Bernoulli **19**(5B), 2013, 2250–2276
DOI: 10.3150/12-BEJ451

Uniform convergence of convolution estimators for the response density in nonparametric regression

ANTON SCHICK[1] and WOLFGANG WEFELMEYER[2]

[1]*Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902, USA.*
E-mail: anton@math.binghamton.edu
[2]*Mathematical Institute, University of Cologne, Weyertal 86-90, 50931 Köln, Germany.*
E-mail: wefelm@math.uni-koeln.de

We consider a nonparametric regression model $Y = r(X) + \varepsilon$ with a random covariate X that is independent of the error ε. Then the density of the response Y is a convolution of the densities of ε and $r(X)$. It can therefore be estimated by a convolution of kernel estimators for these two densities, or more generally by a local von Mises statistic. If the regression function has a nowhere vanishing derivative, then the convolution estimator converges at a parametric rate. We show that the convergence holds uniformly, and that the corresponding process obeys a functional central limit theorem in the space $C_0(\mathbb{R})$ of continuous functions vanishing at infinity, endowed with the sup-norm. The estimator is not efficient. We construct an additive correction that makes it efficient.

Keywords: density estimator; efficient estimator; efficient influence function; functional central limit theorem; local polynomial smoother; local U-statistic; local von Mises statistic; monotone regression function

1. Introduction

Smooth functionals of densities can be estimated by plug-in estimators, and densities of functions of two or more random variables can be estimated by local von Mises statistics. Such estimators often converge at the parametric rate $n^{1/2}$. The response density of a nonparametric regression model can be written in both ways, but it also involves an additional infinite-dimensional parameter, the regression function. As explained below, this usually leads to a slower convergence rate of response density estimators, *except* when the regression function is strictly monotone in the strong sense that it has a nowhere vanishing derivative. In the latter case, we can again obtain the rate $n^{1/2}$.

Specifically, consider the nonparametric regression model $Y = r(X) + \varepsilon$ with a one-dimensional random covariate X that is independent of the unobservable error variable ε. We impose the following assumptions:

(F) The error variable ε has mean zero, a moment of order greater than $8/3$, and a density f, and there are bounded and integrable functions f' and f'' such that $f(z) = \int_{-\infty}^{z} f'(x)\,\mathrm{d}x$ and $f'(z) = \int_{-\infty}^{z} f''(x)\,\mathrm{d}x$ for $z \in \mathbb{R}$.

1350-7265

(G) The covariate X is *quasi-uniform* on the interval $[0, 1]$ in the sense that its density g is bounded and bounded away from zero on the interval and vanishes outside. Furthermore, g is of bounded variation.

(R) The unknown regression function r is twice continuously differentiable on $[0, 1]$, and r' is strictly positive on $[0, 1]$.

Assume that $(X_1, Y_1), \dots, (X_n, Y_n)$ are n independent copies of (X, Y). We are interested in estimating the density h of the response Y. An obvious estimator is the kernel estimator

$$\frac{1}{n}\sum_{j=1}^{n} K_b(y - Y_j), \qquad y \in \mathbb{R},$$

where $K_b(t) = K(t/b)/b$ for some kernel K and some bandwidth b. Under the above assumptions on f and g, the density h has a Lipschitz-continuous second derivative as demonstrated in Section 2. Thus, if the kernel has compact support and is of order three, and the bandwidth b is chosen proportional to $n^{-1/7}$, then the mean squared error of the kernel estimator is of order $n^{-6/7}$. This means that the estimator has the nonparametric rate $n^{3/7}$ of convergence.

The above kernel estimator neglects the structure of the regression model. We shall see that by exploiting this structure one can construct estimators that have the faster (parametric) rate $n^{1/2}$ of convergence. For this we observe that the density h is the convolution of the error density f and the density q of $r(X)$. The latter density is given by

$$q(z) = \frac{g(r^{-1}(z))}{r'(r^{-1}(z))}, \qquad z \in \mathbb{R}.$$

By our assumptions on r and g, the density q is quasi-uniform on the interval $[r(0), r(1)]$, which is the image of $[0, 1]$ under r. Furthermore, q is of bounded variation. The convolution representation $h = f * q$ suggests a plug-in estimator or *convolution estimator* $\hat{h} = \hat{f} * \hat{q}$ based on estimators $\hat{f}$ and $\hat{q}$ of f and q, for example the kernel estimators

$$\hat{f}(x) = \frac{1}{n}\sum_{j=1}^{n} k_b(x - \hat{\varepsilon}_j) \quad \text{and} \quad \hat{q}(x) = \frac{1}{n}\sum_{j=1}^{n} k_b\big(x - \hat{r}(X_j)\big), \qquad x \in \mathbb{R},$$

with *nonparametric residuals* $\hat{\varepsilon}_j = Y_j - \hat{r}(X_j)$. Setting $K = k * k$, the convolution estimator $\hat{h}(y)$ has the form of a *local von Mises statistic*

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_b\big(y - \hat{\varepsilon}_i - \hat{r}(X_j)\big).$$

In Section 3, we show that the estimator $\hat{h}$ is root-n consistent in the sup-norm and obeys a functional central limit theorem in the space $C_0(\mathbb{R})$ of all continuous function on $\mathbb{R}$ that vanish at plus and minus infinity. As an auxiliary result, Section 2 treats the case of a *known* regression function r. When r is *unknown*, we estimate it by a local quadratic smoother. The required properties of this smoother are proved in Section 5. The convergence rate of $\hat{h}$ follows from a stochastic expansion which in turn is implied by equations (3.1)–(3.4). These equations are proved in Sections 6–9.

Plug-in estimators in nonparametric settings are often efficient; see, for example, Bickel and Ritov [2], Laurent [8], Chaudhuri *et al.* [3] and Efromovich and Samarov [4]. In Section 4, we first calculate the asymptotic variance bound and the efficient influence function for estimators of $h(y)$. Surprisingly our estimator $\hat{h}(y)$ is not efficient unless the error distribution happens to be normal. We construct an additive correction term $\hat{C}(y)$ such that $\hat{h}(y) - \hat{C}(y)$ is efficient for $h(y)$. This estimator again obeys a uniform stochastic expansion and a functional central limit theorem in $C_0(\mathbb{R})$. The proof of this result is given in Section 10.

The estimator $\hat{h}$ used here goes back to Frees [6]. He observed that densities of some (known) transformations $T(X_1, \dots, X_m)$ of $m \geq 2$ independent and identically distributed random variables $X_1, \dots, X_m$ can be estimated pointwise at the parametric rate by a local U-statistic. Saavedra and Cao [15] consider the transformation $T(X_1, X_2) = X_1 + \varphi X_2$ with $\varphi \neq 0$. Schick and Wefelmeyer [19] and [20] obtain this rate in the sup-norm and in L_1-norms for transformations of the form $T(X_1, \dots, X_m) = T_1(X_1) + \cdots + T_m(X_m)$ and $T(X_1, X_2) = X_1 + X_2$. Giné and Mason [7] obtain such functional results in L_p-norms for $1 \leq p \leq \infty$ and general transformations $T(X_1, \dots, X_m)$. The results of Nickl [12] and [13] are also applicable in this context.

The same convergence rates have been obtained for convolution estimators or local von Mises statistics of the stationary density of linear processes. Saavedra and Cao [14] treat pointwise convergence for a first-order moving average process. Schick and Wefelmeyer [18] and [17] consider higher-order moving average processes and convergence in L_1, and Schick and Wefelmeyer [21] and [22] obtain parametric rates in the sup-norm and in L_1 for estimators of the stationary density of invertible linear processes. Analogous pointwise convergence results for response density estimators in nonlinear regression (with responses missing at random) and in nonparametric regression are in Müller [9] and Støve and Tjøstheim [25], respectively. Escanciano and Jacho-Chávez [5] consider the nonparametric regression model and show uniform convergence, on compact sets, of their local U-statistic. Their results allow for a multivariate covariate X, but require the density of $r(X)$ to be bounded and Lipschitz.

In the above applications to regression models and time series, and also in the present paper, the (auto-)regression function is assumed to have a nonvanishing derivative. This assumption is essential. Suppose there is a point x at which the regression function behaves like $r(y) = r(x) + c(y - x)^\nu + o(|y - x|^\nu)$, for y to the left or right of x, with $\nu \geq 2$. Then the density q of $r(X)$ has a strong peak at $r(x)$. This slows down the rate of the convolution density estimator or local von Mises statistic for $h = f * q$. For densities of transformations $T(X_1, X_2) = |X_1|^\nu + |X_2|^\nu$ of independent and identically distributed random variables, see Schick and Wefelmeyer [24] and [23] and the review paper by Müller *et al.* [11]. In their simulations, Escanciano and Jacho-Chávez [5] consider the regression function $r(x) = \sin(2\pi x)$ and a covariate following a Beta distribution. This choice does not fit their assumptions because the density of $r(X)$ is neither bounded nor Lipschitz. Indeed, for $x = 1/4$ and $x = 3/4$, the regression function behaves as above with $\nu = 2$. In this case, the convolution density estimator does not have the rate $\sqrt{n}$, but at best the slower rate $\sqrt{n/\log n}$.

2. Known regression function

We begin by proving an auxiliary result for the (unrealistic) case that the regression function r is *known*. Then we can observe the error $\varepsilon = Y - r(X)$, and we can apply the results for known transformations cited in Section 1. We obtain a root-n consistent estimator of the response density h by the local von Mises statistic

$$\tilde{h}(y) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_b\big(y - \varepsilon_i - r(X_j)\big), \qquad y \in \mathbb{R}.$$

In the following, we specify conditions under which the convergence holds in $C_0(\mathbb{R})$. We shall assume that K is the convolution $k * k$ for some continuous third-order kernel k with compact support. Then we can write

$$\tilde{h}(y) = \tilde{f} * \tilde{q}(y), \qquad y \in \mathbb{R},$$

where

$$\tilde{f}(x) = \frac{1}{n}\sum_{j=1}^{n} k_b(x - \varepsilon_j) \quad \text{and} \quad \tilde{q}(x) = \frac{1}{n}\sum_{j=1}^{n} k_b\big(x - r(X_j)\big), \qquad x \in \mathbb{R}.$$

Setting $f_b = f * k_b$ and $q_b = q * k_b$, we have the decomposition

$$\tilde{f} * \tilde{q} = f_b * q_b + f_b * (\tilde{q} - q_b) + q_b * (\tilde{f} - f_b) + (\tilde{f} - f_b) * (\tilde{q} - q_b).$$

Note that $f_b * q_b = f * q * k_b * k_b = h * K_b$. Since q is of bounded variation and is quasi-uniform on $[r(0), r(1)]$, we may and do assume that q is of the form

$$q(x) = \int_{u \le x} \phi(u)\nu(\mathrm{d}u), \qquad x \in \mathbb{R},$$

where ν is a finite measure with $\nu(\mathbb{R} - [r(0), r(1)]) = 0$, and ϕ is a measurable function such that $|\phi| \le 1$. This allows us to write

$$h(y) = \int f(y - x) q(x)\,\mathrm{d}x = \int F(y - u)\phi(u)\nu(\mathrm{d}u),$$

where F is the distribution function corresponding to the error density f. Indeed,

$$\begin{aligned} h(y) &= \int q(y - x) f(x)\,\mathrm{d}x \\ &= \int\int_{u \le y - x} f(x)\phi(u)\nu(\mathrm{d}u)\,\mathrm{d}x \\ &= \int\int_{x \le y - u} f(x)\,\mathrm{d}x\phi(u)\nu(\mathrm{d}u). \end{aligned}$$

The properties of f now yield that h is three times differentiable with bounded derivatives

$$h'(y) = \int f(y-u)\phi(u)\nu(\mathrm{d}u), \qquad y \in \mathbb{R}, \tag{2.1}$$

$$h''(y) = \int f'(y-u)\phi(u)\nu(\mathrm{d}u), \qquad y \in \mathbb{R}, \tag{2.2}$$

$$h'''(y) = \int f''(y-u)\phi(u)\nu(\mathrm{d}u), \qquad y \in \mathbb{R}. \tag{2.3}$$

As k is of order three, so is K. Thus, it follows from a standard argument that

$$\|h * K_b - h\| = \sup_{y\in\mathbb{R}} \bigl|h * K_b(y) - h(y)\bigr| \le Cb^3$$

for some constant C.

Next, we note that $f_b * (\tilde{q} - q_b) = H_1 * K_b$ with

$$H_1(y) = \frac{1}{n}\sum_{j=1}^{n}\bigl(f\bigl(y - r(X_j)\bigr) - h(y)\bigr), \qquad y \in \mathbb{R}.$$

Similarly, $q_b * (\tilde{f} - f_b) = H_2 * K_b$ with

$$H_2(y) = \frac{1}{n}\sum_{j=1}^{n}\bigl(q(y - \varepsilon_j) - h(y)\bigr), \qquad y \in \mathbb{R}.$$

As shown in Schick and Wefelmeyer [21], $n^{1/2}H_1$ converges in $C_0(\mathbb{R})$ to a centered Gaussian process with covariance function

$$\Gamma_1(s,t) = \mathrm{Cov}\bigl(f\bigl(s - r(X)\bigr), f\bigl(t - r(X)\bigr)\bigr), \qquad s,t \in \mathbb{R},$$

and the following approximation holds,

$$\|H_1 * K_b - H_1\| = \mathrm{o}_p\bigl(n^{-1/2}\bigr).$$

We can write

$$H_2(y) = \int \bigl(\mathbb{F}(y-x) - F(y-x)\bigr)\phi(x)\nu(\mathrm{d}x), \qquad y \in \mathbb{R},$$

where $\mathbb{F}$ is the empirical distribution function based on the errors $\varepsilon_1, \dots, \varepsilon_n$,

$$\mathbb{F}(t) = \frac{1}{n}\sum_{j=1}^{n}\mathbf{1}[\varepsilon_j \le t], \qquad t \in \mathbb{R}.$$

Setting $\Delta = n^{1/2}(\mathbb{F} - F)$ and writing $\|\cdot\|_1$ for the L_1-norm, we obtain for each $\delta > 0$ the inequalities

$$\begin{aligned} T_1(\delta) &= \sup_{|y_1-y_2|\le\delta} n^{1/2}\bigl|H_2 * K_b(y_1) - H_2 * K_b(y_2)\bigr| \\ &\le \sup_{|y_1-y_2|\le\delta} \iint \bigl|\Delta(y_1 - x - bu) - \Delta(y_2 - x - bu)\bigr|\bigl|K(u)\bigr| \,\mathrm{d}u\,\nu(\mathrm{d}x) \\ &\le \|K\|_1 \nu(\mathbb{R}) \sup_{|y_1-y_2|\le\delta} \bigl|\Delta(y_1) - \Delta(y_2)\bigr|. \end{aligned}$$

Similarly, we obtain the inequalities

$$\begin{aligned} T_2(M) &= \sup_{|y|>2M} n^{1/2}\bigl|H_2 * K_b(y)\bigr| \\ &\le \sup_{|y|>2M} \iint \bigl|\Delta(y - x - bu)\bigr|\bigl|K(u)\bigr| \,\mathrm{d}u\,\nu(\mathrm{d}x) \\ &\le \|K\|_1 \nu(\mathbb{R}) \sup_{|y|>M} \bigl|\Delta(y)\bigr| \end{aligned}$$

for all M such that $-M < r(0) - bB < r(1) + bB < M$, where the constant B is such that the interval $[-B, B]$ contains the support of K. From these inequalities, the characterization of compactness as given in Corollary 4 of Schick and Wefelmeyer [21], and the properties of the empirical process, we obtain tightness of the process $n^{1/2}H_2 * K_b$ in $C_0(\mathbb{R})$. We also have

$$n^{1/2}\|H_2 * K_b - H_2\| \le \|K\|_1 \nu(\mathbb{R}) \sup_{|y_1-y_2|\le bB} \bigl|\Delta(y_1) - \Delta(y_2)\bigr|.$$

It is now easy to conclude that $n^{1/2}H_2 * K_b$ converges in $C_0(\mathbb{R})$ to a centered Gaussian process with covariance function

$$\Gamma_2(s,t) = \mathrm{Cov}\bigl(q(s-\varepsilon), q(t-\varepsilon)\bigr), \qquad s, t \in \mathbb{R}.$$

Finally, we have

$$\bigl\|(\tilde{f} - f_b) * (\tilde{q} - q_b)\bigr\| \le \|\tilde{f} - f_b\|_2 \|\tilde{q} - q_b\|_2 = \mathrm{O}_p\bigl((nb)^{-1}\bigr),$$

where $\|\cdot\|_2$ denotes the L_2-norm.

The above yield the following result.

Theorem 1. *Suppose* (F), (G) *and* (R) *hold, the kernel K is the convolution $k * k$ of some continuous third-order kernel k with compact support, and the bandwidth b satisfies $nb^6 \to 0$ and $nb^2 \to \infty$. Then $n^{1/2}(\tilde{h} - h)$ converges in distribution in the space $C_0(\mathbb{R})$ to a centered Gaussian process with covariance function $\Gamma_1 + \Gamma_2$. Moreover,*

$$\|\tilde{h} - h - H_1 - H_2\| = \mathrm{o}_p\bigl(n^{-1/2}\bigr).$$

3. Unknown regression function

Our main result concerns the case of an *unknown* regression function r. Then we do not observe the random variables $\varepsilon_i = Y_i - r(X_i)$ and $r(X_j)$. In the local von Mises statistic $\tilde{h}$ of Section 2, we therefore replace r by a nonparametric estimator $\hat{r}$, substitute the residual $\hat{\varepsilon}_i = Y_i - \hat{r}(X_i)$ for the error ε_i, and plug in surrogates $\hat{r}(X_j)$ for $r(X_j)$. The resulting estimator for $h = f * q$ is then

$$\hat{h}(y) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K_b\big(y - \hat{\varepsilon}_i - \hat{r}(X_j)\big), \qquad y \in \mathbb{R}.$$

Our estimator $\hat{r}$ will be a local quadratic smoother. More precisely, for a fixed x in [0, 1], we estimate $r(x)$ by the first coordinate $\hat{r}(x) = \hat{\beta}_1(x)$ of the weighted least squares estimator

$$\hat{\beta}(x) = \arg\max_{\beta} \frac{1}{nc} \sum_{j=1}^{n} w\left(\frac{X_j - x}{c}\right)\left(Y_j - \beta^\top \psi\left(\frac{X_j - x}{c}\right)\right)^2,$$

where $\psi(x) = (1, x, x^2)^\top$, the weight function w is a three times continuously differentiable symmetric density with compact support $[-1, 1]$, and the bandwidth c is proportional to $n^{-1/4}$. This means that we undersmooth, since an optimal bandwidth for estimating a twice differentiable regression function is proportional to $n^{-1/5}$.

We assume that K is the convolution $k * k$ for some twice continuously differentiable third-order kernel k with compact support. Then we can write our estimator for h as the convolution

$$\hat{h}(y) = \hat{f} * \hat{q}(y), \qquad y \in \mathbb{R},$$

of the residual-based kernel estimator of f,

$$\hat{f}(x) = \frac{1}{n} \sum_{j=1}^{n} k_b(x - \hat{\varepsilon}_j), \qquad x \in \mathbb{R},$$

with the surrogate-based kernel estimator of q,

$$\hat{q}(x) = \frac{1}{n} \sum_{j=1}^{n} k_b\big(x - \hat{r}(X_j)\big), \qquad x \in \mathbb{R}.$$

Similarly as in Section 2, we have the decomposition

$$\hat{f} * \hat{q} = f_b * q_b + f_b * (\hat{q} - q_b) + q_b * (\hat{f} - f_b) + (\hat{f}_b - f) * (\hat{q} - q_b).$$

Let us introduce

$$\bar{\varepsilon} = \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j$$

and

$$H_3(y) = \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j \big(f'\big(y - r(X_j)\big) - h'(y)\big), \qquad y \in \mathbb{R}.$$

We can write H_3 as the convolution $M * f''$ with

$$M(z) = \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j \big(\mathbf{1}\big[r(X_j) \le z\big] - Q(z)\big), \qquad z \in \mathbb{R},$$

where Q denotes the distribution function of $r(X)$. Write $\sigma^2 = E[\varepsilon^2]$ for the error variance. Since $nE[M^2(z)]$ equals $\sigma^2 Q(z)(1 - Q(z))$ and $(1 - Q)Q$ is integrable, we obtain from Corollary 4 in Schick and Wefelmeyer [21] and the remark after it that $n^{1/2}H_3$ converges in distribution in $C_0(\mathbb{R})$ to a centered Gaussian process with covariance function $\sigma^2\Gamma_3$, where

$$\Gamma_3(s,t) = \mathrm{Cov}\big(f'\big(s - r(X)\big), f'\big(t - r(X)\big)\big), \qquad s,t \in \mathbb{R}.$$

Note that f' and f'' are bounded and integrable and therefore square-integrable.

We shall show in Sections 6–9 that

$$\big\| q_b * (\hat{f} - \tilde{f}) - \bar{\varepsilon} h' \big\| = \mathrm{o}_p\big(n^{-1/2}\big), \tag{3.1}$$

$$\big\| f_b * (\hat{q} - \tilde{q}) + \bar{\varepsilon} h' + J \big\| = \mathrm{o}_p\big(n^{-1/2}\big), \tag{3.2}$$

$$\|\hat{f} - f_b\|_2^2 = \mathrm{O}_p\left(\frac{1}{nb}\right), \tag{3.3}$$

$$\|\hat{q} - q_b\|_2^2 = \mathrm{o}_p(b). \tag{3.4}$$

The last two statements require also $nb^4/\log^4 n \to \infty$. These four statements and Theorem 1 yield our main result.

Theorem 2. *Suppose* (F), (G) *and* (R) *hold*, *the kernel K is the convolution $k * k$ of some twice continuously differentiable third-order kernel k with compact support*, *and the bandwidth b satisfies $nb^6 \to 0$ and $nb^4/\log^4 n \to \infty$. Let $\hat{r}$ be the local quadratic estimator for a weight function w that is a three times continuously differentiable symmetric density with compact support* $[-1, 1]$, *and for a bandwidth c proportional to $n^{-1/4}$. Then $n^{1/2}(\hat{h} - h)$ converges in distribution in the space $C_0(\mathbb{R})$ to a centered Gaussian process with covariance function $\Gamma = \Gamma_1 + \Gamma_2 + \sigma^2\Gamma_3$. Moreover*, *we have the uniform stochastic expansion*

$$\|\hat{h} - h - H_1 - H_2 - H_3\| = \mathrm{o}_p\big(n^{-1/2}\big). \tag{3.5}$$

We should point out that $\Gamma(s,t) = \mathrm{Cov}(H(s), H(t))$ for $s, t \in \mathbb{R}$, where

$$H(y) = f\big(y - r(X)\big) + q(y - \varepsilon) - \varepsilon\big(f'\big(y - r(X)\big) - h'(y)\big), \qquad y \in \mathbb{R}.$$

4. An efficient estimator

In this section, we treat the question of efficient estimation for h. For the theory of efficient estimation of real-valued functionals on nonparametric statistical models, we refer to Theorem 2 in Section 3.3 of the monograph by Bickel *et al.* [1]. It follows from (3.5) that the estimator $\hat{h}(y)$ has influence function

$$I_y(X,Y) = q(y-\varepsilon) - h(y) + f\bigl(y - r(X)\bigr) - h(y) - \varepsilon\bigl(f'\bigl(y - r(X)\bigr) - h'(y)\bigr).$$

We shall now show that this differs in general from the efficient influence function. The latter can be calculated as the projection of $I_y(X,Y)$ onto the tangent space of the nonparametric regression model considered here. The tangent space consists of all functions of the form

$$\alpha(X) + \beta(\varepsilon) + \gamma(X)\ell(\varepsilon),$$

where the function α satisfies $\int \alpha(x)g(x)\,\mathrm{d}x = 0$ and $\int \alpha^2(x)g(x)\,\mathrm{d}x < \infty$, the function β satisfies $\int \beta(z)f(z)\,\mathrm{d}z = 0 = \int z\beta(z)f(z)\,\mathrm{d}y$ and $\int \beta^2(z)f(z)\,\mathrm{d}z < \infty$, and the function γ satisfies $\int \gamma^2(x)g(x)\,\mathrm{d}x < \infty$; see Schick [16] for details. The projection of the influence function onto the tangent space is

$$\begin{aligned} I_y^*(X,Y) &= \bigl[f\bigl(y - r(X)\bigr) - h(y)\bigr] + \bigl[q(y-\varepsilon) - h(y) - d(y)\ell(\varepsilon)\bigr] \\ &\quad + \left[d(y) - \frac{1}{J}\bigl(f'\bigl(y - r(X)\bigr) - h'(y)\bigr)\right]\ell(\varepsilon). \end{aligned}$$

Here $\ell = -f'/f$ denotes the score function for location, $J = \int \ell^2(y)f(y)\,\mathrm{d}y$ is the Fisher information, which needs to be finite for efficiency considerations, and $d(y)$ is the expectation $E[q(y-\varepsilon)\varepsilon]$. For later use, we set

$$\lambda(y) = \frac{\ell(y)}{J} - y, \qquad y \in \mathbb{R}.$$

To see that $I_y^*(X,Y)$ is indeed the projection of the influence function onto the tangent space, we note that $I_y^*(X,Y)$ belongs to the tangent space and that the difference

$$I_y(X,Y) - I_y^*(X,Y) = \bigl(f'\bigl(y - r(X)\bigr) - h'(y)\bigr)\lambda(\varepsilon)$$

is orthogonal to the tangent space. For this, one uses the well-known identities $E[\ell(\varepsilon)] = 0$ and $E[\varepsilon\ell(\varepsilon)] = 1$.

We have $I_y(X,Y) = I_y^*(X,Y)$ if and only if $\lambda = 0$, which in turn holds if and only if f is a mean zero normal density. Consequently, our estimator is efficient for normal errors, but not for other errors.

In order to see why our estimator for $h(y)$ is not efficient in general, consider for simplicity the case of known f and g. The efficient influence function is then $-f'(y - r(X))\ell(\varepsilon)/J$. Thus, an estimator $\hat{h}(y)$ of $h(y)$ is efficient if it satisfies the stochastic expansion

$$\hat{h}(y) = h(y) - \frac{1}{n}\sum_{j=1}^{n} \frac{1}{J} f'\bigl(y - r(X_j)\bigr)\ell(\varepsilon_j) + \mathrm{o}_p\bigl(n^{-1/2}\bigr).$$

A candidate would be obtained by replacing, in the relevant terms on the right-hand side, the unknown r by an estimator $\hat{r}$, resulting in the estimator

$$\int f\bigl(y-\hat{r}(x)\bigr)g(x)\,\mathrm{d}x - \frac{1}{n}\sum_{j=1}^{n}\frac{1}{J}f'\bigl(y-\hat{r}(X_j)\bigr)\ell\bigl(Y_j-\hat{r}(X_j)\bigr).$$

This shows that a correction term to the plug-in estimator $\int f(y-\hat{r}(x))g(x)\,\mathrm{d}x$ is required for efficiency.

In the general situation, with f, g and r unknown, we must construct a stochastic term $\hat{C}(y)$ such that

$$\hat{C}(y) = \frac{1}{n}\sum_{j=1}^{n}\bigl(f'\bigl(y-r(X_j)\bigr)-h'(y)\bigr)\lambda(\varepsilon_j) + \mathrm{o}_p\bigl(n^{-1/2}\bigr). \tag{4.1}$$

Then the estimator $\hat{h}(y)-\hat{C}(y)$ has influence function $I_y^*(X,Y)$,

$$\hat{h}(y)-\hat{C}(y) = \frac{1}{n}\sum_{j=1}^{n} I_y^*(X_j,Y_j) + \mathrm{o}_p\bigl(n^{-1/2}\bigr), \tag{4.2}$$

and hence is efficient. We shall construct $\hat{C}(y)$ such that (4.1), and hence (4.2), hold uniformly in y. This implies a functional central limit theorem in $C_0(\mathbb{R})$ also for the improved estimator $\hat{h}-\hat{C}$. We mention that tightness of $n^{1/2}C$, with

$$C(y) = \frac{1}{n}\sum_{j=1}^{n}\lambda(\varepsilon_j)\bigl(f'\bigl(y-r(X_j)\bigr)-h'(y)\bigr), \qquad y\in\mathbb{R},$$

is verified by the same argument as used for $n^{1/2}H_3$.

To construct the correction term, we use sample splitting. Let m denote the integer part of $n/2$. Let $\hat{r}_1$ and $\hat{r}_2$ denote the local quadratic smoothers constructed from the observations $(X_1,Y_1),\dots,(X_m,Y_m)$ or $(X_{m+1},Y_{m+1}),\dots,(X_n,Y_n)$, both with the same bandwidth c as before. Define residuals $\hat{\varepsilon}_{i,j}=Y_j-\hat{r}_i(X_j)$ for $i=1,2$ and $j=1,\dots,n$, and kernel density estimators

$$\hat{f}_1(z) = \frac{1}{m}\sum_{j=1}^{m}\kappa_a(z-\hat{\varepsilon}_{1,j}), \qquad \hat{f}_2(z) = \frac{1}{n-m}\sum_{j=m+1}^{n}\kappa_a(z-\hat{\varepsilon}_{2,j})$$

and

$$\hat{f}_3(z) = \frac{1}{n}\sum_{j=1}^{m}\kappa_a(z-\hat{\varepsilon}_{2,j}) + \frac{1}{n}\sum_{j=m+1}^{n}\kappa_a(z-\hat{\varepsilon}_{1,j}),$$

where $\kappa_a(x) = \kappa(x/a)/a$ for some bandwidth a and a density κ fulfilling Condition K of Schick [16], such as the logistic kernel. Then we can estimate $\ell(z)$ by

$$\hat{\ell}_1(z) = -\frac{\hat{f}_1'(z)}{a + \hat{f}_1(z)} \quad \text{and} \quad \hat{\ell}_2(z) = -\frac{\hat{f}_2'(z)}{a + \hat{f}_2(z)},$$

the Fisher information J by

$$\hat{J} = \frac{1}{n}\sum_{j=1}^{m} \hat{\ell}_2^2(\hat{\varepsilon}_{1,j}) + \frac{1}{n}\sum_{j=m+1}^{n} \hat{\ell}_1^2(\hat{\varepsilon}_{2,j}),$$

and $\lambda(z)$ by

$$\hat{\lambda}_i(z) = \frac{\hat{\ell}_i(z)}{\hat{J}} - z, \qquad i = 1, 2.$$

Finally, we take $\hat{C}(y) = \hat{C}_1(y) + \hat{C}_2(y)$ with

$$\hat{C}_1(y) = \frac{1}{n}\sum_{j=1}^{m}\left(\hat{f}_3'\big(y - \hat{r}_2(X_j)\big) - \frac{1}{m}\sum_{i=1}^{m} \hat{f}_3'\big(y - \hat{r}_2(X_i)\big)\right)\hat{\lambda}_2(\hat{\varepsilon}_{1,j})$$

and

$$\hat{C}_2(y) = \frac{1}{n}\sum_{j=m+1}^{n}\left(\hat{f}_3'\big(y - \hat{r}_1(X_j)\big) - \frac{1}{n-m}\sum_{i=m+1}^{n} \hat{f}_3'\big(y - \hat{r}_1(X_i)\big)\right)\hat{\lambda}_1(\hat{\varepsilon}_{2,j}).$$

We have the following result, which is proved in Section 10.

Theorem 3. *Suppose* (F), (G) *and* (R) *hold*, f *has finite Fisher information* J, *and the bandwidth* a *satisfies* $a \to 0$ *and* $a^8 n \to \infty$. *Then we have the stochastic expansion* $\|\hat{C} - C\| = o_p(n^{-1/2})$.

Theorems 2 and 3 imply that the improved estimator $\hat{h} - \hat{C}$ has the uniform stochastic expansion

$$\sup_{y\in\mathbb{R}}\left|\hat{h}(y) - \hat{C}(y) - h(y) - \frac{1}{n}\sum_{j=1}^{n} I_y^*(Y_j, X_j)\right| = o_p\big(n^{-1/2}\big)$$

and is efficient. As mentioned above, if the errors happen to be normally distributed, then $\lambda = 0$. Therefore, $C = 0$ so that $\hat{C}$ collapses in the sense that $\|\hat{C}\| = o_p(n^{-1/2})$.

5. Properties of the local quadratic smoother

The weighted least squares estimator $\hat{\beta}(x)$ satisfies the normal equation

$$\bar{W}(x)\hat{\beta}(x) = \bar{V}(x)$$

with

$$\bar{W}(x) = \frac{1}{nc}\sum_{j=1}^{n} w\left(\frac{X_j - x}{c}\right)\psi\left(\frac{X_j - x}{c}\right)\psi^\top\left(\frac{X_j - x}{c}\right),$$

$$\bar{V}(x) = \frac{1}{nc}\sum_{j=1}^{n} w\left(\frac{X_j - x}{c}\right)Y_j\psi\left(\frac{X_j - x}{c}\right).$$

Subtracting from both sides of the normal equation the term $\bar{W}(x)\beta(x)$ with

$$\beta(x) = \big(r(x), cr'(x), c^2 r''(x)/2\big)^\top,$$

we arrive at the equality

$$\bar{W}(x)\big(\hat{\beta}(x) - \beta(x)\big) = A(x) + B(x),$$

where

$$A(x) = \frac{1}{nc}\sum_{j=1}^{n} w\left(\frac{X_j - x}{c}\right)\varepsilon_j\psi\left(\frac{X_j - x}{c}\right),$$

$$B(x) = \frac{1}{nc}\sum_{j=1}^{n} w\left(\frac{X_j - x}{c}\right)R(X_j, x)\psi\left(\frac{X_j - x}{c}\right),$$

and

$$\begin{aligned} R(X_j, x) &= r(X_j) - r(x) - r'(x)(X_j - x) - \frac{1}{2}r''(x)(X_j - x)^2 \\ &= \int_0^1 (X_j - x)^2\big(r''\big(x + s(X_j - x)\big) - r''(x)\big)(1 - s)\,\mathrm{d}s. \end{aligned}$$

Since r'' is uniformly continuous on [0, 1], we see that

$$\sup_{0\le x\le 1}\big|B(x)\big| = \mathrm{o}_p\big(c^2\big) = \mathrm{o}_p\big(n^{-1/2}\big).$$

It follows from the proof of Lemma 1 in Müller *et al.* [10] that

$$\sup_{0\le x\le 1}\big|A(x)\big|^2 = \mathrm{O}_p\left(\frac{\log n}{nc}\right)$$

and

$$\sup_{0\le x\le 1}\big|\bar{W}(x) - W(x)\big|^2 = \mathrm{O}_p\left(\frac{\log n}{nc}\right)$$

with

$$W(x) = E\big[\bar{W}(x)\big] = \int g(x+cu)\psi(u)\psi^\top(u)w(u)\,\mathrm{d}u.$$

Since g is quasi-uniform on [0, 1], there is an η with $0 < \eta < 1$ for which

$$\eta < \inf_{|v|=1} v^\top W(x)v \le \sup_{|v|=1} v^\top W(x)v < \frac{1}{\eta} \tag{5.1}$$

holds for all x in [0, 1] and all $c < 1/2$. From this we obtain the expansion

$$\sup_{0\le x\le 1}\big|\bar{W}^{-1}(x) - W^{-1}(x)\big|^2 = \mathrm{O}_p\bigg(\frac{\log n}{nc}\bigg),$$

where M^{-1} denotes a generalized inverse of a matrix M if its inverse does not exist. Combining the above, we obtain that

$$\sup_{0\le x\le 1}\big|\hat{r}(x) - r(x) - D(x)\big(A(x)+B(x)\big)\big| = \mathrm{O}_p\bigg(\frac{\log n}{nc}\bigg), \tag{5.2}$$

where $D(x)$ is the first row of $W^{-1}(x)$. For later use, we note that $|D(x)|^2 \le 3/\eta^2$ for all x in [0, 1] and $c \le 1/2$. We also have

$$\sup_{0\le x\le 1}\big|\hat{r}(x) - r(x) - \hat{\varrho}(x)\big| = \mathrm{o}_p\big(n^{-1/2}\big), \tag{5.3}$$

where

$$\hat{\varrho}(x) = D(x)A(x) = \frac{1}{nc}\sum_{j=1}^n w\bigg(\frac{X_j - x}{c}\bigg)\varepsilon_j D(x)\psi\bigg(\frac{X_j - x}{c}\bigg).$$

It is easy to check that

$$\int \hat{\varrho}^2(x)g(x)\,\mathrm{d}x = \mathrm{O}_p\bigg(\frac{1}{nc}\bigg),$$

$$\frac{1}{n}\sum_{j=1}^n \hat{\varrho}^2(X_j) = \mathrm{O}_p\bigg(\frac{1}{nc}\bigg),$$

$$\sup_{0\le x\le 1}\big|\hat{\varrho}(x)\big|^2 = \mathrm{O}_p\bigg(\frac{\log n}{nc}\bigg).$$

Thus, we obtain

$$\frac{1}{n}\sum_{j=1}^n \big(\hat{r}(X_j) - r(X_j)\big)^2 = \mathrm{O}_p\bigg(\frac{1}{nc}\bigg), \tag{5.4}$$

$$\int \big(\hat{r}(x) - r(x)\big)^4 g(x)\,\mathrm{d}x = \mathrm{O}_p\bigg(\frac{\log n}{n^2c^2}\bigg). \tag{5.5}$$

Let χ be a square-integrable function. Then the function γ defined by

$$\gamma(t) = \int \bigl(\chi(x-t) - \chi(x)\bigr)^2 \,\mathrm{d}x = \int \bigl(\chi(x+t) - \chi(x)\bigr)^2 \,\mathrm{d}x, \qquad t \in \mathbb{R},$$

is bounded by $4\|\chi\|_2^2$ and satisfies $\gamma(t) \to 0$ as $t \to 0$. Using this and the fact that w has support $[-1, 1]$, we derive

$$\begin{aligned} E\biggl[\biggl(\int \bigl(\chi(X \pm cu) - \chi(X)\bigr)u^i w(u)\,\mathrm{d}u\biggr)^2\biggr] &\le E\biggl[\int \bigl(\chi(X \pm cu) - \chi(X)\bigr)^2 w(u)\,\mathrm{d}u\biggr] \\ &\le \|g\| \int \gamma(cu) w(u)\,\mathrm{d}u \\ &\to 0. \end{aligned}$$

Applying this with $\chi = g$, we can conclude

$$E\bigl[\bigl|W(X) - g(X)\Psi\bigr|^2\bigr] \to 0,$$

where $\Psi = \int \psi(u)\psi^\top(u) w(u)\,\mathrm{d}u$. From this and (5.1), we derive that

$$E\bigl[\bigl|g(X)W^{-1}(X) - \Psi^{-1}\bigr|^2\bigr] \to 0.$$

In particular, with $e = (1, 0, 0)^\top$,

$$E\bigl[\bigl|g(X)D(X) - e^\top\Psi^{-1}\bigr|^2\bigr] \to 0.$$

Let us set

$$\begin{aligned} t(X) &= \int g(X - cu)D(X - cu)\psi(u)w(u)\,\mathrm{d}u \\ &= \int \bigl(g(X - cu)D(X - cu) - g(X)D(X)\bigr)\psi(u)w(u)\,\mathrm{d}u \\ &\quad + \bigl(g(X)D(X) - e^\top\Psi^{-1}\bigr)\Psi e + 1. \end{aligned}$$

Then we have

$$\begin{aligned} E\bigl[\bigl(t(X) - 1\bigr)^2\bigr] &\le 6E\biggl[\int \bigl|g(X - cu)D(x - cu) - g(X)D(X)\bigr|^2 w(u)\,\mathrm{d}u\biggr] \\ &\quad + 2E\bigl[\bigl|g(X)D(X) - e^\top\Psi^{-1}\bigr|^2\bigr]|\Psi e|^2 \\ &\to 0, \end{aligned}$$

since $|gD|$ is square-integrable. This can be used to show that

$$\int \hat{\varrho}(x)g(x)\,\mathrm{d}x = \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j t(X_j) = \bar{\varepsilon} + \mathrm{o}_p\big(n^{-1/2}\big).$$

In view of (5.3), this yields

$$\int \big(\hat{r}(x) - r(x)\big)g(x)\,\mathrm{d}x = \bar{\varepsilon} + \mathrm{o}_p\big(n^{-1/2}\big). \tag{5.6}$$

6. Proof of (3.1)

Since q is of bounded variation, we can write $q_b * (\hat{f} - \tilde{f}) = \hat{H}_2 * K_b$, where

$$\hat{H}_2(y) = \frac{1}{n}\sum_{j=1}^{n} \int \big(\hat{\mathbb{F}}(y-z) - \mathbb{F}(y-z)\big)\phi(z)\nu(z), \qquad y \in \mathbb{R},$$

with $\hat{\mathbb{F}}$ denoting the empirical distribution function based on the residuals $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$,

$$\hat{\mathbb{F}}(t) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}[\hat{\varepsilon}_j \le t], \qquad t \in \mathbb{R}.$$

It was shown in Müller *et al.* [10] that

$$\|\hat{\mathbb{F}} - \mathbb{F} - \bar{\varepsilon} f\| = \mathrm{o}_p\big(n^{-1/2}\big).$$

From this and the representation (2.1) of h', we immediately derive the expansion

$$\big\|\hat{H}_2 - \bar{\varepsilon}h'\big\| = \mathrm{o}_p\big(n^{-1/2}\big).$$

This lets us conclude that

$$\big\|q_b * (\hat{f} - \tilde{f}) - \bar{\varepsilon}h'\big\| = \mathrm{o}_p\big(n^{-1/2}\big).$$

7. Proof of (3.2)

Since f' and f'' are bounded, a Taylor expansion and the bounds (5.3) and (5.4) yield the uniform expansion

$$\sup_{y\in\mathbb{R}}\left|\frac{1}{n}\sum_{j=1}^{n}\Big(f\big(y-\hat{r}(X_j)\big) - f\big(y-r(X_j)\big) + f'\big(y-r(X_j)\big)\hat{\varrho}(X_j)\Big)\right| = \mathrm{o}_p\big(n^{-1/2}\big).$$

Now set

$$S_1(y) = \frac{1}{n}\sum_{j=1}^{n} f'\bigl(y - r(X_j)\bigr)\hat{\varrho}(X_j),$$

$$S_2(y) = \int f'\bigl(y - r(x)\bigr)\hat{\varrho}(x)g(x)\,\mathrm{d}x,$$

$$S_3(y) = \frac{1}{n(n-1)}\sum_{i\neq j} f'\bigl(y - r(X_j)\bigr)\varepsilon_i D(X_j)v_c(X_i - X_j),$$

$$S = \frac{1}{n(n-1)}\sum_{i\neq j}\varepsilon_i\left(D(X_j)v_c(X_i - X_j) - \int D(x)v_c(X_i - x)g(x)\,\mathrm{d}x\right)$$

with

$$v_c(z) = w(z/c)\psi(z/c)/c.$$

Then we have

$$\left\| S_1 - \frac{n-1}{n}S_3\right\| \le \|f'\|\frac{1}{n^2}\sum_{j=1}^{n}\bigl|\varepsilon_j D(X_j)v_c(0)\bigr| = \mathrm{O}_p\left(\frac{1}{nc}\right).$$

In view of $h' = f'' * Q$, we have the identity

$$S_3(y) - S_2(y) - h'(y)S = \int f''(z)U(y-z)\,\mathrm{d}z$$

with

$$U(z) = \frac{1}{n(n-1)}\sum_{i\neq j}\varepsilon_i\Bigl(\bigl(\mathbf{1}\bigl[r(X_j)\le z\bigr] - Q(z)\bigr)D(X_j)v_c(X_i - X_j)$$
$$- \int\bigl(\mathbf{1}\bigl[r(x)\le z\bigr] - Q(z)\bigr)D(x)v_c(X_i - x)g(x)\,\mathrm{d}x\Bigr).$$

The terms in the sum have mean zero and are uncorrelated, with second moments bounded by $\sigma^2\mathbf{1}[r(0)\le z\le r(1)]E[|D(X_2)v_c(X_1 - X_2)|^2]$. Thus, we have

$$n(n-1)\int E\bigl[U^2(z)\bigr]\mathrm{d}z \le \sigma^2\bigl(r(1) - r(0)\bigr)E\bigl[\bigl|D(X_2)v_c(X_1 - X_2)\bigr|^2\bigr] = \mathrm{O}(1/c),$$

from which we derive

$$\|S_3 - S_2 - h'S\| \le \|f''\|_2\|U\|_2 = \mathrm{o}_p\bigl(n^{-1/2}\bigr).$$

Similarly, one has $cn(n-1)E[S^2] = \mathrm{O}(1)$ and obtains

$$\|h'S\| = \mathrm{o}_p\bigl(n^{-1/2}\bigr).$$

Next we have $S_2 = N * f''$, where

$$\begin{aligned}
N(z) &= \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j \int \mathbf{1}\big[r(x) \le z\big] D(x) v_c(X_j - x) g(x)\,\mathrm{d}x \\
&= \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j \int \mathbf{1}\big[r(X_j - cu) \le z\big] g(X_j - cu) D(X_j - cu)\psi(u) w(u)\,\mathrm{d}u \\
&= N_1(z) + N_2(z) + N_3(z) + Q(z)N
\end{aligned}$$

with

$$\begin{aligned}
N_1(z) &= \int \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j \big(\mathbf{1}\big[r(X_j - cu) \le z\big] - \mathbf{1}\big[r(X_j) \le z\big]\big) \\
&\qquad\qquad \times g(X_j - cu) D(X_j - cu)\psi(u) w(u)\,\mathrm{d}u, \\
N_2(z) &= \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j \mathbf{1}\big[r(X_j) \le z\big], \\
N_3(z) &= \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j \big(t(X_j) - 1\big)\big(\mathbf{1}\big[r(X_j) \le z\big] - Q(z)\big), \\
N &= \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j \big(t(X_j) - 1\big).
\end{aligned}$$

It is easy to check that $N_2 * f'' = \bar{\varepsilon} h' + H_3$. Recall the identity $Q * f'' = h'$. Using these identities, we see that

$$\big\|S_2 - \bar{\varepsilon} h' - H_3\big\| \le \big\|h'\big\| |N| + \big\|f''\big\|_2 \big(\|N_1\|_2 + \|N_3\|_2\big).$$

We show now that the right-hand side is of order $o_p(n^{-1/2})$. First, we calculate

$$nE\big[N^2\big] = \sigma^2 E\big[\big(t(X) - 1\big)^2\big] \to 0.$$

Second, using the abbreviation $T(u,z) = \mathbf{1}[r(X - cu) \le z] - \mathbf{1}[r(X) \le z]$, we have

$$\begin{aligned}
n\int E\big[N_1^2(z)\big]\,\mathrm{d}z &= \sigma^2 \int E\bigg[\bigg(\int T(u,z) g(X - cu) D(X - cu)\psi(u) w(u)\,\mathrm{d}u\bigg)^2\bigg]\,\mathrm{d}z \\
&\le \sigma^2 \int E\bigg[\int \big(T(u,z) g(X - cu) D(X - cu)\psi(u)\big)^2 w(u)\,\mathrm{d}u\bigg]\,\mathrm{d}z \\
&\le \sigma^2 \int E\bigg[\int T^2(u,z)\,\mathrm{d}z \big(g(X - cu) D(X - cu)\psi(u)\big)^2\bigg] w(u)\,\mathrm{d}u
\end{aligned}$$

$$\le \sigma^2 \int E\big[\big|r(X-cu)-r(X)\big|\big(g(X-cu)D(X-cu)\psi(u)\big)^2\big]w(u)\,\mathrm{d}u$$
$$\to 0.$$

Third, we derive

$$\begin{aligned} n\int E\big[N_3^2(z)\big]\,\mathrm{d}z &= \sigma^2 \int E\big[\big(t(X)-1\big)^2\big(\mathbf{1}\big[r(X)\le z\big]-Q(z)\big)^2\big]\,\mathrm{d}z \\ &= \sigma^2 E\bigg[\big(t(X)-1\big)^2 \int \big(\mathbf{1}\big[r(X)\le z\big]-Q(z)\big)^2\,\mathrm{d}z\bigg] \\ &\le \sigma^2\big(r(1)-r(0)\big)E\big[\big(t(X)-1\big)^2\big] \\ &\to 0. \end{aligned}$$

We can now conclude that $\|S_2-\bar{\varepsilon}h'-H_3\| = \mathrm{o}_p(n^{-1/2})$.

The above relations show that $\|R+\bar{\varepsilon}h'+H_3\| = \mathrm{o}_p(n^{-1/2})$, where

$$R(y) = \frac{1}{n}\sum_{j=1}^{n}\big(f\big(y-\hat{r}(X_j)\big)-f\big(y-r(X_j)\big)\big).$$

Note that $f_b * (\hat{q}-\tilde{q}) = R * K_b$. Thus, the desired (3.2) follows from the bound

$$\begin{aligned} &\big\| f_b * (\hat{q}-\tilde{q}) + \bar{\varepsilon}h' + H_3 \big\| \\ &\quad \le \big\|\big(R+\bar{\varepsilon}h'+H_3\big) * K_b\big\| + \big\|\big(\bar{\varepsilon}h'+H_3\big) * K_b - \bar{\varepsilon}h' - H_3\big\| \\ &\quad \le \big\|R+\bar{\varepsilon}h'+H_3\big\|\,\|K\|_1 + \big\|\big(\bar{\varepsilon}h'+H_3\big) * K_b - \bar{\varepsilon}h' - H_3\big\| \end{aligned}$$

and the tightness of $n^{1/2}(\bar{\varepsilon}h'+H_3)$ in $C_0(\mathbb{R})$.

8. Proof of (3.3)

Without loss of generality, we assume that $c < 1/2$. Then we have the inequality

$$\big|D(x)v_c(X-x)\big| \le \frac{3}{\eta c} w\bigg(\frac{X-x}{c}\bigg), \qquad 0 \le x \le 1. \tag{8.1}$$

Let us set $\hat{a} = \hat{r} - r$, and, for a subset C of $\{1,\dots,n\}$,

$$\hat{a}_C(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}[j \notin C]\big(\varepsilon_j + R(X_j,x)\big)D(x)v_c(X_j-x).$$

Note that $\hat{a}_{\varnothing}(x) = D(x)(A(x)+B(x))$. For $l = 1,\dots,n$ with $l \ne C$ we have

$$\big|\hat{a}_{C\cup l}(x) - \hat{a}_C(x)\big| \le \frac{1}{n}\big|\varepsilon_l + R(X_l,x)\big|\big|D(x)\big|\big|v_c(X_l-x)\big| \le \frac{3}{\eta}\frac{|\varepsilon_l| + c^2\omega(c)}{nc} w\bigg(\frac{X_l-x}{c}\bigg),$$

where

$$\omega(c) = \sup\{|r''(x) - r''(y)|: x, y \in [0,1], |x-y| \le c\}.$$

We abbreviate $\hat{a}_{\{i\}}$ by $\hat{a}_i$ and $\hat{a}_{\{i,j\}}$ by $\hat{a}_{i,j}$. The above inequality and (5.2) yield the rates

$$\frac{1}{n}\sum_{j=1}^{n}\bigl(\hat{a}(X_j) - \hat{a}_j(X_j)\bigr)^2 = \mathrm{O}_p\biggl(\frac{\log^2 n}{n^2c^2}\biggr), \tag{8.2}$$

$$\frac{1}{n}\sum_{j=1}^{n}\int\bigl(\hat{a}(x) - \hat{a}_j(x)\bigr)^2 g(x)\,\mathrm{d}x = \mathrm{O}_p\biggl(\frac{\log^2 n}{n^2c^2}\biggr), \tag{8.3}$$

$$E\bigl[\bigl(\hat{a}_1(X_1) - \hat{a}_{1,2}(X_1)\bigr)^2\bigr] = \mathrm{O}_p\biggl(\frac{1}{n^2c}\biggr). \tag{8.4}$$

Let us now set

$$\bar{T}(z) = \frac{1}{n}\sum_{j=1}^{n} T_j(z,\hat{a}) \quad \text{and} \quad \bar{T}_*(z) = \frac{1}{n}\sum_{j=1}^{n} T_j(z,\hat{a}_j),$$

where

$$T_j(z,a) = k_b\bigl(z - \varepsilon_j + a(X_j)\bigr) - \iint k_b\bigl(z - y + a(x)\bigr) f(y)g(x)\,\mathrm{d}y\,\mathrm{d}x$$

for a continuous function a. It follows from the properties of k that

$$\int\Biggl(\frac{1}{m}\sum_{i=1}^{m}\bigl(k_b(x-x_i) - k_b(x-y_i)\bigr)\Biggr)^2\mathrm{d}x \le b^{-3}\|k'\|_2^2\frac{1}{m}\sum_{i=1}^{m}(x_i - y_i)^2 \tag{8.5}$$

for real numbers $x_1,\dots,x_m$ and $y_1,\dots,y_m$. This inequality and statements (8.2) and (8.3) yield the rate

$$\int\bigl(\bar{T}(z) - \bar{T}_*(z)\bigr)^2\,\mathrm{d}z = \mathrm{O}_p\biggl(\frac{\log^2 n}{b^3n^2c^2}\biggr) = \mathrm{o}_p\biggl(\frac{1}{nb}\biggr).$$

The last step used the fact that $nc^2b^2/\log^2 n$ is of order $n^{1/2}b^2/\log^2 n$ and tends to infinity. In addition, we have

$$nE\bigl[\bar{T}_*^2(z)\bigr] = E\bigl[T_1^2(z,\hat{a}_1)\bigr] + (n-1)E\bigl[T_1(z,\hat{a}_1)T_2(z,\hat{a}_2)\bigr].$$

Conditioning on $\xi = (\varepsilon_2, X_2, \dots, \varepsilon_n, X_n)$, we see that

$$E\bigl[T_1(z,\hat{a}_1)T_2(z,\hat{a}_{1,2})\bigr] = E\bigl[T_2(z,\hat{a}_{1,2})E\bigl(T_1(z,\hat{a}_1)|\xi\bigr)\bigr] = 0.$$

Similarly one verifies that $E[T_1(z,\hat{a}_{1,2})T_2(z,\hat{a}_2)]$ and $E[T_1(z,\hat{a}_{1,2})T_2(z,\hat{a}_{1,2})]$ are zero. An application of the Cauchy–Schwarz inequality shows that

$$E\bigl[T_1(z,\hat{a}_1)T_2(z,\hat{a}_2)\bigr] = E\bigl[\bigl(T_1(z,\hat{a}_1) - T_1(z,\hat{a}_{1,2})\bigr)\bigl(T_2(z,\hat{a}_2) - T_2(z,\hat{a}_{1,2})\bigr)\bigr]$$

is bounded by $E[(T_1(z,\hat{a}_1)-T_1(z,\hat{a}_{1,2}))^2]$ which in turn is bounded by

$$E\bigl[\bigl(k_b\bigl(z-\varepsilon_1-\hat{a}_1(X_1)\bigr)-k_b\bigl(z-\varepsilon_1-\hat{a}_{1,2}(X_1)\bigr)\bigr)^2\bigr].$$

With the help of (8.4) and (8.5), we thus obtain the bound

$$\int E\bigl[\bar{T}_*^2(z)\bigr]\,\mathrm{d}z \le \frac{\|k\|_2^2}{nb}+\frac{(n-1)}{nb^3}\|k'\|_2^2E\bigl[\bigl(\hat{a}_1(X_1)-\hat{a}_{1,2}(X_1)\bigr)^2\bigr]=\mathrm{O}\biggl(\frac{1}{nb}\biggr).$$

It follows that we have the rate $nb\|\bar{T}\|_2^2=\mathrm{O}_p(1)$.

Now we set

$$\hat{f}_*(z)=\iint k_b\bigl(z-y+\hat{a}(x)\bigr)f(y)\,\mathrm{d}y\,g(x)\,\mathrm{d}x=\int f_b\bigl(z+\hat{a}(x)\bigr)g(x)\,\mathrm{d}x.$$

Since $\hat{f}-\hat{f}_*$ equals $\bar{T}$, we have

$$\|\hat{f}-\hat{f}_*\|_2^2=\mathrm{O}_p\biggl(\frac{1}{nb}\biggr). \tag{8.6}$$

A Taylor expansion yields the bound

$$\int\biggl(\hat{f}_*(z)-f_b(z)-f_b'(z)\int\hat{a}(x)g(x)\,\mathrm{d}x\biggr)^2\mathrm{d}z\le\bigl\|f_b''\bigr\|_2^2\int\hat{a}^4(x)g(x)\,\mathrm{d}x.$$

We have $\|f_b'\|_2=\|f'*k_b\|_2\le\|f'\|_2\|k_b\|_1=\|f'\|_2\|k\|_1$ and $\|f_b''\|_2\le\|f''\|_2\|k\|_1$. Using these bounds, (5.5) and (5.6), we obtain the rate

$$\|\hat{f}_*-f_b\|_2^2=\mathrm{O}_p\biggl(\frac{1}{n}\biggr). \tag{8.7}$$

The desired result (3.3) follows from (8.6) and (8.7).

9. Proof of (3.4)

We assume again that $c<1/2$ and set

$$\hat{q}_*(z)=\int k_b\bigl(z-r(x)-\hat{a}(x)\bigr)g(x)\,\mathrm{d}x,\qquad T'(z,a)=\int k_b'\bigl(z-r(x)\bigr)a(x)g(x)\,\mathrm{d}x.$$

An argument similar to the one leading to (8.6) yields

$$\|\hat{q}-\hat{q}_*\|_2^2=\mathrm{O}_p\biggl(\frac{1}{nb}\biggr). \tag{9.1}$$

Note that $\|k_b'\|_2^2 = \mathrm{O}(b^{-3})$ and $\|k_b''\|_2^2 = \mathrm{O}(b^{-5})$. A Taylor expansion and (5.5) yield

$$\int \big(\hat{q}_*(z) - q_b(z) - T'(z,\hat{a})\big)^2 \, \mathrm{d}z \le \|k_b''\|_2^2 \int \hat{a}^4(x) g(x) \, \mathrm{d}x = \mathrm{O}_p\bigg(\frac{\log n}{b^5 n^2 c^2}\bigg) = \mathrm{o}_p\bigg(\frac{1}{nb^3}\bigg).$$

In view of (5.3), we find

$$\int \big(T'(z,\hat{a}) - T'(z,\hat{\varrho})\big)^2 \, \mathrm{d}z \le \|k_b'\|_2^2 \int \big(\hat{a}(x) - \hat{\varrho}(x)\big)^2 g(x) \, \mathrm{d}x = \mathrm{o}_p\bigg(\frac{1}{nb^3}\bigg).$$

Finally, we write

$$\begin{aligned} T'(z,\hat{\varrho}) &= \frac{1}{n}\sum_{j=1}^n \varepsilon_j \int k_b'\big(z - r(x)\big) D(x) v_c(X_j - x) g(x) \, \mathrm{d}x \\ &= \frac{1}{n}\sum_{j=1}^n \varepsilon_j \int \Big(k_b'\big(z - r(x)\big) - k_b'\big(z - r(X_j)\big)\Big) D(x) v_c(X_j - x) g(x) \, \mathrm{d}x \\ &\quad + \frac{1}{n}\sum_{j=1}^n \varepsilon_j k_b'\big(z - r(X_j)\big) t(X_j). \end{aligned}$$

In view of (8.1), we have the bound

$$\int \big|D(x) v_c(X - x)\big| g(x) \, \mathrm{d}x \le \frac{3}{\eta}\|g\|.$$

This inequality and an application of the Cauchy–Schwarz inequality yield the bound

$$n \int E\big[\big(T'(z,\hat{\varrho})\big)^2\big] \, \mathrm{d}z \le 2\sigma^2 \bigg(\frac{3\|g\|}{\eta} E[U] + \|k_b'\|_2^2 E\big[t^2(X)\big]\bigg)$$

with

$$\begin{aligned} U &= \iint \Big(k_b'\big(z - r(x)\big) - k_b'\big(z - r(X)\big)\Big)^2 \big|D(x) v_c(X - x)\big| g(x) \, \mathrm{d}x \, \mathrm{d}z \\ &\le \|k_b''\|_2^2 \frac{3}{\eta} \int \big(r(X) - r(x)\big)^2 \frac{1}{c} w\bigg(\frac{X - x}{c}\bigg) g(x) \, \mathrm{d}x. \end{aligned}$$

In the last step we used (8.1) and the analog of (8.5) with k_b' in place of k_b. Since r is Lipschitz on [0, 1], we obtain $E[U] = \mathrm{O}(b^{-5}c^2) = \mathrm{o}(b^{-3})$. The above relations show that

$$\|\hat{q}_* - q_b\|_2^2 = \mathrm{O}_p\bigg(\frac{1}{nb^3}\bigg) = \mathrm{o}_p(b). \tag{9.2}$$

The desired (3.4) follows from (9.1) and (9.2).

10. Proof of Theorem 3

It suffices to show that $n^{1/2}\|\hat{C}_i - C_i\| = o_p(1)$ for $i = 1, 2$, with

$$C_1(y) = \frac{1}{n}\sum_{j=1}^{m}\bigl(f'\bigl(y - r(X_j)\bigr) - h'(y)\bigr)\lambda(\varepsilon_j)$$

and $C_2 = C - C_1$. Since the two cases are similar, we prove only the case $i = 1$.

We begin by writing $n^{1/2}C_1 = N * f''$ and $n^{1/2}\hat{C}_1 = \hat{N} * \hat{f}_3''$ where

$$N(z) = N(z, \lambda) = \frac{1}{\sqrt{n}}\sum_{j=1}^{m}\lambda(\varepsilon_j)\bigl(\mathbf{1}\bigl[r(X_j) \le z\bigr] - Q(z)\bigr)$$

and

$$\hat{N}(z) = \hat{N}(z, \hat{\lambda}_2) = \frac{1}{\sqrt{n}}\sum_{j=1}^{m}\hat{\lambda}_2(\hat{\varepsilon}_{2,j})\bigl(\mathbf{1}\bigl[\hat{r}_2(X_j) \le z\bigr] - \hat{Q}(z, \hat{r}_2)\bigr)$$

with

$$\hat{Q}(z, \rho) = \frac{1}{m}\sum_{j=1}^{m}\mathbf{1}\bigl[\rho(X_j) \le z\bigr].$$

In view of $E[\int N^2(z)\,\mathrm{d}z] = E[\lambda^2(\varepsilon)]\int Q(z)(1 - Q(z))\,\mathrm{d}z < \infty$ and the bound

$$n^{1/2}\|\hat{C}_1 - C_1\| \le \|\hat{N} - N\|_2\bigl\|\hat{f}_3''\bigr\|_2 + \|N\|_2\bigl\|\hat{f}_3'' - f''\bigr\|_2$$

it suffices to show

$$\|\hat{N} - N\|_2 = o_p(1) \tag{10.1}$$

and

$$\bigl\|\hat{f}_3'' - f''\bigr\|_2 = o_p(1). \tag{10.2}$$

Let us first prove (10.2). With $\hat{\Delta}_i = \hat{r}_i - r$, we have $\hat{\varepsilon}_{i,j} = \varepsilon_j - \hat{\Delta}_i(X_j)$ for $i = 1, 2$ and $j = 1, \dots, n$. Then we can write

$$\hat{f}_3''(z) - f''(z) = (m/n)D_1(z) + \bigl(1 - (m/n)\bigr)D_2(z)$$

with

$$D_1(z) = \frac{1}{m}\sum_{j=1}^{m}\bigl(\kappa_a''\bigl(z - \varepsilon_j + \hat{\Delta}_2(X_j)\bigr) - f''(z)\bigr),$$

$$D_2(z) = \frac{1}{n-m}\sum_{j=m+1}^{n}\bigl(\kappa_a''\bigl(z - \varepsilon_j + \hat{\Delta}_1(X_j)\bigr) - f''(z)\bigr).$$

Let $\mathbb{E}_2$ denote the conditional expectation given $X_{m+1}, Y_{m+1}, \dots, X_n, Y_n$. Using the square-integrability of f'' and a standard argument, we find that

$$\begin{aligned}
\mathbb{E}_2\left[\int D_1^2(z)\,\mathrm{d}z\right] &\le m^{-1}\int \big(\kappa_a''(z)\big)^2\,\mathrm{d}z \\
&\quad + \iiint \big(f''\big(z-\hat\Delta_2(x)-au\big) - f''(z)\big)^2 \kappa(u)\,\mathrm{d}u\,g(x)\,\mathrm{d}x\,\mathrm{d}z \\
&= \mathrm{O}\big(m^{-1}a^{-5}\big) + \mathrm{o}_p(1).
\end{aligned}$$

Thus, $\|D_1\|_2 = \mathrm{o}_p(1)$. Similarly, one verifies $\|D_2\|_2 = \mathrm{o}_p(1)$, and we obtain (10.2).

To prove (10.1), we set

$$\bar N(z) = \bar N(z, \hat\lambda_2) = \frac{1}{\sqrt{n}}\sum_{j=1}^m \int \hat\lambda_2\big(y - \hat\Delta_2(X_j)\big) f(y)\,\mathrm{d}y\big(\mathbf{1}\big[\hat r_2(X_j) \le z\big] - \hat Q(z, \hat r_2)\big)$$

and shall verify

$$\|\hat N - \bar N - N\|_2 = \mathrm{o}_p(1) \quad \text{and} \quad \|\bar N\|_2 = \mathrm{o}_p(1).$$

We can write

$$\hat N - \bar N - N = \frac{\hat L - \bar L - L}{\hat J} + \left(\frac{1}{\hat J} - \frac{1}{J}\right)L - (\hat M - \bar M - M)$$

with $\hat L(z) = \hat N(z, \hat\ell_2)$, $\bar L(z) = \bar N(z, \hat\ell_2)$, $L(z) = N(z, \ell)$, $\hat M(z) = \hat N(z, \mathrm{id})$, $\bar M(z) = \bar N(z, \mathrm{id})$ and $M(z) = N(z, \mathrm{id})$ where id denotes the identity map on $\mathbb{R}$. Now let $\mathbb{E}$ denote the conditional expectation given $X_1, \dots, X_n, Y_{m+1}, \dots, Y_n$. Then we find

$$\mathbb{E}\big(\|\hat L - \bar L - L\|_2^2\big) \le \frac{1}{n}\sum_{j=1}^m \big(2\Lambda(X_j)R_{1,j} + 2JR_{2,j}\big) \tag{10.3}$$

with

$$\begin{aligned}
\Lambda(x) &= \int \big(\hat\ell_2\big(y - \hat\Delta_2(x)\big) - \ell(y)\big)^2 f(y)\,\mathrm{d}y, \\
R_{1,j} &= \int \big(\mathbf{1}\big[\hat r_2(X_j) \le z\big] - \hat Q(z, \hat r_2)\big)^2\,\mathrm{d}z, \\
R_{2,j} &= \int \big(\mathbf{1}\big[\hat r_2(X_j) \le z\big] - \hat Q(z, \hat r_2) - \mathbf{1}\big[r(X_j) \le z\big] + Q(z)\big)^2\,\mathrm{d}z.
\end{aligned}$$

By the properties of the quadratic smoother, we have

$$\frac{1}{n}\sum_{j=1}^n \hat\Delta_2^2(X_j) = \mathrm{O}_p\big(n^{-3/4}\big) \quad \text{and thus} \quad \frac{1}{n}\sum_{j=1}^n \big|\hat\Delta_2(X_j)\big| = \mathrm{O}_p\big(n^{-3/8}\big). \tag{10.4}$$

Several applications of the Cauchy–Schwarz inequality yield the bound

$$\frac{1}{n}\sum_{j=1}^{m} R_{2,j} \le \left(\frac{3}{n}+\frac{3}{m}\right)\sum_{j=1}^{m}\int \big(\mathbf{1}\big[\hat{r}_2(X_j)\le z\big]-\mathbf{1}\big[r(X_j)\le z\big]\big)^2\,\mathrm{d}z + 3\int \big(\hat{Q}(z,r)-Q(z)\big)^2\,\mathrm{d}z.$$

Now we use the identity $(\mathbf{1}[u\le z]-\mathbf{1}[v\le z])^2=\mathbf{1}[u<z\le v]$, valid for $u\le v$, and (10.4), to conclude

$$\frac{1}{n}\sum_{j=1}^{m} R_{2,j} \le \frac{6}{m}\sum_{j=1}^{m}\big|\hat{\Delta}_2(X_j)\big| + \mathrm{O}_p\big(n^{-1/2}\big)=\mathrm{O}_p\big(n^{-3/8}\big). \tag{10.5}$$

Using the above identity and the uniform consistency of $\hat{r}_2$, we obtain

$$\max_{1\le j\le m} R_{1,j} \le \max_{1\le j\le m}\frac{1}{m}\sum_{i=1}^{m}\int \big(\mathbf{1}\big[\hat{r}_2(X_j)\le z\big]-\mathbf{1}\big[\hat{r}_2(X_i)\le z\big]\big)^2\,\mathrm{d}z=\mathrm{O}_p(1). \tag{10.6}$$

By Lemma 10.1 in Schick [16] there is a constant c_* so that

$$\frac{1}{n}\sum_{j=1}^{m}\int \big(\hat{\ell}_2\big(y-\hat{\Delta}_2(X_j)\big)-\hat{\ell}_2(y)\big)^2 f(y)\,\mathrm{d}y \le \frac{c_*}{a^4 n}\sum_{j=1}^{m}\hat{\Delta}_2^2(X_j), \tag{10.7}$$

$$\int \big(\hat{\ell}_2(y)-\ell(y)\big)^2 f(y)\,\mathrm{d}y \le \frac{c_*}{a^6 m}\sum_{j=m+1}^{n}\hat{\Delta}_2^2(X_j) + \mathrm{O}_p\left(\frac{1}{a^6 m}\right)+\mathrm{o}_p(1). \tag{10.8}$$

From (10.3)–(10.8) and $a^8 n\to\infty$, we obtain $\|\hat{L}-\bar{L}-L\|_2=\mathrm{o}_p(1)$. A similar argument yields $\|\hat{M}-\bar{M}-M\|_2=\mathrm{o}_p(1)$. Using (10.7), (10.8) and the operator $\mathbb{E}$, we obtain

$$\frac{1}{n}\sum_{j=1}^{m}\big(\hat{\ell}_2(\hat{\varepsilon}_{2,j})-\ell(\varepsilon_j)\big)^2=\mathrm{o}_p(1).$$

It is now easy to see that $\hat{J}$ is a consistent estimator of J. This completes the proof of $\|\hat{N}-\bar{N}-N\|_2=\mathrm{o}_p(1)$.

We are left to verify $\|\bar{N}\|_2=\mathrm{o}_p(1)$. Using the definition of $\hat{Q}(z,\hat{r}_2)$, we can write

$$\bar{N}(z)=\frac{1}{\sqrt{n}}\sum_{j=1}^{m}\big(\mathbf{1}\big[\hat{r}_2(X_j)\le z\big]-\hat{Q}(z,\hat{r}_2)\big)\left(\hat{\Delta}_2(X_j)+\frac{1}{\hat{J}}\hat{\omega}(X_j)\right),$$

where

$$\hat{\omega}(X_j) = \int \bigl(\hat{\ell}_2\bigl(y - \hat{\Delta}_2(X_j)\bigr) - \hat{\ell}_2(y)\bigr) f(y)\,\mathrm{d}y = \int \hat{\ell}_2(y)\bigl(f\bigl(y + \hat{\Delta}_2(X_j)\bigr) - f(y)\bigr)\,\mathrm{d}y.$$

A Taylor expansion yields

$$f\bigl(y + \hat{\Delta}_2(X_j)\bigr) - f(y) - \hat{\Delta}_2(X_j) f'(y) = \hat{\Delta}_2^2(X_j) \int_0^1 (1-s) f''\bigl(y + s\hat{\Delta}_2(X_j)\bigr)\,\mathrm{d}s.$$

Since $\hat{\ell}_2$ is bounded by c_*/a, we obtain

$$\bigl|\hat{\omega}(X_j) + \hat{\Delta}_2(X_j)\hat{J}_2\bigr| \le \frac{c_*}{a}\hat{\Delta}_2^2(X_j) \int \bigl|f''(y)\bigr|\,\mathrm{d}y$$

with $\hat{J}_2 = \int \hat{\ell}_2(y)\ell(y) f(y)\,\mathrm{d}y = J + \mathrm{o}_p(1)$. Now set

$$\hat{\Upsilon}(z) = \frac{1}{\sqrt{n}} \sum_{j=1}^m \bigl(\mathbf{1}\bigl[\hat{r}_2(X_j) \le z\bigr] - \hat{Q}(z, \hat{r}_2)\bigr)\hat{\Delta}_2(X_j),$$

$$\Upsilon(z) = \frac{1}{\sqrt{n}} \sum_{j=1}^m \bigl(\mathbf{1}\bigl[r(X_j) \le z\bigr] - Q(z)\bigr)\hat{\Delta}_2(X_j).$$

Using the Minkowski inequality and the statements (10.4)–(10.6), we derive

$$\bigl\|\bar{N} - (1 - \hat{J}_2/\hat{J})\hat{\Upsilon}\bigr\|_2 \le \frac{c_*\|f''\|_1}{a\sqrt{n}} \sum_{j=1}^m R_{1,j}^{1/2}\hat{\Delta}_2^2(X_j) = \mathrm{O}_p\bigl(a^{-1}n^{-1/4}\bigr) = \mathrm{o}_p(1),$$

$$\|\hat{\Upsilon} - \Upsilon\|_2 \le \frac{1}{\sqrt{n}} \sum_{j=1}^m R_{2,j}^{1/2}\bigl|\hat{\Delta}_2(X_j)\bigr| \le n^{1/2}\Biggl(\frac{1}{n}\sum_{j=1}^m R_{2,j}\frac{1}{n}\sum_{j=1}^m \hat{\Delta}_2^2(X_j)\Biggr)^{1/2} = \mathrm{o}_p(1).$$

Using the inequality $|\mathbf{1}[r(x) \le z] - Q(z)| \le \mathbf{1}[r(0) \le z \le r(1)]$, valid for all $0 \le x \le 1$ and $z \in \mathbb{R}$, we obtain

$$\begin{aligned}\mathbb{E}_2\bigl[\bigl\|\Upsilon - \mathbb{E}_2[\Upsilon]\bigr\|_2^2\bigr] &\le \frac{m}{n} \iint \bigl(\mathbf{1}\bigl[r(x) \le z\bigr] - Q(z)\bigr)^2 \hat{\Delta}_2^2(x) g(x)\,\mathrm{d}x\,\mathrm{d}z \\ &\le \bigl(r(1) - r(0)\bigr) \int \hat{\Delta}_2^2(x) g(x)\,\mathrm{d}x = \mathrm{o}_p(1).\end{aligned}$$

Now introduce

$$I(z, \rho) = \int \bigl(\mathbf{1}\bigl[r(x) \le z\bigr] - Q(z)\bigr)\rho(x) g(x)\,\mathrm{d}x.$$

Then we have $\mathbb{E}_2[\Upsilon(z)] = n^{-1/2} m I(z, \hat{\Delta}_2)$. In view of the above and $1 - \hat{J}_2/\hat{J} = \mathrm{o}_p(1)$, the desired property $\|\bar{N}\|_2 = \mathrm{o}_p(1)$ will follow if we show $\|I(\cdot, \hat{\Delta}_2)\|_2 = \mathrm{O}_p(n^{-1/2})$. The latter is equivalent to showing $\|I(\cdot, \hat{r} - r)\|_2 = \mathrm{O}_p(n^{-1/2})$. In view of (5.3), we have

$$\big\|I(\cdot, \hat{r} - r) - I(\cdot, \hat{\varrho})\big\|_2 = \big\|I(\cdot, \hat{r} - r - \hat{\varrho})\big\|_2 = \mathrm{o}_p\big(n^{-1/2}\big).$$

We can express $I(z, \hat{\varrho})$ as the average

$$\frac{1}{n}\sum_{j=1}^{n} \varepsilon_j \tau(z, X_j)$$

with

$$\begin{aligned} \tau(z, X_j) &= \int \big(\mathbf{1}\big[r(x) \le z\big] - Q(z)\big)\frac{1}{c} w\bigg(\frac{X_j - x}{c}\bigg) D(x)\psi\bigg(\frac{X_j - x}{c}\bigg) g(x)\,\mathrm{d}x \\ &= \int \big(\mathbf{1}\big[r(X_j - cu) \le z\big] - Q(z)\big) w(u) D(X_j - cu)\psi(u) g(X_j - cu)\,\mathrm{d}u. \end{aligned}$$

Since $|\tau(z, X_j)|$ is bounded by a constant times $\mathbf{1}[r(0) \le z \le r(1)]$, we conclude

$$nE\big[\big\|I(\cdot, \hat{\varrho})\big\|_2^2\big] = \int \sigma^2 E\big[\tau^2(z, X)\big]\,\mathrm{d}z = \mathrm{O}(1).$$

The above shows that $\|I(\cdot, \hat{r} - r)\|_2 = \mathrm{O}_p(n^{-1/2})$, and the proof is finished.

Acknowledgements

We thank the referee for suggesting to discuss the question of efficiency. This resulted in adding the present Sections 4 and 10. The research of Anton Schick was supported in part by NSF Grant DMS 09-06551.

References

[1] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer. Reprint of the 1993 original. MR1623559

[2] Bickel, P.J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393. MR1065550

[3] Chaudhuri, P., Doksum, K. and Samarov, A. (1997). On average derivative quantile regression. *Ann. Statist.* **25** 715–744. MR1439320

[4] Efromovich, S. and Samarov, A. (2000). Adaptive estimation of the integral of squared regression derivatives. *Scand. J. Statist.* **27** 335–351. MR1777508

[5] Escanciano, J.C. and Jacho-Chávez, D.T. (2012). $\sqrt{n}$-uniformly consistent density estimation in nonparametric regression models. *J. Econometrics* **167** 305–316. MR2892077

[6] Frees, E.W. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.* **89** 517–525. MR1294078

[7] Giné, E. and Mason, D.M. (2007). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.* **35** 1105–1145. MR2341700

[8] Laurent, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** 659–681. MR1394981

[9] Müller, U.U. (2012). Estimating the density of a possibly missing response variable in nonlinear regression. *J. Statist. Plann. Inference* **142** 1198–1214. MR2879764

[10] Müller, U.U., Schick, A. and Wefelmeyer, W. (2009). Estimating the error distribution function in nonparametric regression with multivariate covariates. *Statist. Probab. Lett.* **79** 957–964. MR2509488

[11] Müller, U.U., Schick, A. and Wefelmeyer, W. (2013). Non-standard behavior of density estimators for functions of independent observations. *Comm. Statist. Theory Methods* **42** 2991–3000.

[12] Nickl, R. (2007). Donsker-type theorems for nonparametric maximum likelihood estimators. *Probab. Theory Related Fields* **138** 411–449. Erratum: *Probab. Theory Related Fields* **141** (2008) 331–332. MR2299714

[13] Nickl, R. (2009). On convergence and convolutions of random signed measures. *J. Theoret. Probab.* **22** 38–56. MR2472004

[14] Saavedra, Á. and Cao, R. (1999). Rate of convergence of a convolution-type estimator of the marginal density of an MA(1) process. *Stochastic Process. Appl.* **80** 129–155. MR1682259

[15] Saavedra, A. and Cao, R. (2000). On the estimation of the marginal density of a moving average process. *Canad. J. Statist.* **28** 799–815. MR1821435

[16] Schick, A. (1993). On efficient estimation in regression models. *Ann. Statist.* **21** 1486–1521. Correction and Addendum: *Ann. Statist.* **23** (1995) 1862–1863. MR1241276

[17] Schick, A. and Wefelmeyer, W. (2004). Functional convergence and optimality of plug-in estimators for stationary densities of moving average processes. *Bernoulli* **10** 889–917. MR2093616

[18] Schick, A. and Wefelmeyer, W. (2004). Root n consistent and optimal density estimators for moving average processes. *Scand. J. Statist.* **31** 63–78. MR2042599

[19] Schick, A. and Wefelmeyer, W. (2004). Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Stat.* **16** 925–935. MR2094747

[20] Schick, A. and Wefelmeyer, W. (2007). Root-n consistent density estimators of convolutions in weighted L_1-norms. *J. Statist. Plann. Inference* **137** 1765–1774. MR2323861

[21] Schick, A. and Wefelmeyer, W. (2007). Uniformly root-n consistent density estimators for weakly dependent invertible linear processes. *Ann. Statist.* **35** 815–843. MR2336870

[22] Schick, A. and Wefelmeyer, W. (2008). Root-n consistency in weighted L_1-spaces for density estimators of invertible linear processes. *Stat. Inference Stoch. Process.* **11** 281–310. MR2438498

[23] Schick, A. and Wefelmeyer, W. (2009). Convergence rates of density estimators for sums of powers of observations. *Metrika* **69** 249–264. MR2481923

[24] Schick, A. and Wefelmeyer, W. (2009). Non-standard behavior of density estimators for sums of squared observations. *Statist. Decisions* **27** 55–73. MR2597426

[25] Støve, B. and Tjøstheim, D. (2012). A convolution estimator for the density of nonlinear regression observations. *Scand. J. Statist.* **39** 282–304. MR2927026

Received December 2011 and revised April 2012