

ROBUST NONPARAMETRIC REGRESSION WITH SIMULTANEOUS SCALE CURVE ESTIMATION¹

BY W. HÄRDLE AND A. B. TSYBAKOV

Universität Bonn and Academy of Sciences of the USSR

Let $\{X_i, Y_i\}_{i=1}^n \subset \mathbf{R}^d \times \mathbf{R}$ be independent identically distributed random variables. If the conditional distribution $F(y|x)$ can be parametrized by $F(y|x) = F_0((y - m(x))/\sigma(x))$ with a fixed and known distribution F_0 , the regression curve $m(x)$ and scale curve $\sigma(x)$ could be estimated by some parametric method. More generally, we assume that F is unknown and consider nonparametric simultaneous M -type estimates of the unknown functions $m(x)$ and $\sigma(x)$, using kernel estimators for the conditional distribution function $F(y|x)$. We show pointwise consistency and asymptotic normality of these estimates. The rate of convergence is optimal in the sense of Stone (1980). The asymptotic bias term of this robust estimate turns out to be the same as for the linear Nadaraya-Watson kernel estimate.

1. Introduction. Let $(X_1, Y_1), (X_2, Y_2), \dots$ be a sequence of independent identically distributed $(d + 1)$ -dimensional random vectors. Assume that the conditional distribution $P\{Y_1 \leq y | X_1 = x\} = F(y|x)$ has the form $F(y|x) = F_0((y - m(x))/(\sigma(x)))$ with a fixed (but unknown) distribution function F_0 . Call $m(\cdot)$ the regression curve and $\sigma(\cdot)$ the scale curve and assume that they are continuous functions on a set $\Xi \subset \mathbf{R}^d$. Our goal is the simultaneous and nonparametric estimation of the regression curve $m(\cdot)$ and the scale curve $\sigma(\cdot)$ from a random sample $(X_1, Y_1), \dots, (X_n, Y_n)$.

There exists a tradition of nonparametric regression [Nadaraya (1964), Watson (1964)], where $m(x)$ is viewed as an expression for the conditional expectation $E(Y|X = x)$ and this $m(x)$ is estimated by a weighted average of the response variables Y . Mild conditions on the distribution of the Y -variables and on the weights ensure convergence of the estimators to the conditional expectation $E(Y|X = x)$, as Stone (1977) has shown. In the discussion to Stone's paper, Brillinger raised the point that a nonlinear M -type estimate of the regression curve might be worthwhile to study in order to achieve desirable robustness properties.

In this paper we consider more generally simultaneous nonparametric estimation of $m(x)$ and $\sigma(x)$ by M -type smoothers. Our approach is closely related to simultaneous M -estimation of location and scale; see Huber [(1981), Chapter 6.4]. Our approach differs in that we have to consider additional bias terms, due to the fact that $m(\cdot)$ and $\sigma(\cdot)$ are unspecified functions and $F(y|x)$ is estimated by the nonparametric kernel method. The simultaneous M -type smoothers of the

Received April 1985; revised May 1987.

¹Research supported by Deutsche Forschungsgemeinschaft, Sonderforschungsbereiche 303 and 123.

AMS 1980 subject classification. 62G05.

Key words and phrases. Robust curve estimation, M -estimation, nonparametric regression, joint estimation of regression and scale curve, optimal rate of convergence.

regression curve and of the scale curve are determined by a system of nonlinear equations. Define for $s \in \mathbb{R}^+$, $t \in \mathbb{R}$, $x \in \Xi$,

$$(1.1) \quad T_1(s, t) = \int \psi\left(\frac{y-t}{s}\right) dF(y|x)$$

and

$$(1.2) \quad T_2(s, t) = \int \chi\left(\frac{y-t}{s}\right) dF(y|x),$$

with ψ and χ some bounded real functions satisfying additional properties to be stated later. We generalize the preceding notion about $m(x)$ and $\sigma(x)$ by assuming that the curves $m(x)$ and $\sigma(x)$ can be represented as simultaneous zeros of T_1 and T_2 , i.e., $T_1(\sigma(x), m(x)) = T_2(\sigma(x), m(x)) = 0$.

The unknown conditional distribution $F(y|x)$ is estimated by the kernel method,

$$F_n(y|x) = \sum_{i=1}^n W_{ni}(x; X_1, \dots, X_n) I(Y_i \leq y).$$

Here $\{W_{ni}\}_{i=1}^n$ denotes a sequence of weights

$$W_{ni}(x; X_1, \dots, X_n) = \frac{K((X_i - x)/h)}{\sum_{j=1}^n K((X_j - x)/h)}$$

with kernel $K: \mathbb{R}^d \rightarrow \mathbb{R}$ and bandwidth sequence $h = h_n \in \mathbb{R}^+$. In analogy to (1.1) and (1.2) the nonparametric estimates $(m_n(x), \sigma_n(x))$ are defined as simultaneous zeros of

$$(1.3) \quad T_{1n}(s, t) = \int \psi\left(\frac{y-t}{s}\right) dF_n(y|x)$$

and

$$(1.4) \quad T_{2n}(s, t) = \int \chi\left(\frac{y-t}{s}\right) dF_n(y|x).$$

Such simultaneous zeros exist as is shown in Theorem 1. Under regularity conditions on the kernel and the functions ψ and χ , we prove strong consistency of $(m_n(x), \sigma_n(x))$ as well as the asymptotic normality of

$$\sqrt{nh^d} \left[\begin{pmatrix} m_n(x) \\ \sigma_n(x) \end{pmatrix} - \begin{pmatrix} m(x) \\ \sigma(x) \end{pmatrix} \right].$$

Numerous examples of functions ψ and χ for the simultaneous M -estimation of location and scatter can be found in the literature on robust estimation. For instance, the well known

$$\begin{aligned} \psi(u) &= -k \vee (k \wedge u), & k > 0, \\ \chi(u) &= c^2 \wedge u^2 - \beta, & 0 < \beta < c^2, \end{aligned}$$

fulfill our assumptions for suitable β [Assumption (A1)]; see Huber [(1981), page

137]. Note that in the case $c = k = \infty$ and $\beta = \int u^2 dF_0(u)$, this class of functions ψ and χ give the Nadaraya–Watson kernel estimate and the natural estimate

$$\sigma(x) = \left[n^{-1} \sum_{i=1}^n W_{ni}(x) (Y_i - m_n(x))^2 \right]^{1/2}$$

for the conditional scale $\sigma(x)$.

The estimation of $m(x)$ alone by M -type estimators has been investigated by several authors. Tsybakov (1982a, b) and Härdle (1984) showed consistency and asymptotic normality. Some Monte Carlo results for kernel “ M -smoothers” are presented in Härdle and Gasser (1984). A recursive M -type regression function estimator was considered by Tsybakov (1982a, b). An M -type smoothing spline was considered by Huber (1979), Cox (1983) and Silverman (1985). An M -type estimation on functional classes was investigated by Nemirovskii, Polyak and Tsybakov (1983).

The results of this paper are relevant for several applications. For instance, in physical chemistry the Raman spectra estimation instrumental noise is considerably reduced by the robust estimator $m_n(x)$; see Bussian and Härdle (1984). In image processing Justusson (1981) applied two-dimensional running medians to image restoration from noisy signals. Hildenbrand and Hildenbrand [(1986), Figure 7] report aberrant observations in an analysis of expenditure curves for potatoes as a function of (normalized) income and use a robust two stage estimation technique. Also in the a posteriori construction of parametric models following a previous nonparametric analysis, a robust nonparametric estimator seems to be desirable. Outliers might mimic nonexistent structure resulting in a biased parametric model.

It has been conjectured that robust smoothers are inclining to oversmooth the data by chopping off existing peaks of the regression curve which finally would result in an increased bias. It turns out (Theorem 2) that this conjecture is not true: The “ M -smoothers” considered here have the same asymptotic bias as their linear relatives such as the Nadaraya–Watson estimator $\int y dF_n(y|x)$. Our representation of $(m(x), \sigma(x))$ as zeros of certain functionals of the conditional distribution $F(y|x)$ introduces a slightly more general class of regression curves than the conditional expectation curve of Y or X . We may also note that even when outliers are absent it is reasonable to complement the nonparametric regression estimate m_n by a suitable estimate of its accuracy σ_n . This was not commonly realized in earlier work on nonparametric regression. In the setting of parametric linear regression, however, robust estimation from heteroscedastic data has been considered by Carroll (1982) via construction of a (linear) nonparametric estimate of the scale curve $\sigma(x)$. There are some open questions. In this paper we do not consider the choice of the bandwidth $h = h_n$ that has to be made in practice. A cross-validatory choice for the Nadaraya–Watson estimator has been proposed by Härdle and Marron (1985). In a forthcoming paper we will present an adaptive bandwidth selection rule that minimizes the maximal risk over specific classes of regression curves. Also the functions ψ and χ have to be

chosen in practice. Our result on the asymptotic normality of (m_n, σ_n) suggests that, as in the classical M -estimation of location and scale, there are estimators that minimize the maximal asymptotic variance over a certain class of distributions. Is it possible to adapt ψ and χ to the underlying F_0 in order to achieve asymptotic efficiency?

2. Simultaneous M -smoothing of regression and scale curve. The following regularity conditions on ψ and χ are needed to ensure consistency of the estimates.

- (A1) The distribution function F_0 is continuous and symmetric. Further every nonempty neighborhood of zero has nonnull F_0 -measure, and $\int \chi(u) dF_0(u) = 0$.
- (A2) The function $\psi(t)$ is continuous, nondecreasing, bounded and odd.
- (A3) The function $\chi(t)$ is continuous, bounded and even, nondecreasing for $t \geq 0$ and strictly increasing in the interval, where $\chi(t) < \chi(\infty)$.
- (A4) The functions $t^{-1}\psi(t)$ and $t^{-2}(\chi(t) - \chi(0))$ are continuous and nonincreasing for $t \geq 0$.
- (A5) There exists a constant $t_0 > 0$ such that $\chi(t_0) > 0$ and $t^{-1}\psi(t) > 0$ for $t \leq t_0$.

The next two conditions specify the class of kernels K and regulate the speed of the bandwidth sequence.

- (A6) The kernel $K: \mathbb{R}^d \rightarrow \mathbb{R}$ is bounded and nonnegative with bounded support and $\int K(u) du \neq 0$.
- (A7) The sequence of bandwidths $h = h_n$ tends to zero such that

$$\begin{aligned} \text{(a)} \quad & nh^d \rightarrow \infty \\ \text{or} \\ \text{(b)} \quad & nh^d / \log n \rightarrow \infty. \end{aligned}$$

Assumption (A7a) is necessary to obtain convergence in probability whereas (A7b) is used to show almost sure convergence. Such conditions on the rate of convergence of h_n are compatible to other smoothing techniques; see the survey article of Collomb (1981). Finally we postulate continuity of the marginal density $f(\cdot)$ of X , the regression curve and the scale curve.

- (A8) The density $f(\cdot)$ of X is continuous and positive in some neighborhood of x .
- (A9) The functions $m(\cdot)$ and $\sigma(\cdot)$ are continuous in some neighborhood of x , and $\sigma(x) > 0$.

THEOREM 1. *Let (A1)–(A9) be satisfied. Then*

- (i) $(m(x), \sigma(x))$ are unique simultaneous zeros of (1.1) and (1.2);
- (ii) there exist simultaneous zeros $(m_n(x), \sigma_n(x))$ of (1.3) and (1.4) with probability tending to 1 as $n \rightarrow \infty$ (a.s. for n sufficiently large) if (A7a) [respectively, (A7b)] holds;

(iii) for any simultaneous zeros $(m_n(x), \sigma_n(x))$ of (1.3) and (1.4),

$$(m_n(x), \sigma_n(x)) \rightarrow (m(x), \sigma(x)), \quad n \rightarrow \infty,$$

in probability (almost surely) if (A7a) [respectively, (A7b)] holds.

The next conditions are refinements of the preceding assumptions and are used to show the asymptotic normality of (m_n, σ_n) .

(A10) The functions ψ and χ are continuously differentiable with bounded derivative and $t\psi'(t)$ and $t\chi'(t)$ are continuous and bounded.

$$(A11) \quad 0 < \varphi_0 = \int \psi'(u) dF_0(u),$$

$$0 < \kappa_0 = \int u\chi'(u) dF_0(u).$$

Note that (A10) implies that the preceding two integrals are finite.

(A12) The functions m and σ are Lipschitz continuous with Lipschitz constants L, L' , respectively. The directional derivatives

$$m'(x; u) = \lim_{\varepsilon \rightarrow 0} \varepsilon^{-1}(m(x + \varepsilon u) - m(x))$$

[and similarly for $\sigma(x)$] exist for all $u \in \mathbb{R}^d$.

Assumption (A12) appears to be the minimal smoothness assumption under which the asymptotic normality may yet be expected. Using the argument of Stone (1980) one can show that under (A12) the squared error optimal pointwise rate of convergence of (m_n, σ_n) to (m, σ) is attained for $h_n \sim n^{-1/(d+2)}$. This is the bandwidth rate for which the squared bias and the variance of the estimate are asymptotically of the same order. Therefore it is reasonable to assume:

(A13) There is a constant $0 \leq \beta < \infty$ such that

$$\lim_{n \rightarrow \infty} h_n n^{1/(d+2)} = \beta.$$

Note that $\beta \neq 0$ corresponds to the optimal rate $\{n^{-1/(d+2)}\}$; see Stone (1980). In the case $\beta = 0$ the bias is of smaller order than the variance. The case $\beta = \infty$ is not considered. In this case the asymptotic variance of (m_n, σ_n) is negligible compared to the bias. The convergence rate of (m_n, σ_n) could be improved by so-called higher order kernels at the expense of assuming higher differentiability of (m, σ) [Härdle and Marron (1985)]. For instance, if m and σ are twice continuously differentiable and a smooth symmetric kernel is used, the rate $\{n^{2/(4+d)}\}$ can be achieved. Indeed, a second order Taylor expansion in (5.9) would result in the rate h_n^{2d} for the bias. Setting $h_n \sim n^{-1/(4+d)}$ yields the faster rate $\{n^{-2/(4+d)}\}$ for (m_n, σ_n) to (m, σ) .

THEOREM 2. Let (A1)–(A13) be satisfied and define $\varphi_2 = \int \psi^2(u) dF_0(u)$ and $\kappa_2 = \int \chi^2(u) dF_0(u)$. Then, as $n \rightarrow \infty$,

$$\sqrt{nh^d} \left[\begin{pmatrix} m_n(x) \\ \sigma_n(x) \end{pmatrix} - \begin{pmatrix} m(x) \\ \sigma(x) \end{pmatrix} \right]$$

is asymptotically normally distributed with mean

$$\beta^{d/2+1} \left(\frac{\int m'(x; u)K(u) du}{\int \sigma'(x; u)K(u) du} \right) / \int K(u) du$$

and covariance matrix

$$\frac{\sigma^2(x) \int K^2(u) du}{f(x) (\int K(u) du)^2} \begin{pmatrix} \varphi_2/\varphi_0^2 & 0 \\ 0 & \kappa_2/\kappa_0^2 \end{pmatrix}.$$

COROLLARY 1. If $\beta \neq 0$, then as $n \rightarrow \infty$,

$$n^{1/(d+2)}(m_n(x) - m(x)) \rightarrow_{\mathcal{D}} N(b_m, V_m),$$

where

$$b_m = \beta \int m'(x; u)K(u) du / \int K(u) du,$$

$$V_m = \frac{\sigma^2(x)}{\beta^d f(x)} \frac{\varphi_2}{\varphi_0^2} \int K^2(u) du / \left(\int K(u) du \right)^2.$$

Also,

$$n^{1/(d+2)}(\sigma_n(x) - \sigma(x)) \rightarrow_{\mathcal{D}} N(b_\sigma, V_\sigma),$$

where

$$b_\sigma = \beta \int \sigma'(x; u)K(u) du / \int K(u) du,$$

$$V_\sigma = \frac{\sigma^2(x)}{\beta^d f(x)} \frac{\kappa_2}{\kappa_0^2} \int K^2(u) du / \left(\int K(u) du \right)^2.$$

3. Preliminary lemmas.

LEMMA 1. Let $\{Q_n(t)\}$ be a sequence of bounded nondecreasing random functions defined on the closed interval $U \subseteq \mathbb{R}$. Suppose that $Q(t)$ is a continuous nondecreasing bounded function on U . Assume:

1. $Q_n(t) \rightarrow Q(t)$, $n \rightarrow \infty$, a.s. (in probability) $\forall t \in U$.
2. If the right endpoint of U is $+\infty$, then

$$\lim_{t \rightarrow \infty} Q_n(t) = \lim_{t \rightarrow \infty} Q(t), \quad \forall n,$$

and if the left endpoint of U is $-\infty$, then

$$\lim_{t \rightarrow -\infty} Q_n(t) = \lim_{t \rightarrow -\infty} Q(t), \quad \forall n.$$

Then

$$\sup_{t \in U} |Q_n(t) - Q(t)| \rightarrow 0, \quad n \rightarrow \infty,$$

a.s. (in probability, respectively).

The proof of Lemma 1 is obtained by the same argument as for the Glivenko–Cantelli theorem.

LEMMA 2. Let F_0 be continuous and let conditions (A6)–(A9) be satisfied. Then

$$\sup_{y \in \mathbb{R}} |F_n(y|x) - F(y|x)| \rightarrow 0, \quad n \rightarrow \infty,$$

in probability (almost surely) if (A7a) [respectively, (A7b)] holds.

PROOF. From Collomb [(1980), Proposition 1 (2)], it follows that $F_n(y|x) \rightarrow F(y|x)$, $n \rightarrow \infty$, $\forall y \in \mathbb{R}$, in probability (almost surely) if (A7a) [respectively, (A7b)] holds. \square

Now Lemma 1 is applied with $Q_n(t) = F_n(t|x)$ and $Q(t) = F(t|x)$ to yield uniform convergence of conditional functions.

LEMMA 3. Let $Q(y, t)$ be continuous in (y, t) and a bounded function of $y \in \mathbb{R}$, $t \in T$, T a compact set in \mathbb{R}^d . If

$$(3.1) \quad \int \varphi(y) F_n(dy|x) \rightarrow \int \varphi(y) F(dy|x), \quad n \rightarrow \infty,$$

a.s. (in probability) for any bounded continuous function φ , then

$$\sup_{t \in T} \left| \int Q(y, t) F_n(dy|x) - \int Q(y, t) F(dy|x) \right| \rightarrow 0,$$

$n \rightarrow \infty$, a.s. (in probability).

PROOF. Consider for brevity only the a.s. case. Let N be a minimal ε -net on T in the Euclidean metric. Let

$$V_n(t) = \int Q(y, t) F_n(dy|x), \quad V(t) = \int Q(y, t) F(dy|x).$$

Then,

$$(3.2) \quad \begin{aligned} \sup_{t \in T} |V_n(t) - V(t)| &\leq \max_{\tilde{t} \in N} |V_n(\tilde{t}) - V(\tilde{t})| \\ &+ \max_{\tilde{t} \in N} \sup_{t: |t - \tilde{t}| \leq \varepsilon} |V_n(t) - V_n(\tilde{t})| \\ &+ \max_{\tilde{t} \in N} \sup_{t: |t - \tilde{t}| \leq \varepsilon} |V(t) - V(\tilde{t})|. \end{aligned}$$

In (3.2), let $n \rightarrow \infty$ and then $\varepsilon \rightarrow 0$. The first summand in (3.2) tends to 0 a.s. as $n \rightarrow \infty$ since (3.1) holds and since $\text{card } N = N(\varepsilon) < \infty$. The third summand tends to 0 as $\varepsilon \rightarrow 0$ by continuity of $V(t)$ on T .

It remains to prove that the second summand tends to 0.

Let

$$\varphi_\varepsilon(y) = \sup_{t, \tilde{t} \in T: |t - \tilde{t}| \leq \varepsilon} |Q(y, t) - Q(y, \tilde{t})|.$$

Then

$$\begin{aligned} \max_{\tilde{t} \in N} \sup_{t: |t-\tilde{t}| \leq \varepsilon} |V_n(t) - V_n(\tilde{t})| &\leq \sup_{t, \tilde{t} \in T: |t-\tilde{t}| \leq \varepsilon} \int |Q(y, t) - Q(y, \tilde{t})| F_n(dy|x) \\ &\leq \int \varphi_\varepsilon(y) F_n(dy|x). \end{aligned}$$

Since Q is continuous in (y, t) then φ_ε is continuous in y . Therefore (3.1) yields

$$(3.3) \quad \limsup_n \max_{\tilde{t} \in N} \sup_{t: |t-\tilde{t}| \leq \varepsilon} |V_n(t) - V_n(\tilde{t})| \leq \int \varphi_\varepsilon(y) F(dy|x).$$

But $\lim_{\varepsilon \rightarrow 0} \varphi_\varepsilon(y) = 0, \forall y$, because Q is continuous in (y, t) . In view of boundedness of φ_ε the right side of (3.3) tends to 0 as $\varepsilon \rightarrow 0$. This completes the proof. \square

4. Proof of Theorem 1. Without loss of generality assume that $m(x) = 0$ and $\sigma(x) = 1$. The assertion (i) of Theorem 1 is deduced from the following lemma.

LEMMA 4.

(4.1) For each t there exists a unique solution $s^*(t)$ of $T_2(s^*(t), t) = 0$.

(4.2) $s^*(t)$ is a continuous function and $\inf_t s^*(t) > 0$.

(4.3) For each $s, T_1(s, t) = 0$ if and only if $t = 0$.

PROOF. The assertions (4.1) and (4.2) are contained in Theorem 1 and Lemma 2 of Maronna (1976). The "if" part of (4.3) follows from the fact that F_0 is symmetric and ψ is odd. The "only if" part of (4.3) follows from monotonicity of ψ and (4.10). (Set $t = \pm \varepsilon$ there to prove by contradiction.) \square

We shall prove the assertions (ii) and (iii) of Theorem 1 in the case when (A7b) holds [the case (A7a) is considered in a similar way].

By (A2) and (A5) the function ψ is monotone and $\psi(\infty) > 0$ and $\psi(-\infty) < 0$. Hence there exists a solution $t_n(s)$ of

$$(4.4) \quad T_{1n}(s, t_n(s)) = 0, \quad \forall s > 0.$$

From Lemma 2 and continuity of F_0 it follows that F_n satisfies condition (E) of Maronna (1976), a.s. for large n . It is easy to verify that conditions (A2)–(A5) coincide with the univariate version of conditions (A)–(D) of Maronna (1976). Therefore we can apply Theorem 2 of Maronna (1976), which yields the assertion (ii) of Theorem 1. In addition there exist some constants $a, A, 0 < a \leq A < \infty$ such that

$$(4.5) \quad a \leq \sigma_n \leq A, \text{ a.s., for } n \text{ sufficiently large.}$$

This follows in the same manner as (5.1) in Maronna [(1976), page 59] (use Lemma 2 instead of the Glivenko–Cantelli theorem there).

LEMMA 5. For any sequence of functions $\{t_n\}$ satisfying (4.4),

$$(4.6) \quad \sup_{a \leq s \leq A} |t_n(s)| \rightarrow 0, \quad \text{a.s., } n \rightarrow \infty.$$

PROOF. Note that for fixed $s_0 > 0$ the function $T_1(s_0, t)$ is nonincreasing in t . Therefore, if for some constants a, A and arbitrarily small $\varepsilon > 0$,

$$(4.7) \quad \inf_{a \leq s \leq A} T_1(s, -\varepsilon) > 0,$$

$$(4.8) \quad \sup_{a \leq s \leq A} T_1(s, +\varepsilon) < 0$$

and

$$(4.9) \quad \sup_{a \leq s \leq A} |T_1(s, \pm\varepsilon) - T_{1n}(s, \pm\varepsilon)| \rightarrow 0, \quad \text{a.s., } n \rightarrow \infty,$$

then

$$\liminf_n \inf_{a \leq s \leq A} T_{1n}(s, -\varepsilon) > 0, \quad \text{a.s.,}$$

$$\limsup_n \sup_{a \leq s \leq A} T_{1n}(s, +\varepsilon) < 0, \quad \text{a.s.,}$$

which entails (4.6).

It remains to show (4.7)–(4.9). We first show (4.9) only for one case; the other cases follow by symmetry. Let $U = [a, A]$ and let

$$Q_n(s) = \int g(y, s) F_n(dy|x), \quad Q(s) = \int g(y, s) F(dy|x),$$

with

$$g(y, s) = \psi\left(\frac{y - \varepsilon}{s}\right) I(y - \varepsilon \leq 0).$$

Note that Q_n and Q are nondecreasing functions; therefore, by Lemmas 1 and 2 and by Billingsley [(1968), Theorem (5.2(iii))], we have that

$$\sup_{s \in U} |Q_n(s) - Q(s)| \rightarrow 0, \quad \text{a.s., } n \rightarrow \infty,$$

which entails (4.9).

It remains to show (4.7) because (4.8) will follow by a symmetry argument. Note that by conditions (A1) and (A2) for all $s \in \mathbb{R}^+$,

$$T_1(s, -\varepsilon) = \int \psi\left(\frac{u + \varepsilon}{s}\right) dF_0(u) \geq T_1(s, 0) = 0.$$

Hence, by continuity of $T_1(s, -\varepsilon)$, it suffices to show

$$(4.10) \quad T_1(s, -\varepsilon) - T_1(s, 0) \neq 0,$$

for all $s \in U$. Assume that (4.10) is not true; then there is an $\tilde{s} \in U$ such that the set $\{u: \psi((u + \varepsilon)/\tilde{s}) \neq \psi(u/\tilde{s})\}$ has F_0 -measure zero. By (A1) it is open and does not contain any neighborhood of zero; therefore, $\psi(\varepsilon/\tilde{s}) = \psi(0) = 0$. This contradicts (A5) and shows (4.10). \square

In order to prove Theorem 1(iii) we first show that for any small $\delta > 0$ there exists a compact interval $I = I_\delta$ centered by 0 such that

$$(4.11) \quad \liminf_n \inf_{t \in I} T_{2n}(s^*(t) - \delta, t) > 0, \quad \text{a.s.},$$

$$(4.12) \quad \limsup_n \sup_{t \in I} T_{2n}(s^*(t) + \delta, t) < 0, \quad \text{a.s.}$$

We show (4.11). The proof of (4.12) is similar. Fix some $\delta \in (0, \inf_t s^*(t))$. Using (4.2) and continuity of ψ one obtains that the function

$$Q(y, t) = \psi((y - t)/(s^*(t) - \delta))$$

is continuous in (y, t) . Therefore $T_2(s^*(t) - \delta, t)$ is continuous in t . In addition, monotonicity of $T_2(s, 0)$ and (4.1) entail that $T_2(s^*(0) - \delta, 0) > 0$. Hence there exists a compact interval I centered by 0 such that

$$(4.13) \quad \inf_{t \in I} T_2(s^*(t) - \delta, t) > 0.$$

By Lemma 3

$$(4.14) \quad \sup_{t \in I} |T_2(s^*(t) - \delta, t) - T_{2n}(s^*(t) - \delta, t)| \rightarrow 0, \quad \text{a.s.}, \quad n \rightarrow \infty,$$

and (4.11) follows from (4.13) and (4.14).

Now observe that $m_n = t_n(\sigma_n)$ by definition. Pulling (4.5) and (4.6) together yields

$$(4.15) \quad m_n \rightarrow 0, \quad \text{a.s.}, \quad n \rightarrow \infty.$$

In particular, $m_n \in I$, a.s. for n sufficiently large, and hence by (4.11) and (4.12)

$$\liminf_n T_{2n}(s^*(m_n) - \delta, m_n) > 0, \quad \text{a.s.},$$

$$\limsup_n T_{2n}(s^*(m_n) + \delta, m_n) < 0, \quad \text{a.s.}$$

These inequalities imply that $\sigma_n - s^*(m_n) \rightarrow 0$, a.s., $n \rightarrow \infty$, since $T_{2n}(s, m_n)$ is monotone in s and $T_{2n}(\sigma_n, m_n) = 0$. Applying (4.12) and (4.15) we finally obtain

$$|\sigma_n - s^*(0)| \leq |\sigma_n - s^*(m_n)| + |s^*(m_n) - s^*(0)| \rightarrow 0, \quad \text{a.s.}, \quad n \rightarrow \infty.$$

Since $s^*(0) = 1 = \sigma(x)$ this completes the proof of Theorem 1(iii).

5. Proof of Theorem 2. To simplify our notation we introduce the parameter $\vartheta = (t, s)$ and the function

$$\Psi(y, \vartheta) = \begin{pmatrix} \psi((y - t)/s) \\ \chi((y - t)/s) \end{pmatrix}.$$

Recall that the point x was fixed. We will write ϑ_n for $(m_n(x), \sigma_n(x))$ and ϑ^* for $(m(x), \sigma(x))$.

Introduce the matrix of derivatives

$$\Psi'(y, \vartheta) = \begin{pmatrix} -1/s\psi'((y-t)/s) & (t-y)/s^2\psi'((y-t)/s) \\ -1/s\chi'((y-t)/s) & (t-y)/s^2\chi'((y-t)/s) \end{pmatrix}.$$

The existence of this matrix in some neighborhood of ϑ^* is guaranteed by condition (A10) and positiveness of $\sigma(x)$. Now

$$\begin{aligned} & \sqrt{nh_n^d} \int \Psi(y, \vartheta^*) F_n(dy|x) \\ (5.1) \quad & = \sqrt{nh_n^d} \int (\Psi(y, \vartheta^*) - \Psi(y, \vartheta_n)) F_n(dy|x) \\ & = \left(\int_0^1 \left\{ \int \Psi'(y, \tau\vartheta^* + (1-\tau)\vartheta_n) F_n(dy|x) \right\} d\tau \right) \sqrt{nh_n^d} (\vartheta^* - \vartheta_n), \end{aligned}$$

if $|\vartheta^* - \vartheta_n|$ is small enough for the existence of $\Psi'(y, \vartheta)$ for $\vartheta: |\vartheta - \vartheta^*| \leq |\vartheta_n - \vartheta^*|$.

Next we shall prove

$$(5.2) \quad \sup_{\{\vartheta: |\vartheta - \vartheta^*| \leq |\vartheta_n - \vartheta^*|\}} \left\| \int \Psi'(y, \vartheta) F_n(dy|x) - \int \Psi'(y, \vartheta^*) F(dy|x) \right\| \rightarrow_P 0, \quad n \rightarrow \infty,$$

where $\|\cdot\|$ is any norm in the space of 2×2 matrices. It suffices to prove (5.2) for all components of matrices separately. Condition (A10) and positiveness of $\sigma(x)$ imply that the components of $\Psi'(y, \vartheta)$ are continuous and bounded in (y, ϑ) for $y \in \mathbb{R}$ and ϑ belonging to some neighborhood of ϑ^* . Hence by Lemma 3,

$$\sup_{|\vartheta - \vartheta^*| \leq \delta} \left\| \int \Psi'(y, \vartheta) F_n(dy|x) - \int \Psi'(y, \vartheta) F(dy|x) \right\| \rightarrow_P 0, \quad n \rightarrow \infty,$$

for sufficiently small $\delta > 0$. This gives

$$(5.3) \quad \sup_{\{\vartheta: |\vartheta - \vartheta^*| \leq |\vartheta_n - \vartheta^*|\}} \left\| \int \Psi'(y, \vartheta) F_n(dy|x) - \int \Psi'(y, \vartheta) F(dy|x) \right\| \rightarrow_P 0, \quad n \rightarrow \infty,$$

since by Theorem 1(iii), $\vartheta_n \rightarrow_P \vartheta^*$ as $n \rightarrow \infty$. In addition,

$$(5.4) \quad \sup_{\{\vartheta: |\vartheta - \vartheta^*| \leq |\vartheta_n - \vartheta^*|\}} \left\| \int \Psi'(y, \vartheta) F(dy|x) - \int \Psi'(y, \vartheta^*) F(dy|x) \right\| \rightarrow_P 0, \quad n \rightarrow \infty,$$

by uniform continuity of $\int \Psi'(y, \vartheta) F(dy|x)$ in some neighborhood of ϑ^* . We see that (5.2) follows from (5.3) and (5.4).

Using (5.2) one obtains

$$(5.5) \quad \begin{aligned} & \int_0^1 \left\{ \int \Psi'(y, \tau\vartheta^* + (1-\tau)\vartheta_n) F_n(dy|x) \right\} d\tau \\ & \rightarrow_P \int \Psi'(y, \vartheta^*) F(dy|x) = \frac{1}{\sigma(x)} \begin{pmatrix} \varphi_0 & 0 \\ 0 & \kappa_0 \end{pmatrix}, \quad n \rightarrow \infty. \end{aligned}$$

We now study the asymptotic distribution of the left-hand side of (5.1). Write

$$(5.6) \quad \sqrt{nh_n^d} \int \Psi(y, \vartheta^*) F_n(dy|x) = \left(\frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) \right)^{-1} G_n,$$

where

$$\begin{aligned} G_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} \eta_{in} \\ \zeta_{in} \end{pmatrix}, \\ \eta_{in} &= \frac{1}{\sqrt{h_n^d}} \Psi\left(\frac{Y_i - m(x)}{\sigma(x)}\right) K\left(\frac{X_i - x}{h_n}\right), \\ \zeta_{in} &= \frac{1}{\sqrt{h_n^d}} \chi\left(\frac{Y_i - m(x)}{\sigma(x)}\right) K\left(\frac{X_i - x}{h_n}\right). \end{aligned}$$

By Cacoullos (1966), under (A6)–(A8),

$$(5.7) \quad \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) \rightarrow_P f(x) \int K(u) du, \quad n \rightarrow \infty.$$

We shall show now that G_n is asymptotically normal. First consider the asymptotics of $E\{G_n\}$. We have

$$\begin{aligned} E\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_{in} \right\} &= \sqrt{n} E \eta_{1n} \\ &= \frac{\sqrt{n}}{\sqrt{h_n^d}} \int \int \psi\left(\frac{m(z) - m(x) + u}{\sigma(x)}\right) K\left(\frac{z - x}{h_n}\right) dF_0\left(\frac{u}{\sigma(z)}\right) f(z) dz \\ &= \sqrt{nh_n^d} \frac{1}{h_n^d} \int K\left(\frac{z - x}{h_n}\right) f(z) \varphi\left(\frac{m(z) - m(x)}{\sigma(x)}, \frac{\sigma(z)}{\sigma(x)}\right) dz, \end{aligned}$$

where $\varphi(a, b) = \int \psi(a + bu) dF_0(u)$.

Let D be diameter of the set $\{z: K(z) \neq 0\}$. It suffices to consider only such z that $|z - x| \leq Dh_n$. For such z it is obvious that $|m(z) - m(x)| \leq LDh_n$ and $|\sigma(z) - \sigma(x)| \leq L'Dh_n$. Thus by continuity of $\varphi'_a(a, b)$ and $\varphi'_b(a, b)$ one obtains

$$\begin{aligned} &\sup_{|z-x| \leq Dh_n} \left| \varphi\left(\frac{m(z) - m(x)}{\sigma(x)}, \frac{\sigma(z)}{\sigma(x)}\right) \right. \\ &\quad \left. - \varphi(0, 1) - \varphi'_a(0, 1) \left(\frac{m(z) - m(x)}{\sigma(x)}\right) - \varphi'_b(0, 1) \left(\frac{\sigma(z)}{\sigma(x)} - 1\right) \right| \\ &= o(h_n), \quad n \rightarrow \infty, \end{aligned}$$

where

$$\varphi(0, 1) = \int \psi(u) dF_0(u) = 0,$$

$$\varphi'_a(0, 1) = \int \psi'(u) dF_0(u) = \varphi_0,$$

$$\varphi'_b(0, 1) = \int u\psi'(u) dF_0(u) = 0.$$

This implies

$$(5.8) \quad \left| \sqrt{n} E \eta_{1n} - \sqrt{nh_n^d} \frac{\varphi_0}{h_n^d} \int \left(\frac{m(z) - m(x)}{\sigma(x)} \right) K \left(\frac{z - x}{h_n} \right) f(z) dz \right| \\ = o\left(h_n \sqrt{nh_n^d}\right) = o(1), \quad n \rightarrow \infty,$$

$$(5.9) \quad \lim_n \sqrt{nh_n^d} \frac{1}{h_n^d} \int (m(z) - m(x)) K \left(\frac{z - x}{h_n} \right) \mu(z) dz \\ = \beta^{d/2+1} \int m'(x; u) K(u) du \mu(x).$$

Together (5.8) and (5.9) yield

$$(5.10) \quad \lim_n \sqrt{n} E \eta_{1n} = \frac{\varphi_0}{\sigma(x)} \beta^{d/2+1} f(x) \int m'(x; u) K(u) du = b_1.$$

Let $\kappa(a, b) = \int \chi(a + bu) dF_0(u)$. Then

$$E \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_{in} \right\} = \sqrt{n} E \xi_{1n} \\ = \frac{\sqrt{n}}{\sqrt{h_n^d}} \int \kappa \left(\frac{m(z) - m(x)}{\sigma(x)}, \frac{\sigma(z)}{\sigma(x)} \right) K \left(\frac{z - x}{h_n} \right) f(z) dz, \\ \sup_{\{z: |z-x| \leq Dh_n\}} \left| \kappa \left(\frac{m(z) - m(x)}{\sigma(x)}, \frac{\sigma(z)}{\sigma(x)} \right) - \kappa(0, 1) - \kappa'_a(0, 1) \left(\frac{m(z) - m(x)}{\sigma(x)} \right) \right. \\ \left. - \kappa'_b(0, 1) \left(\frac{\sigma(z)}{\sigma(x)} - 1 \right) \right| = o(h_n), \quad n \rightarrow \infty,$$

where

$$\kappa(0, 1) = \int \chi(u) dF_0(u) = 0,$$

$$\kappa'_a(0, 1) = \int \chi'(u) dF_0(u) = 0,$$

$$\kappa'_b(0, 1) = \int u\chi'(u) dF_0(u) = \kappa_0.$$

Similarly to (5.10) one proves

$$(5.11) \quad \lim_n \sqrt{n} E \zeta_{1n} = \frac{\kappa_0}{\sigma(x)} \beta^{d/2+1} \int \sigma'(x; u) K(u) du f(x) = b_2.$$

Note that by (5.10) and (5.11),

$$(5.12) \quad E \eta_{1n} = O\left(\frac{1}{\sqrt{n}}\right), \quad E \zeta_{1n} = O\left(\frac{1}{\sqrt{n}}\right), \quad n \rightarrow \infty.$$

Consider the asymptotics of the covariance matrix of G_n . In view of (5.12),

$$(5.13) \quad \begin{aligned} \lim_n \text{var} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_{in} \right) &= \lim_n [E \eta_{1n}^2 - (E \eta_{1n})^2] = \lim_n E \eta_{1n}^2 \\ &= \lim_n \frac{1}{h_n^d} \int \int \psi^2 \left(\frac{m(z) - m(x) + u}{\sigma(x)} \right) dF_0 \left(\frac{u}{\sigma(z)} \right) \\ &\quad \times K^2 \left(\frac{z - x}{h_n} \right) f(z) dz \\ &= f(x) \varphi_2 \int K^2(u) du = \sigma_1^2. \end{aligned}$$

Similarly to (5.13),

$$(5.14) \quad \lim_n \text{var} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_{in} \right) = f(x) \kappa_2 \int K^2(u) du = \sigma_2^2.$$

The components of G_n are asymptotically uncorrelated because

$$(5.15) \quad \begin{aligned} &\lim_n \frac{1}{n} \sum_{i=1}^n E \{ (\eta_{in} - E \eta_{in}) (\zeta_{in} - E \zeta_{in}) \} \\ &= \lim_n \frac{1}{n} \sum_{i=1}^n E \{ \eta_{in} \zeta_{in} \} \\ &= \lim_n \frac{1}{h_n^d} \int \int \psi \left(\frac{m(z) - m(x) + u}{\sigma(x)} \right) \chi \left(\frac{m(z) - m(x) + u}{\sigma(x)} \right) dF_0 \left(\frac{u}{\sigma(z)} \right) \\ &\quad \times K^2 \left(\frac{z - x}{h_n} \right) f(z) dz \\ &= f(x) \int K^2(u) du \int \chi(u) \psi(u) dF_0(u) = 0 \end{aligned}$$

[recall that $\psi(u)$ is odd and $\chi(u)$ is even].

We now prove

$$(5.16) \quad G_n \rightarrow_{\mathcal{D}} \begin{pmatrix} \eta \\ \zeta \end{pmatrix}, \quad n \rightarrow \infty,$$

where $\eta \sim \mathcal{N}(b_1, \sigma_1^2)$, $\zeta \sim \mathcal{N}(b_2, \sigma_2^2)$ and $\text{cov}\{\eta, \zeta\} = 0$. In view of (5.10), (5.11),

(5.13)–(5.15) and Theorem 7.7 of Billingsley (1968) (Cramér–Wold device), it is sufficient to prove that linear combinations of components of G_n satisfy the Lyapunov condition of the central limit theorem.

Since ψ and χ are bounded we obtain

$$(5.17) \quad |\eta_{1n}|, |\zeta_{1n}| \leq \frac{C}{\sqrt{h_n^d}} K\left(\frac{X_i - x}{h_n}\right),$$

where $C > 0$ is some constant. Let $a_1, a_2 \in \mathbb{R}$ be arbitrary. The Lyapunov condition for linear combinations follows from

$$\begin{aligned} & \sum_{i=1}^n E\left(\frac{a_1}{\sqrt{n}}(\eta_{in} - E\eta_{in}) + \frac{a_2}{\sqrt{n}}(\zeta_{in} - E\zeta_{in})\right)^4 \\ & \leq \frac{8}{n} \left(a_1^4 E(\eta_{1n} - E\eta_{1n})^4 + a_2^4 E(\zeta_{1n} - E\zeta_{1n})^4\right) \\ & \leq \frac{64}{n} \left(a_1^4 E\eta_{1n}^4 + a_2^4 E\zeta_{1n}^4\right) + O\left(\frac{1}{n^3}\right) = O\left(\frac{1}{nh_n^d}\right) = o(1), \quad n \rightarrow \infty, \end{aligned}$$

where we used (5.12), (5.17) and the elementary inequality $(a + b)^4 \leq 8(a^4 + b^4)$. This proves (5.16). The assertion of Theorem 2 follows from (5.1), (5.5)–(5.7) and (5.16).

REFERENCES

- BILLINGSLEY, R. (1968). *Convergence of Probability Measures*. Wiley, New York.
- BUSSIAN, B. and HÄRDLE, W. (1984). Robust smoothing applied to white noise and outlier contaminated Raman spectra. *Appl. Spectroscopy* **38** 309–313.
- CACOULOS, T. (1966). Estimation of a multivariate density. *Ann. Inst. Statist. Math.* **18** 178–189.
- CARROLL, R. (1982). Adapting for heteroscedasticity in linear models. *Ann. Statist.* **10** 1224–1233.
- COLLOMB, G. (1980). Estimation non paramétrique des probabilités conditionnelles. *C. R. Acad. Sci. Paris Ser. A* **291** 427–430.
- COLLOMB, G. (1981). Estimation non paramétrique de la régression: Revue bibliographique. *Internat. Statist. Rev.* **49** 75–93.
- COX, D. D. (1983). Asymptotics for M -type smoothing splines. *Ann. Statist.* **11** 530–551.
- HÄRDLE, W. (1984). Robust regression function estimation. *J. Multivariate Anal.* **14** 169–180.
- HÄRDLE, W. and GASSER, T. (1984). Robust nonparametric function fitting. *J. Roy. Statist. Soc. Ser. B* **46** 42–51.
- HÄRDLE, W. and MARRON, S. (1985). Optimal bandwidth selection for nonparametric kernel regression. *Ann. Statist.* **13** 1465–1481.
- HILDENBRAND, K. and HILDENBRAND, W. (1986). On the mean income effect: A data analysis of the UK Family Expenditure Survey. In *Contributions to Mathematical Economics* (W. Hildenbrand and A. Mas-Colell, eds.). North-Holland, Amsterdam.
- HUBER, P. (1979). Robust smoothing. In *Robustness in Statistics* (R. L. Launer and G. N. Wilkinson, eds.) 33–47. Academic, New York.
- HUBER, P. (1981). *Robust Statistics*. Wiley, New York.
- JUSTUSSON, B. I. (1981). Median filtering: Statistical properties. In *Two-Dimensional Digital Signal Processing. Transforms and Median Filters. 2*. Springer, Berlin.
- MARONNA, R. A. (1976). Robust M -estimators of multivariate location and scatter. *Ann. Statist.* **4** 51–67.
- NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

- NEMIROVSKIĬ, A. S., POLYAK, B. T. and TSYBAKOV, A. B. (1983). Estimators of maximum likelihood type for nonparametric regression. *Soviet Math. Dokl.* **28** 788–792.
- SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). *J. Roy. Statist. Soc. Ser. B* **47** 1–52.
- STONE, C. J. (1977). Consistent nonparametric regression (with discussion). *Ann. Statist.* **5** 595–645.
- STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.
- TSYBAKOV, A. B. (1982a). Nonparametric signal estimation when there is incomplete information on the noise distribution. *Problemy Peredači Informacii* **18**(2) 44–60. English translation in *Problems Inform. Transmission* **18** 116–130.
- TSYBAKOV, A. B. (1982b). Robust estimates of a function. *Problemy Peredači Informacii* **18**(3) 39–52. English translation in *Problems Inform. Transmission* **18** 190–201.
- TSYBAKOV, A. B. (1983). Convergence of nonparametric robust algorithms of reconstruction of functions. *Avtomat. i Telemekh.* (12) 66–76. English translation in *Avtomat. Remote Control* **44** 1582–1591.
- WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā* **26** 359–372.

DEPARTMENT OF ECONOMICS
UNIVERSITY OF BONN
ADENAUERALLEE 24-26
D-5300 BONN 1
WEST GERMANY

INSTITUTE FOR PROBLEMS OF
INFORMATION TRANSMISSION
ACADEMY OF SCIENCES OF THE USSR
ERMOLOVOVSTR. 19
101 447 MOSCOW GSP-4
USSR