= \iint \left[\{g(g) - g_1(x)\}^2 + \sigma^2 \right] L_0(u)^3\, f_U(x - q - hu)\, f_X(q)\, dq\, du, \\ A_5 &= \iint L_0(u)^3\, f_U(x - q - hu)\, f_X(q)\, dq\, du. \end{aligned} \tag{E.21}$$

Using (E.17), (E.18), (E.20) and (E.21), we find that

$$\begin{aligned} \gamma_h &= (A_2 - 3\, A_4\, A_0\, \delta w + 3\, \sigma^2\, a_3)\, A_0^{-3} (1 + 3\, A_0^{-1}\, a_2\, \delta w) + O\{\delta^2 + (h/n)^{1/2}\} \\ &= (A_2 + 3\, \sigma^2\, a_3 + 3\, A_0^{-1}\, A_2\, a_2\, \delta w - 3\, A_0\, A_4\, \delta w + 9\, \sigma^2\, a_3\, A_0^{-1}\, a_2\, \delta w)\, A_0^{-3} \\ &\quad + O\{\delta^2 + (h/n)^{1/2}\}. \end{aligned} \tag{E.22}$$

Next we replace, in (E.12) and (E.13), $\nu_h\, w$ by $A_0\, w$, where A_0 is as at (5.4). This involves replacing w by $A_0\, w/\nu_h$ on the left-hand side of (E.14); $A_0\, w/\nu_h$ can be written as follows by using (E.18) and Taylor expansion:

$$A_0\, w/\nu_h = w\left\{1 + A_0^{-1}\, a_2\, \delta w + (\tfrac{1}{2}\, A_0^{-2}\, a_2^2 - A_3)(\delta w)^2\right\} + O(\delta^3). \tag{E.23}$$

On combining (E.12), (E.14), (E.22) and (E.23), and recalling the definition of δ at (5.2), we obtain,

$$\begin{aligned}
&\Big| P\Big[(nh)^{1/2}\, R(K_U)^{-1/2}\, \{\hat{g}(x) - g_1(x)\} \le A_0\, w\Big] \\
&\quad = P(A_0 w/\nu_h) + \theta\, P(\Delta_f \le -E\hat{f}) \\
&\quad = \Phi(A_0 w/\nu_h) + (nh)^{-1/2}\, \tfrac{1}{6}\, \gamma_h \left\{1 - (A_0 w/\nu_h)^2\right\} \phi(A_0 w/\nu_h) \\
&\qquad\quad + \theta\, P(\Delta_f \le -E\hat{f}) + O\big\{(nh)^{-1}\big\} \\
&\quad = \Phi(w) + \Big\{A_0^{-1}\, a_2\, \delta w + \big(\tfrac{1}{2}\, A_0^{-2}\, a_2^2 - A_3\big)\, (\delta w)^2\Big\}\, w\, \phi(w) \\
&\qquad\quad + \tfrac{1}{2}\, \big(A_0^{-1}\, a_2\, \delta w\big)^2\, w^2\, \phi'(w) \\
&\qquad\quad + (nh)^{-1/2}\, \tfrac{1}{6}\, \big(A_2 + 3\, \sigma^2\, a_3\big)\, A_0^{-3}\, \big(1 - w^2\big)\, \phi(w) \\
&\qquad\quad + \theta\, P(\Delta_f \le -E\hat{f}) + O\big\{\delta^3 + (nh)^{-1} + \delta\, (nh)^{-1/2}\big\} \\
&\quad = \Phi(w) + \Big\{A_0^{-1}\, a_2\, \delta w + \tfrac{1}{2}\, \big(A_0^{-2}\, a_2^2 - A_1\big)\, (\delta w)^2 \\
&\qquad\quad - \tfrac{1}{2}\, A_0^{-2}\, a_2^2\, (\delta w)^2\, w^2\Big\}\, w\, \phi(w) \\
&\qquad\quad + (nh)^{-1/2}\, \tfrac{1}{6}\, \big(A_2 + 3\, \sigma^2\, a_3\big)\, A_0^{-3}\, \big(1 - w^2\big)\, \phi(w) \\
&\qquad\quad + \theta\, P(\Delta_f \le -E\hat{f}) + O\big\{\delta^3 + (nh)^{-1} + \delta\, (nh)^{-1/2} \\
&\qquad\quad + \delta^2\, h\, (K_U)^{-1}\big\}\,.
\end{aligned} \tag{E.24}$$

Since $\delta\, (nh)^{-1/2} = \delta^2\, (K_U)^{-1/2} \gg \delta^2\, (K_U)^{-1}$ then the term above in $\delta^2\, h\, (K_U)^{-1}$ is redundant. Also, by Hölder's inequality,

$$P(\Delta_f \le -E\hat{f}) \le (E\hat{f})^{-4}\, E\big(\Delta_f^4\big) = O\big(\delta^4\big)\,.$$

Result (5.8) follows from these results and (E.24).

E.2 Proof of Corollary 5.1

Below we use the notation R_j, for $j = 1, \ldots, 8$, to denote quantities that satisfy $|R_j| \le C_j\{\delta^3 + (nh)^{-1} + \delta\, (nh)^{-1/2}\}$ for some finite constant C_j. By Taylor expansion of the left hand side of (5.8), we can write $w_{1\alpha} = z_{1\alpha} + \Delta + R_1$, where

$$\Delta = \delta\chi_1 + \delta^2\chi_2 + (nh)^{-1/2}\chi_3,$$

and the χ_js are functions of $w_{1\alpha}$, bounded uniformly in n, and in the argument of the functions on any compact interval. To determine the χ_js, note that, using the right hand side of (5.8), we find

$$P_1(w_{1\alpha}) = \Phi(w_{1\alpha}) + \delta\, p_1(w_{1\alpha})\phi(w_{1\alpha}) + \delta^2\, p_2(w_{1\alpha})\phi(w_{1\alpha}) + (nh)^{-1/2}p_3(w_{1\alpha})\phi(w_{1\alpha}) + R_2.$$

Taylor expanding each term of the right hand side of the above equality around z_α, and recalling that $P_1(w_{1\alpha}) = \alpha$, we deduce that

$$\begin{aligned}\alpha =&\alpha + \{\delta\chi_1 + \delta^2\chi_2 + (nh)^{-1/2}\chi_3\}\phi(z_{1\alpha}) - \frac{\delta^2\chi_1^2}{2} z_{1\alpha}\phi(z_{1\alpha}) + \delta p_1(z_{1\alpha})\phi(z_{1\alpha}) \\ &+ \delta^2\chi_1(p_1\phi)'(z_{1\alpha}) + \delta^2\, p_2(z_{1\alpha})\phi(z_{1\alpha}) + (nh)^{-1/2}p_3(z_{1\alpha})\phi(z_{1\alpha}) + R_4.\end{aligned}$$

This implies that

$$\begin{aligned}&\delta\chi_1 = -\delta p_1(z_{1\alpha}) + R_5,\\ &\delta^2\chi_2 - \frac{\delta^2\chi_1^2}{2} z_{1\alpha} + \delta^2\chi_1\{p_1'(z_{1\alpha}) - z_{1\alpha}p_1(z_{1\alpha})\} + \delta^2\, p_2(z_{1\alpha}) = R_6,\\ &(nh)^{-1/2}\chi_3 = -p_3(z_{1\alpha}) + R_7.\end{aligned}$$

In particular,

$$\begin{aligned}w_{1\alpha} =& z_{1\alpha} - \delta p_1(z_{1\alpha}) + \frac{\delta^2 p_1^2(z_{1\alpha})}{2} z_{1\alpha} + \delta^2 p_1(z_{1\alpha})\{p_1'(z_{1\alpha}) - z_{1\alpha}p_1(z_{1\alpha})\} - \delta^2\, p_2(z_{1\alpha}) \\ &- (nh)^{-1/2}p_3(z_{1\alpha}) + R_8 \\ =& z_{1\alpha} - \delta p_1(z_{1\alpha}) - \delta^2\, p_2(z_{1\alpha}) + \delta^2 p_4(z_{1\alpha}) - (nh)^{-1/2}p_3(z_{1\alpha}) + R_8,\end{aligned}$$

if we define $p_4(z_{1\alpha}) = p_1^2(z_{1\alpha})z_{1\alpha}/2 + p_1(z_{1\alpha})\{p_1'(z_{1\alpha}) - z_{1\alpha}p_1(z_{1\alpha})\}$. Recalling that $p_1(z) = A_0^{-1}a_2\, z^2$, we conclude that

$$p_4(z) = \tfrac{1}{2}\,(A_0^{-1}a_2)^2\, z^5 + 2\,(A_0^{-1}a_2)^2\, z^3 - (A_0^{-1}a_2)^2\, z^5 = \tfrac{1}{2}\,(A_0^{-1}a_2)^2\,(4 - z^2)\, z^3.$$

This establishes (5.9).

E.3 Consistency of confidence band

For a related problem in the nonparametric density case, see (4.101) of Hall (1992c). Four bandwidths are involved in our procedure: h, h_0, h_1 and h_2. Let b denote any one of these. We assume here that Conditions A, B and C are satisfied and that

$$b = o(1) \text{ and } \delta(b) = o(1). \tag{E.25}$$

To simplify the discussion, we also assume that the curves g and f_X satisfy sufficient smoothness conditions (typically, that they have enough bounded derivatives) required for our results. We assume that K is a second order kernel.

To establish (2.10) we shall prove that

$$P\{\hat{g}(x;h) - g(x) \leqslant \hat{t}_\alpha\} = \alpha + o(1). \tag{E.26}$$

It is not hard to check that we have $\hat{t}_\alpha = \delta(h_2)\hat{A}_0\hat{w}_\alpha + \delta(h_2)\hat{A}_0\hat{\eta}$, where $\hat{w}_\alpha$ is defined in (5.14) and $\hat{\eta} \equiv \{\hat{g}_1(x;h_2) - \hat{g}(x;h_0)\}/\{\delta(h_2)\hat{A}_0(x)\}$. Using Corollary 5.1, we have $\hat{w}_\alpha = z_\alpha + o_p(1)$. Moreover, under Conditions A, B, C and (E.25), $\hat{A}_0 \xrightarrow{P} A_0 > 0$, and if we assume that $h \sim h_2$, then $\delta(h)/\delta(h_2) \to 1$. Under these conditions we deduce that

$$\delta(h_2)\hat{A}_0\hat{w}_\alpha = \delta(h)A_0 z_\alpha + o_P\{\delta(h)\}.$$

To analyse $\hat{\eta}$ we note that, for second order kernels, $g_1(x) - g(x) = h^2 b_1(x) + o(h^2)$, where b_1 depends on g', g'', f_X and f'_X, but not on h. Moreover, $\hat{g}_1(x;h_2) - \hat{g}(x;h_0) = h_2^2\,\hat{b}_1(x;h_0,h_1) + o_p(h_2^2)$ is the version of $g_1(x;h) - g(x)$ with $g^{(j)}(x)$ and $f_X^{(j)}(x)$ replaced by $\hat{g}^{(j)}(x;h_0)$ and $\hat{f}_X^{(j)}(x;h_1)$, respectively, these being the jth derivatives of $\hat{g}(x;h_0)$ and $\hat{f}_X(x;h_1)$. Under smoothness conditions on f_X and g, we have $\hat{f}_X^{(j)}(x;h_1) = f_X^{(j)}(x) + O(h_1^2) + O_p\{\delta(h_1)/h_1^j\}$ for $j = 0, 1$, and $\hat{g}^{(j)}(x;h_0) = g^{(j)}(x) + O(h_0^2) + O_p\{\delta(h_0)/h_0^j\}$ for $j = 1, 2$.

Therefore, if h_0 and h_1 are of standard size, i.e. $h_0^2 \asymp \delta(h_0)$ and $h_1^2 \asymp \delta(h_1)$, then $\hat{b}_1$ does not consistently estimate b_1, but if we take h and h_2 smaller than usual, i.e. $h^2 \sim h_2^2 = o\{\delta(h)\}$, then we have $\hat{t}_\alpha = \delta(h)A_0 z_\alpha + o_P\{\delta(h)\}$ and $\hat{g}(x;h) - g_1(x) = \hat{g}(x;h) - g(x) + o\{\delta(h)\}$. Therefore, using Theorem 5.1,

$$P\{\hat{g}(x;h) - g(x) \leqslant \hat{t}_\alpha\} = P\{\hat{g}(x;h) - g_1(x) \leqslant \delta(h)A_0 z_\alpha\} + o(1) = \alpha + o(1).$$

On the other hand, if we take h, h_1 and h_2 of conventional size, i.e. $h^2 \asymp \delta(h)$, $h_1^2 \asymp \delta(h_1)$ and $h_2^2 \asymp \delta(h_2)$, and choose h_0 of larger than conventional size, so that $\delta(h_0)/h_0^2 = o(1)$, then $\hat{b}_1$ consistently estimates b_1, and $\delta(h_2)\hat{A}_0\hat{\eta} = g_1(x) - g(x) + o_P\{\delta(h)\}$. We deduce that $\hat{t}_\alpha = \delta(h)A_0 z_\alpha + g_1(x) - g(x) + o_P\{\delta(h)\}$, and, using Theorem 5.1,

$$P\{\hat{g}(x;h) - g(x) \leqslant \hat{t}_\alpha\} = P\{\hat{g}(x;h) - g_1(x) \leqslant \delta(h)A_0 z_\alpha\} + o(1) = \alpha + o(1).$$

F Complements to the numerical section

Figure F1 shows graphs of estimated coverage probabilities for $x \in [-5, 5]$, calculated from 100 samples and derived using our bootstrap method in the LAP and EST cases, and the

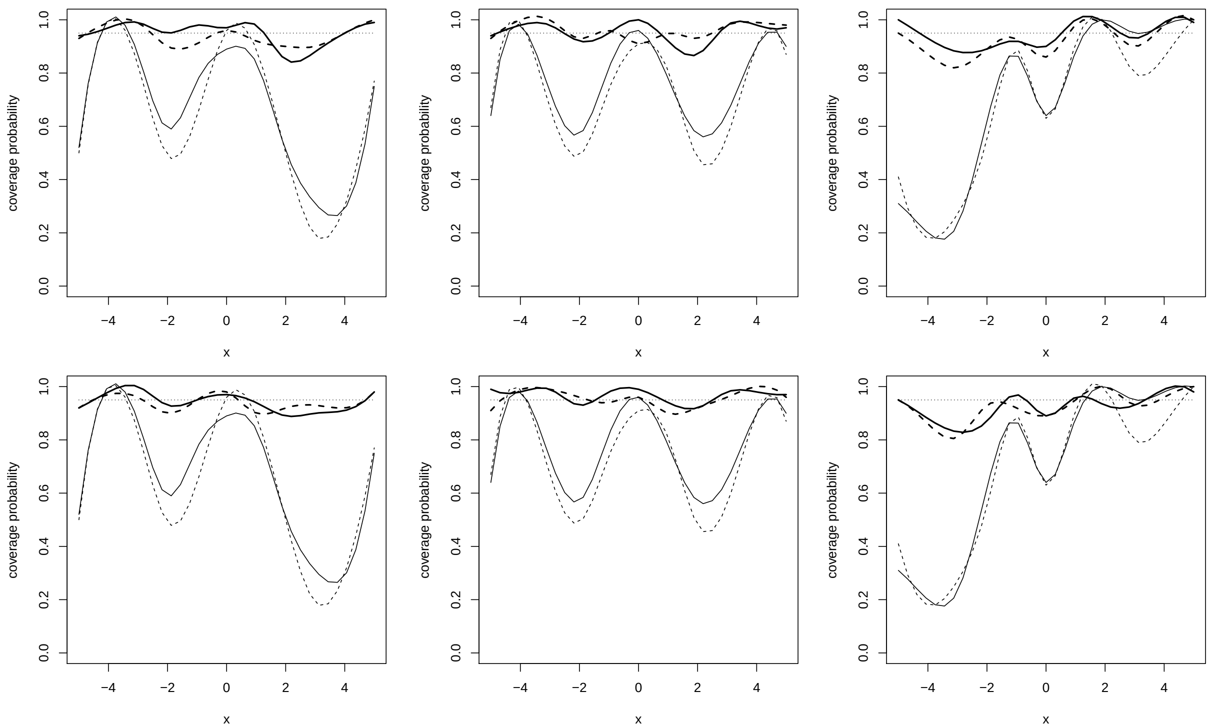

Figure F1: Coverage probability curves derived from the naive approach with $n = 100$ (thin dashed line) and $n = 200$ (thin solid line), and from our bootstrap method with $n = 100$ (thick dashed line) and $n = 200$ (thick solid line), in the LAP case (row 1) and the EST case (row 2); left: $g = g_4$, and error E_D; middle: $g = g_1$ and error E_C; right: $g = g_3$ and error E_E. The horizontal dotted line indicates 0.95 for reference.

naive approach, for samples of size $n = 100$ and 200, and for various combinations of curves and error distributions corresponding to cases (d)–(f) of Figure 1 in the paper.

Figures F2–F4 show, in the KNE case, the confidence bands constructed by our bootstrap method, for the four samples ranked 20, 40, 60 and 80, using the ranking defined in Section 4.3. The figures show the graphs for $n = 100$ and $n = 200$, and for various combinations of regression curves g and error distributions. The figures also show the estimator $\hat{g}$ calculated as in (2.4).

To illustrate the effect of estimating the error or misspecifying it, Figures F5–F10 show, for various combinations of regression curves g and error distributions error distribution, and for $n = 100$ and $n = 200$, the confidence bands constructed by our bootstrap method for four contaminated samples as above, in each of the KNE, LAP and EST settings. Note that the four confidence bands shown in each setting were not computed using the same sample: rather, for each setting we used the samples ranked 20, 40, 60 and 80. As could be expected, the graphs show that the confidence bands corresponding to the KNE case are usually the

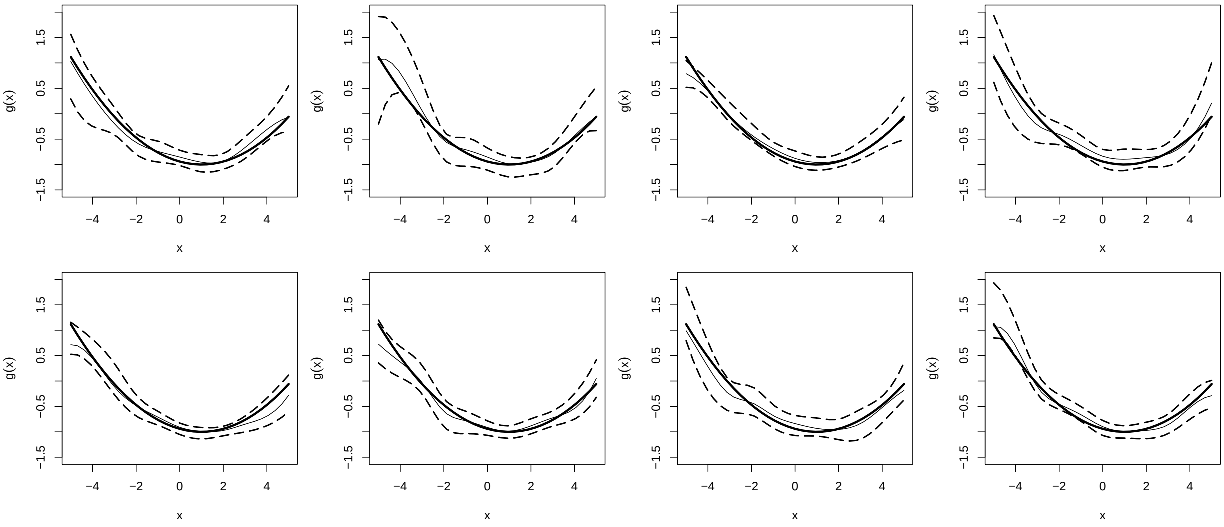

Figure F2: $\hat{g}$ (thin solid line), $g = g_2$ (thick solid line) and 95% confidence bands (dashed lines) for g constructed from the samples ranked 20, 40, 60 and 80 (from left to right), when $n = 100$ (row 1) or 200 (row 2) with error distribution E_A, in the KNE case.

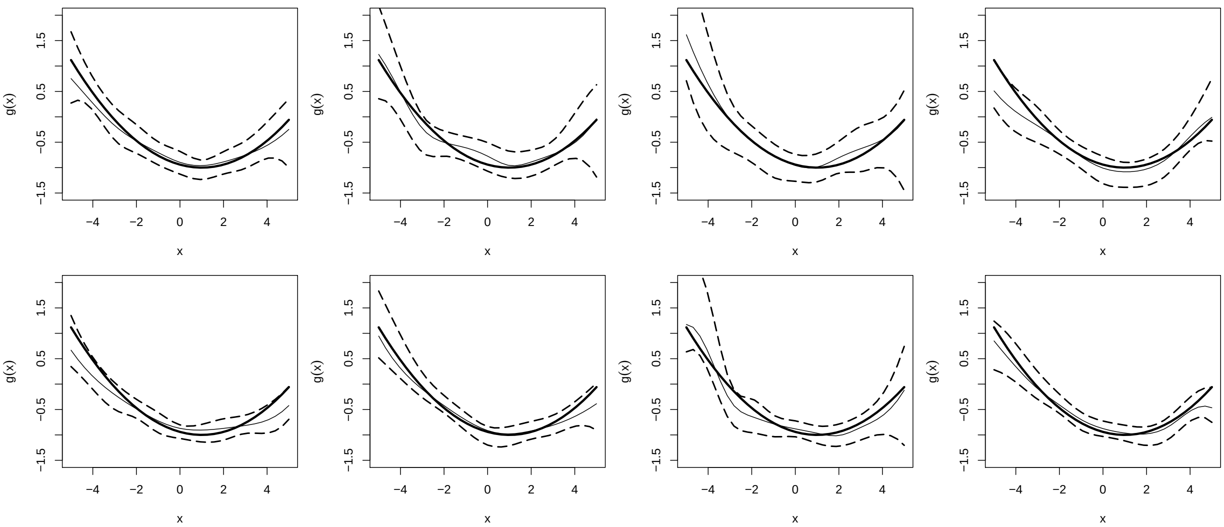

Figure F3: $\hat{g}$ (thin solid line), $g = g_2$ (thick solid line) and 95% confidence bands (dashed lines) for g constructed from the samples ranked 20, 40, 60 and 80 (from left to right), when $n = 100$ (row 1) or 200 (row 2) with error distribution E_B, in the KNE case.

most attractive ones, but the bands obtained in the LAP and KNE case are also quite good. In particular, they have good coverage without being too wide.

Additional References

[1] GNEDENKO, B.V. AND KOLMOGOROV, A.N. (1954). *Limit Distributions for Sums*

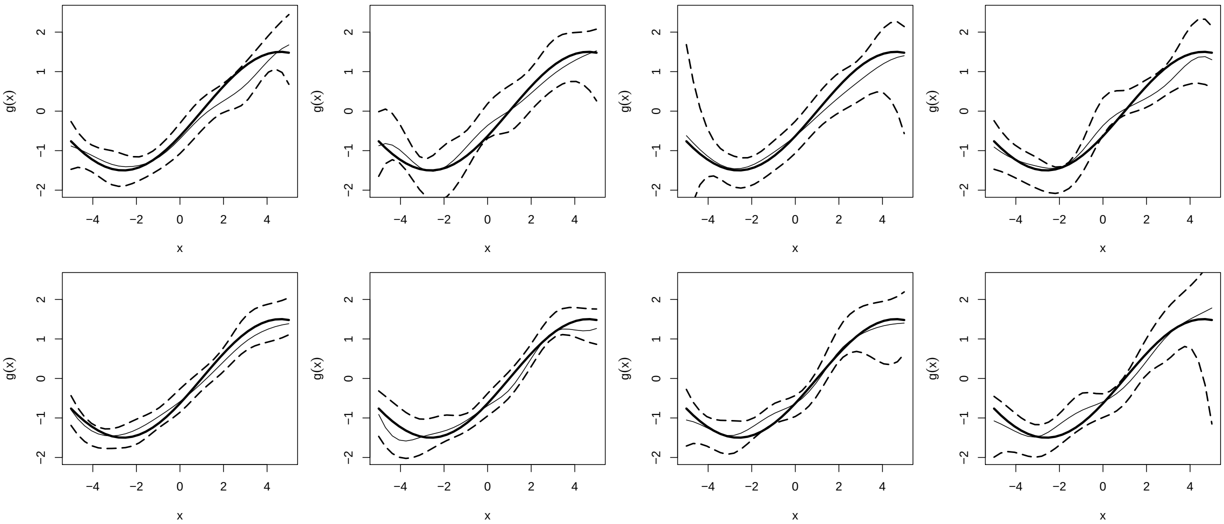

Figure F4: $\hat{g}$ (thin solid line), $g = g_4$ (thick solid line) and 95% confidence bands (dashed lines) for g constructed from the samples ranked 20, 40, 60 and 80 (from left to right), when $n = 100$ (row 1) or 200 (row 2) and with error distribution E_B, in the KNE case.

of Independent Random Variables. Addison-Wesley, Reading, Mass.

[2] PETROV, V.V. (1975). *Sums of Independent Random Variables.* Springer, Berlin.

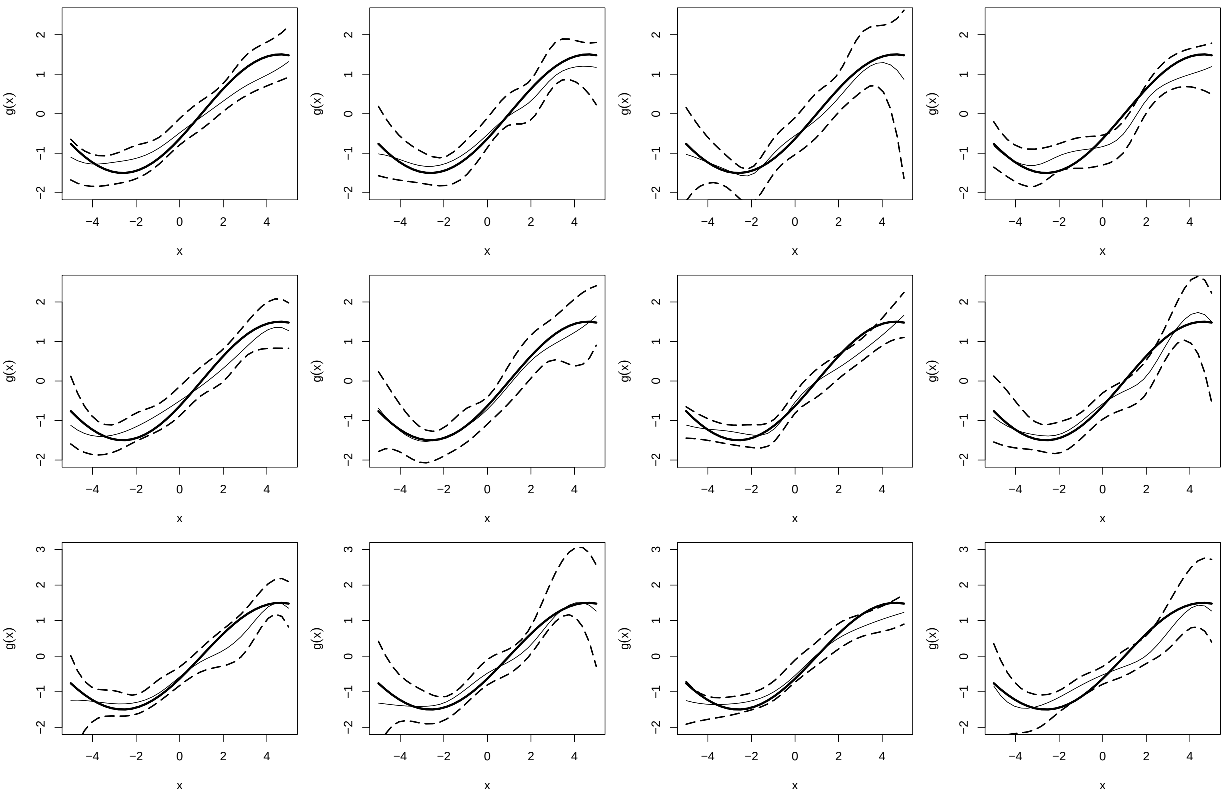

Figure F5: $\hat{g}$ (thin solid line), $g = g_4$ (thick solid line) and 95% confidence bands (dashed lines) for g constructed from the samples ranked 20, 40, 60 and 80 (from left to right), when $n = 100$, with error distribution E_D. Row 1: KNE case, row 2: LAP case, row 3: EST case.

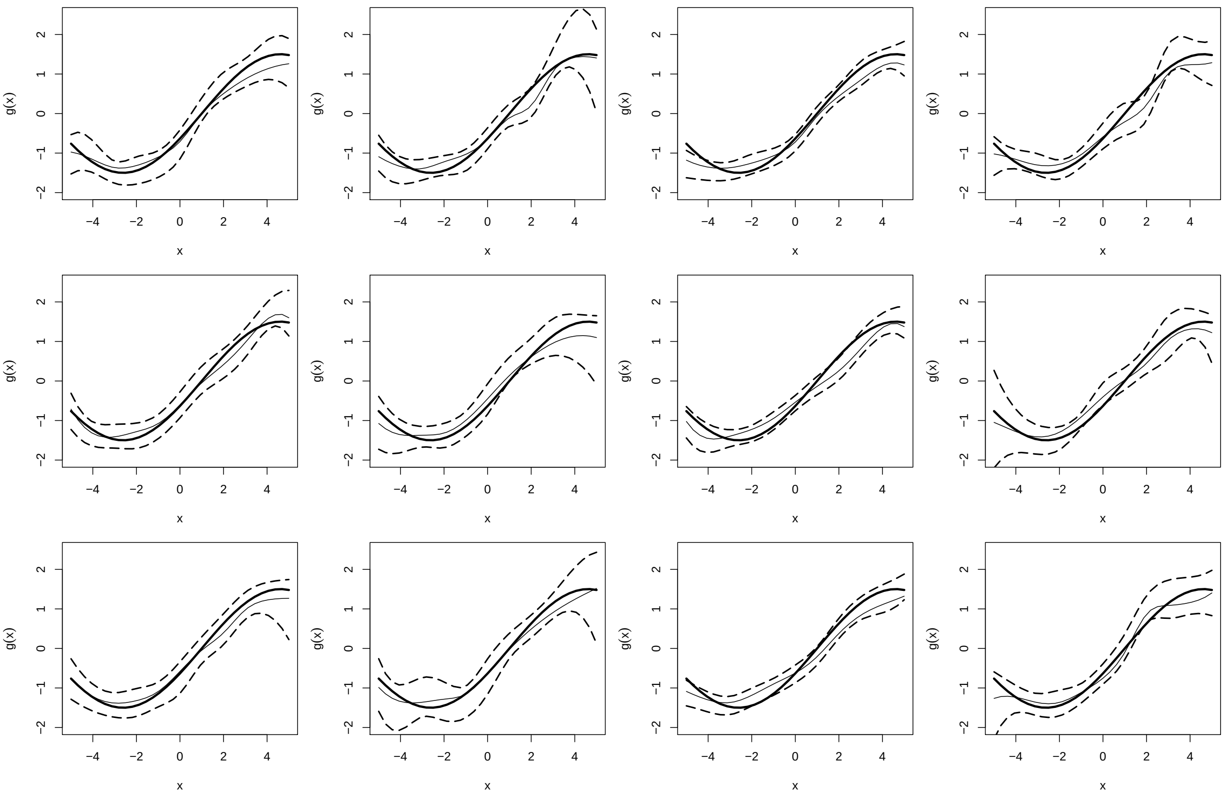

Figure F6: $\hat{g}$ (thin solid line), $g = g_4$ (thick solid line) and 95% confidence bands (dashed lines) for g constructed from the samples ranked 20, 40, 60 and 80 (from left to right), when $n = 200$, with error distribution E_D. Row 1: KNE case, row 2: LAP case, row 3: EST case.

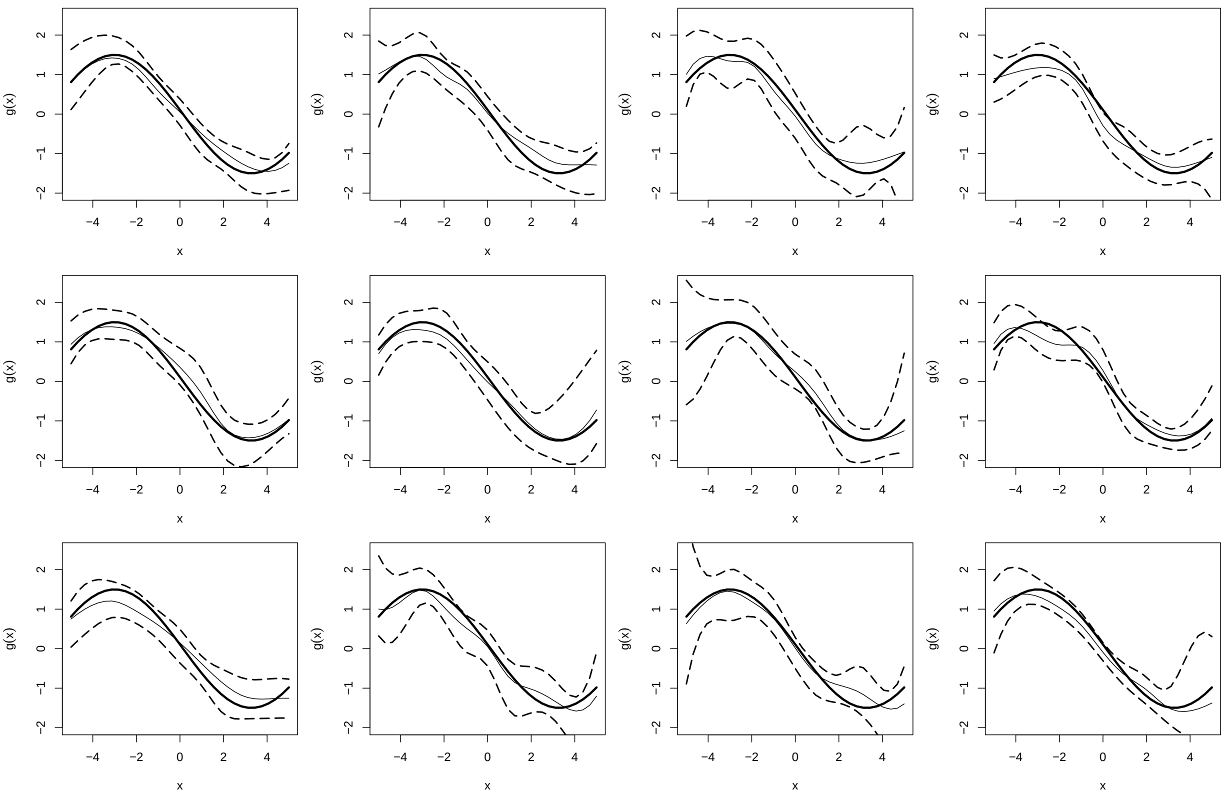

Figure F7: $\hat{g}$ (thin solid line), $g = g_1$ (thick solid line) and 95% confidence bands (dashed lines) for g constructed from the samples ranked 20, 40, 60 and 80 (from left to right), when $n = 100$ and with error distribution E_C. Row 1: KNE case, row 2: LAP case, row 3: EST case.

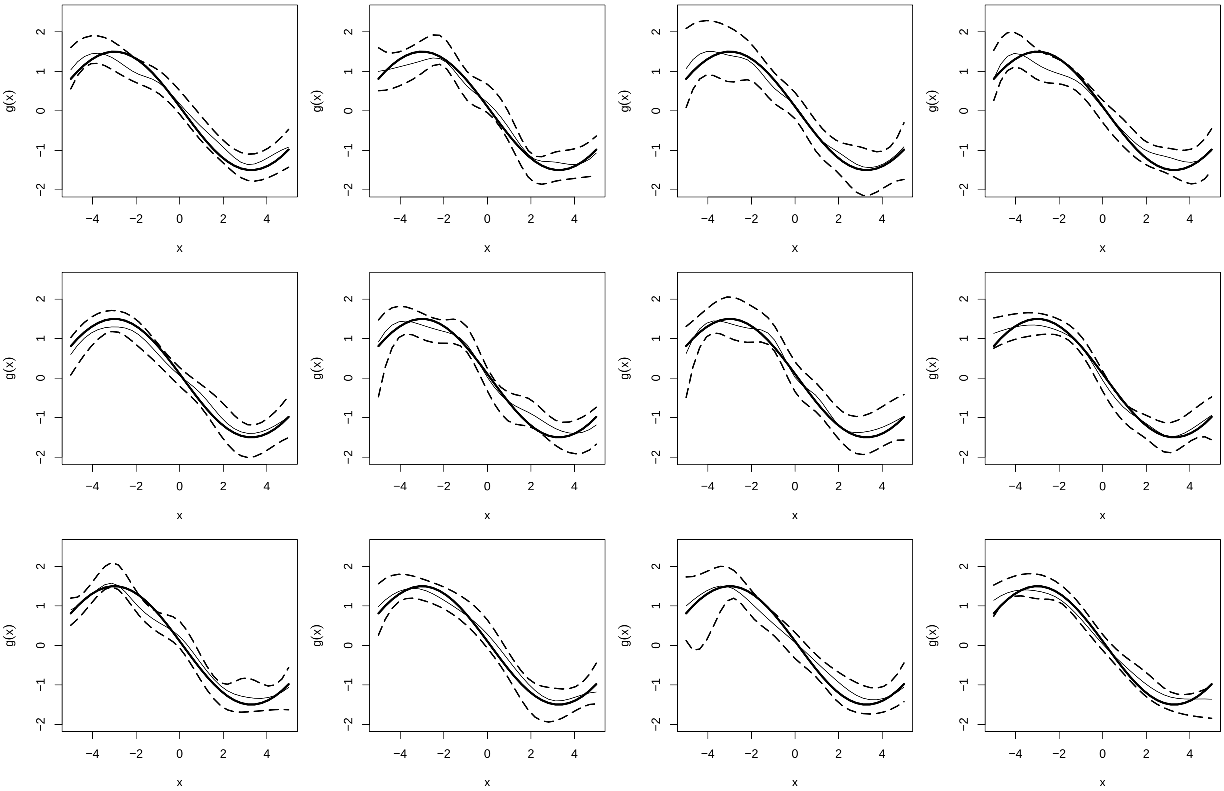

Figure F8: $\hat{g}$ (thin solid line), $g = g_1$ (thick solid line) and 95% confidence bands (dashed lines) for g constructed from the samples ranked 20, 40, 60 and 80 (from left to right), when $n = 200$ and with error distribution E_C. Row 1: KNE case, row 2: LAP case, row 3: EST case.

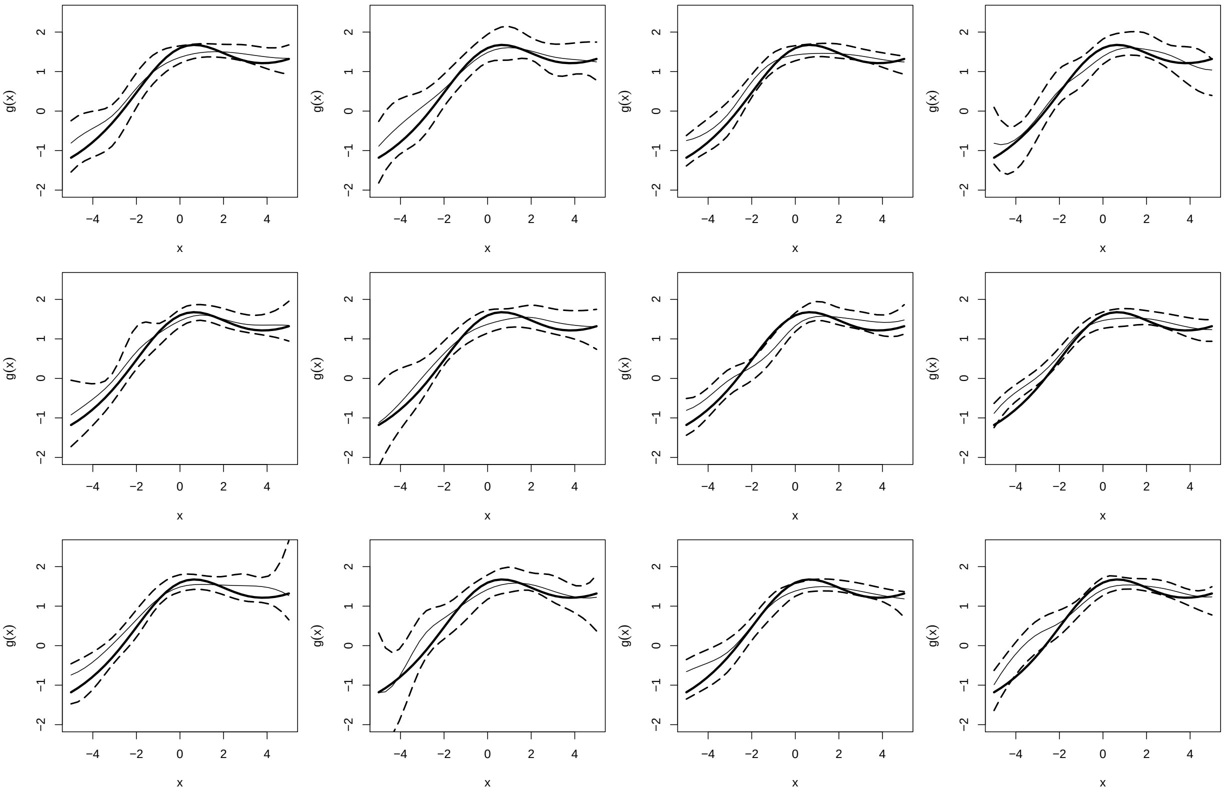

Figure F9: $\hat{g}$ (thin solid line), $g = g_3$ (thick solid line) and 95% confidence bands (dashed lines) for g constructed from the samples ranked 20, 40, 60 and 80 (from left to right), when $n = 100$ and with error distribution E_E. Row 1: KNE case, row 2: LAP case, row 3: EST case.

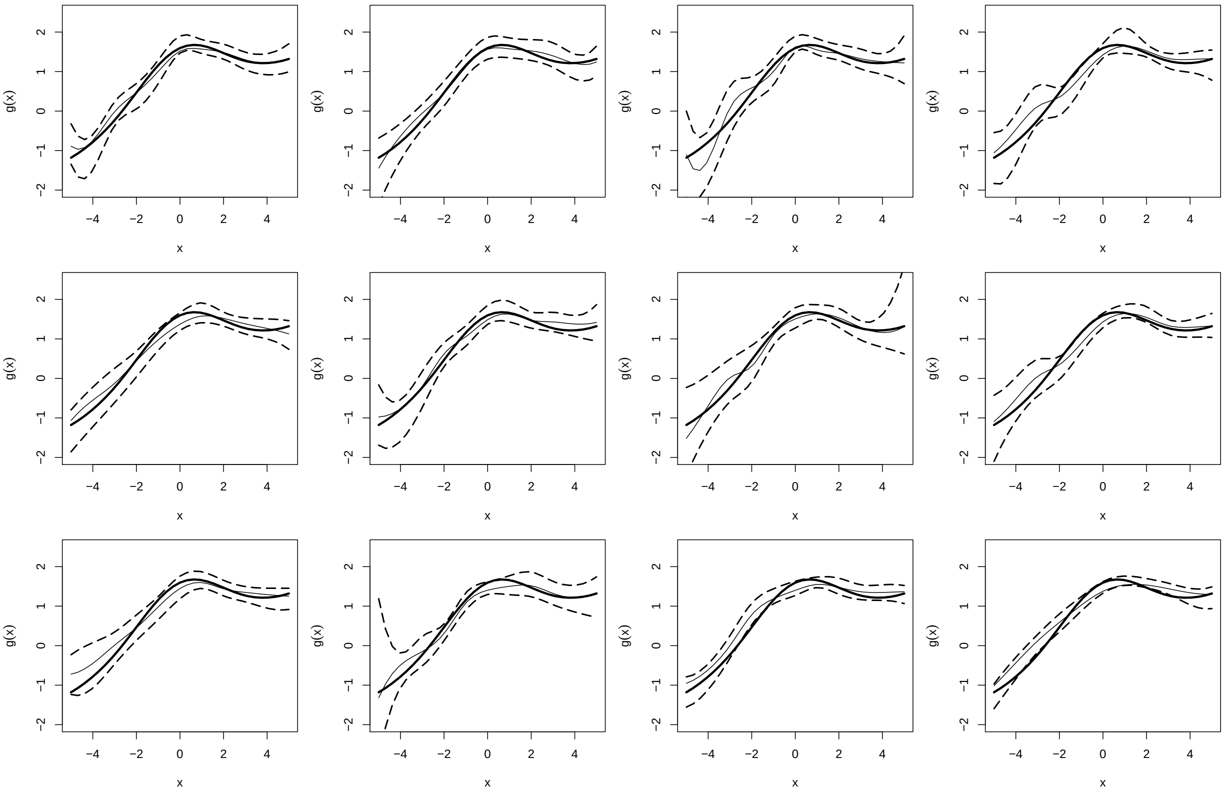

Figure F10: $\hat{g}$ (thin solid line), $g = g_3$ (thick solid line) and 95% confidence bands (dashed lines) for g constructed from the samples ranked 20, 40, 60 and 80 (from left to right), when $n = 200$ and with error distribution E_E. Row 1: KNE case, row 2: LAP case, row 3: EST case.