The Annals of Statistics
1987, Vol. 15, No. 1, 182–201

VARIABLE BANDWIDTH KERNEL ESTIMATORS OF REGRESSION CURVES

BY HANS-GEORG MÜLLER AND ULRICH STADTMÜLLER

University of Marburg and University of Ulm

In the model

$$Y_i = g(t_i) + \varepsilon_i, \qquad i = 1, \dots, n,$$

where Y_i are given observations, ε_i i.i.d. noise variables and t_i nonrandom design points, kernel estimators for the regression function $g(t)$ with variable bandwidth (smoothing parameter) depending on t are proposed. It is shown that in terms of asymptotic integrated mean squared error, kernel estimators with such a local bandwidth choice are superior to the ordinary kernel estimators with global bandwidth choice if optimal bandwidths are used. This superiority is maintained in a certain sense if optimal local bandwidths are estimated in a consistent manner from the data, which is proved by a tightness argument. The finite sample behavior of a specific local bandwidth selection procedure based on the Rice criterion for global bandwidth choice [Rice (1984)] is investigated by simulation.

1. Introduction. We consider kernel estimates of regression curves in the fixed design case. Estimators of this kind have been proposed by Priestley and Chao (1972), Cheng and Lin (1981) and Gasser and Müller (1984) and constitute an attractive alternative to smoothing splines.

Observations $Y_1, \dots, Y_n$ are generated according to

$$Y_i = g(t_i) + \varepsilon_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where $g \in \mathrm{Lip}([0,1])$ is the curve to be estimated and ε_i are i.i.d. noise variables satisfying $E\varepsilon_1 = 0$, $E\varepsilon_1^2 = \sigma^2 < \infty$. The design $t_1, \dots, t_n$ is fixed in advance as $t_i = i/n$. In order to avoid consideration of boundary effects, we assume that data are available on both sides beyond 0 and 1. Also, the assumption of equidistance is not essential for our main results.

As kernel estimator of $g(t)$ we consider

$$\hat{G}(t, b) := \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{t - t_i}{b}\right) Y_i. \tag{1.2}$$

This estimator depends on the kernel function K and a sequence of bandwidths $b = b(n)$ which has to satisfy

$$b \to 0, \qquad nb^2 \to \infty, \quad \text{as } n \to \infty. \tag{1.3}$$

It is a well-known phenomenon that estimators (1.2) exhibit an increased bias near peaks of g [compare Gasser et al. (1984)]. Since the bias near peaks may

Received April 1984; revised March 1985.

AMS 1980 *subject classifications.* Primary 62G05; secondary 65J02, 65D10.

Key words and phrases. Nonparametric kernel regression, consistent bandwidth choice, Rice criterion, local bandwidths, asymptotic optimality, tightness in C, Gaussian limiting process.

contribute especially a large portion to the integrated mean squared error (IMSE) of the estimate, IMSE might be reduced by decreasing the bandwidth and therefore local MSE near peaks, and increasing the bandwidth in flat parts of the curve. These ideas motivate the choice of locally varying bandwidths. The corresponding kernel estimator is

$$\hat{g}(t, b_t) := \frac{1}{nb_t} \sum_{i=1}^{n} K\left(\frac{t - t_i}{b_t}\right) Y_i, \tag{1.4}$$

given some function $t \to b_t$, $t \in [0,1]$.

In the following, the notion of optimality always refers to minimization of the leading terms of asymptotic expressions (2.5) of the MSE/IMSE. We prove the following result in Section 2: Let b^*, b_t^* be the optimal bandwidths for (1.2) and (1.4) and assume that $\tilde{b}_t^*$ is a truncated version of b_t^* (there is a cut-off if b_t^* gets too large). Then

$$\lim_{n \to \infty} n^{2k/(2k+1)} \mathrm{IMSE}\big(\hat{g}(\cdot, \tilde{b}_t^*)\big) \le \lim_{n \to \infty} n^{2k/(2k+1)} \mathrm{IMSE}\big(\hat{G}(\cdot, b^*)\big), \tag{1.5}$$

if $g \in \mathscr{C}^k([0,1])$ for some $k \ge 2$ and $K \in \mathscr{M}_{0,k}$, where for $0 \le \nu \le k$

$$\mathscr{M}_{\nu,k} := f \in \mathrm{Lip}([-1,1]),\ \int_{-1}^{1} f(x) x^j \, dx = \begin{cases} 0, & j = 0, \ldots, k-1, \\ & j \ne \nu, \\ (-1)^\nu \nu!, & j = \nu, \\ \ne 0, & j = k. \end{cases}$$

While inequality (1.5) is expected to hold on general grounds since the class of variable bandwidth estimators contains the class of fixed bandwidth estimators, it will be of interest to derive explicit expressions for the terms on both sides in (2.8)–(2.10).

In Section 3, we show that if $\hat{b}_t$ is a consistent estimator of b_t^*, i.e., $\hat{b}_t / b_t^* \to_p 1$, as $n \to \infty$, then $\hat{g}(\cdot, \hat{b}_t)$ behaves asymptotically as well as $\hat{g}(\cdot, b_t^*)$. By a tightness argument, we prove that $n^{k/(2k+1)}(\hat{g}(t, \tau n^{-1/(2k+1)}) - g(t))$ (t fixed) converges in distribution to a Gaussian limit process on $\mathscr{C}([r, s])$ for some τ-interval $[r, s]$, and hence any consistent estimator for b_t^* will be asymptotically efficient (Theorem 3.1). Results of the type, $\mathrm{IMSE}(\hat{g}(t, \hat{b}_t))/\mathrm{IMSE}(\hat{G}(t, b^*)) \le 1$, can be inferred under additional assumptions (Theorems 3.2 and 3.3). Weak conditions under which these results apply are given in Theorem 4.1 (Section 4).

In Section 5, we propose a specific method for local bandwidth choice with nice finite sample properties, applying an asymptotic relation between optimal bandwidths for estimators of different orders of derivatives. This relationship allows to derive consistent procedures for bandwidth choice for the estimation of derivatives from consistent procedures for the estimation of the curve itself. A different proposal for bandwidth choice for derivatives is due to Rice (1986). The starting point of our procedure is a consistent method of global bandwidth choice for the curve itself introduced by Rice (1984). Other methods of global bandwidth choice, e.g., the testgraph method of Silverman (1978), cross-validation or generalized cross-validation [Craven and Wahba (1979)] and related methods [Rice (1983)] could be used in the initial step of our method, too. Section 6 contains a brief discussion.

The choice of variable bandwidths is also implied by nearest-neighbor estimators [Stone (1977)], but these are not applicable in our situation which is based on an equidistant design fixed in advance. Other proposals in density estimation consider bandwidths varying across the sample [Breiman et al. (1977) and Abramson (1982)], whereas we discuss variation of bandwidth with respect to local curvature at t. Our approach is related to work on asymptotic efficiency of estimated bandwidths at a point in density estimation, due to Woodroofe (1970), Krieger and Pickands (1981) and Abramson (1982a).

2. Auxiliary results. Let $k \geq 2$, $0 \leq \nu \leq k$, be given and assume that $g \in \mathscr{C}^k([0,1])$ and $K \in \mathscr{M}_{0,k}$.

The proposal of Rice (1984) for global bandwidth choice in the case of the estimator (1.2) is to choose the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{G}(t_i, b)\right)^2 - \hat{\sigma}^2 + \frac{2\hat{\sigma}^2}{nb}K(0), \tag{2.1}$$

w.r.t. b. Here, $\hat{\sigma}$ denotes any consistent estimator of σ [obtained, e.g., by fitting constants to successive triples of points, see Rice (1984)]. We denote this minimizer by $\hat{b}$. According to Rice, the estimator $\hat{b}$ is consistent in the sense that

$$\frac{\hat{b} - b^*}{b^*} \to_p 0, \quad \text{as } n \to \infty, \tag{2.2}$$

where b^* denotes the optimal global bandwidth. Rice states this result for $k = 2$, but the proof may be easily extended to cover the case $k > 2$. The kernels have to satisfy certain smoothness properties and the existence of eight moments of the noise variables is required.

Since the method of local bandwidth choice to be introduced in Section 4 requires estimators of g'', we extend the estimator (1.2) to the estimation of $g^{(\nu)}(t)$ [compare Gasser and Müller (1984)]:

$$\hat{G}_\nu(t, b) := \frac{1}{nb^{\nu+1}}\sum_{i=1}^{n} K_\nu\left(\frac{t - t_i}{b}\right)Y_i, \tag{2.3}$$

where $K_\nu \in \mathscr{M}_{\nu,k}$. Such a kernel is said to be of order k. In the same way, $\hat{g}$ may be generalized to $\hat{g}_\nu$.

LEMMA 2.1. *Assume that* $nb^{2\nu+2}/(\log n)^2 \to \infty$ *and* K_ν *is Lipschitz continuous on* $\mathbb{R}$. *Then*

$$\sup_{t\in[0,1]}\left|\hat{G}_\nu(t, b) - G^{(\nu)}(t)\right| \to 0, \qquad n \to \infty \quad a.s.$$

PROOF. The assertion follows from Theorem 3.2 of Müller (1983), choosing $m = 1$, $B = [0,1]$ and $r = 2$. That the requirements for Theorem 3.2 are satisfied under the present assumptions follows from the proof of Corollary 4.1C, from Lemma 3.2 and from the first part of the proof of Corollary 4.5 of Müller (1983). That $\sup_{t\in[0,1]}|G^{(\nu)}(t) - E\hat{G}_\nu(t, b)| \to 0$ under $K_\nu \in \mathscr{M}_{\nu,\nu}$, $g \in \mathscr{C}^\nu([0,1])$, may also be easily verified directly by making use of the fact that the support of K_ν is compact. □

Related results can be found in Cheng and Lin (1981) and Stadtmüller (1982, 1986).

In the following, we assume that $k > \nu$. Define

$$\beta_{\nu,k} := (k!)^{-1}\int K_\nu(x)x^k\,dx \quad \text{and} \quad v_{\nu,k} := \int K_\nu^2(x)\,dx.$$

Standard calculations involving a Taylor expansion yield

$$E\hat{g}_\nu(t, b_t) - g^{(\nu)}(t) = b_t^{k-\nu}\big(\beta_{\nu,k}g^{(k)}(t) + o(1)\big) + O\big((nb_t^\nu)^{-1}\big) \tag{2.4}$$

and

$$\begin{aligned} &E\big(\hat{g}_\nu(t, b_t) - g^{(\nu)}(t)\big)^2 \\ &\quad = \frac{\sigma^2}{nb_t^{2\nu+1}}(v_{\nu,k} + o(1)) + b_t^{2(k-\nu)}\big(\beta_{\nu,k}^2 g^{(k)}(t)^2 + o(1)\big), \end{aligned} \tag{2.5}$$

and we obtain the locally optimal bandwidth

$$b_t^* = \left(\frac{2\nu+1}{2(k-\nu)}\frac{v_{\nu,k}}{\beta_{\nu,k}^2}\frac{\sigma^2}{g^{(k)}(t)^2}\frac{1}{n}\right)^{1/(2k+1)}, \tag{2.6}$$

provided that $g^{(k)}(t) \neq 0$.

Analogously, (2.5) holds for $\hat{G}_\nu(t, b)$, and the o-terms are uniform in $t \in [0,1]$ by (1.3). In this case we may therefore integrate (2.5) over $[0,1]$ and obtain the optimal bandwidth (assuming that $\int_0^1 g^{(k)}(t)^2\,dt \neq 0$) w.r.t. the IMSE, namely,

$$b^* = \left(\frac{2\nu+1}{2(k-\nu)}\frac{v_{\nu,k}}{\beta_{\nu,k}^2}\frac{\sigma^2}{\int_0^1 g^{(k)}(t)^2\,dt}\frac{1}{n}\right)^{1/(2k+1)}, \tag{2.7}$$

in analogy to (2.6). Inserting (2.7) into the integral of (2.5) again yields

$$\begin{aligned} &E\int_0^1\big(\hat{G}_\nu(t, b^*) - g^{(\nu)}(t)\big)^2\,dt \\ &\quad = n^{-2(k-\nu)/(2k+1)}\Bigg[c(\nu,k)\big(\sigma^2 v_{\nu,k}\big)^{2(k-\nu)/(2k+1)} \\ &\qquad \times\left(\beta_{\nu,k}^2\int_0^1 g^{(k)}(t)^2\,dt\right)^{(2\nu+1)/(2k+1)} + o(1)\Bigg], \end{aligned} \tag{2.8}$$

where $c(\nu, k)$ is a constant depending on ν, k only.

In order to derive a similar result for $\hat{g}_\nu$, we note that $(g^{(k)})^2$ is limited above but can be zero, and therefore we truncate b_t^* (2.6) and define $\tilde{b}_t^* :=$

$(d/n)^{1/(2k+1)} \wedge b_t^*$, where $d > 0$ is a constant s.t. $(d/n)^{1/(2k+1)} > b^*$ [see (2.7)]. In practical applications, the choice of d has to be a compromise between this requirement and the necessity to avoid boundary effects by choosing d not too large. Given an estimate $\hat{b}^*$ of b^*, our recommendation is to choose d s.t. $(d/n)^{1/(2k+1)} = 2\hat{b}^*$.

Define $S_1 := \{t \in [0,1]: \tilde{b}_t^* = b_t^*\}$ and $S_2 := [0,1] \setminus S_1$. After insertion of b_t^* into (2.5), o-terms are uniform over $[0,1]$ since $n \inf_t \tilde{b}_t^{*2} \to \infty$, $\sup_t \tilde{b}_t^* \to 0$ as $n \to \infty$. Therefore we may integrate to obtain

$$E\int_{S_1} \left(\hat{g}_\nu(t, \tilde{b}_t^*) - g^{(\nu)}(t)\right)^2 dt$$

$$(2.9) \quad = n^{-2(k-\nu)/(2k+1)}\Bigg[c(\nu, k)\left(\sigma^2 v_{\nu,k}\right)^{2(k-\nu)/(2k+1)} \times \left(\beta_{\nu,k}\right)^{2(2\nu+1)/(2k+1)} \int_{S_1} g^{(k)}(t)^{2(2\nu+1)/(2k+1)}\, dt + o(1)\Bigg]$$

and

$$E\int_{S_2} \left(\hat{g}_\nu(t, \tilde{b}_t^*) - g^{(\nu)}(t)\right)^2 dt$$

$$(2.10) \quad = n^{-2(k-\nu)/(2k+1)}\Bigg[d^{-(2\nu+1)/(2k+1)} v_{\nu,k} \int_{S_2} \sigma^2\, dt + d^{2(k-\nu)/(2k+1)} \beta_{\nu,k}^2 \int_{S_2} g^{(k)}(t)^2\, dt + o(1)\Bigg].$$

If $\inf_t |g^{(k)}(t)| > 0$ and d is chosen large enough, the right-hand side of (2.10) vanishes since then $S_2 = \varnothing$.

By Hölder's inequality it follows from (2.8) and (2.9) that (1.5) holds for integration performed on S_1 and also for integration performed on S_2, since there $b^* < (d/n)^{1/(2k+1)} \le b_t^*$. This yields some insight into the possible gain of $\mathrm{IMSE}(\hat{g}_\nu(\cdot, b_t^*))$ as compared to $\mathrm{IMSE}(\hat{G}_\nu(\cdot, b^*))$.

Let $b^{(\nu)*}$ denote the asymptotically optimal global bandwidth for the estimator $\hat{G}_\nu$. If we compare estimators $\hat{G}_0, \hat{G}_\nu$ with kernels $K_0 \in \mathscr{M}_{0,k}$, $K_\nu \in \mathscr{M}_{\nu,k}$, then we obtain from (2.7) the asymptotic relation

$$(2.11) \qquad \frac{b^{(0)*}}{b^{(\nu)*}} = \left(\frac{v_{0,k}}{v_{\nu,k}} \frac{\beta_{\nu,k}^2}{\beta_{0,k}^2} \frac{2(k-\nu)}{2\nu+1} \frac{1}{2k}\right)^{1/(2k+1)} =: d_{\nu,k} \quad \text{say.}$$

The right-hand side $d_{\nu,k}$ is a known constant depending on the kernel functions. Therefore, this relation may be used to derive a consistent estimate of $b^{(\nu)*}$ from a consistent estimate of $b^{(0)*}$.

3. Asymptotic behavior of kernel estimators with data-based local bandwidth selection. In this section we show that for any consistent estimator $\hat{b}_t$ of b_t^*, $\hat{g}(t, \hat{b}_t)$ behaves asymptotically as well as $\hat{g}(t, b_t^*)$, as far as convergence in probability is concerned. Throughout this section we assume that $\nu < k$, $K_\nu \in \mathscr{M}_{\nu,k}$, K_ν' exists, $K_\nu' \in \text{Lip}([-1,1]$ and $g \in \mathscr{C}^k([0,1])$. For simplicity we consider the case $\nu = 0$ only.

LEMMA 3.1. *Assume that* $g^{(k)}(t) \neq 0$ *for some* $t \in [0,1]$. *By* (2.6) *there exists* $0 < \tau^* < \infty$ *s.t.* $b_t^* = \tau^* n^{-1/(2k+1)}$. *Assume that* r, s *with* $0 < r < \tau^* < s < \infty$ *are fixed constants. Then, writing* $b(\tau) := \tau n^{-1/(2k+1)}$,

$$Y_n(\tau) := n^{k/(2k+1)}(\hat{g}(t, b(\tau)) - g(t))$$

is a random function on $\mathscr{C}([r,s])$ *and*

$$(3.1)\quad Y_n(\tau) \to_D Y(\tau) = \frac{\sigma}{\tau}\int K\left(\frac{x}{\tau}\right) dW(x) + \tau^k \beta_{0,k} g^{(k)}(t), \quad \text{as } n \to \infty,$$

on $\mathscr{C}([r,s])$ *endowed with the sup-norm, where* $W(\cdot)$ *is a standard Wiener process.*

REMARKS. (i) In case $K(x) = \frac{1}{2}\chi_{[-1,1]}(x)$, (3.1) can be seen by Donsker's invariance principle, and the first term of the limit process can be written as $(W(\tau)/\tau)\sigma$ [compare Krieger and Pickands (1981) who investigated this case in density estimation but used a different approach based on Poissonization].

(ii) A related tightness argument was used by Abramson (1982a) in density estimation.

PROOF. We split $Y_n(\tau)$ into two parts,

$$n^{k/(2k+1)}(\hat{g}(t, b(\tau)) - E\hat{g}(t, b(\tau)))$$
$$+ n^{k/(2k+1)}(E\hat{g}(t, b(\tau)) - g(t)) =: Z_n(\tau) + B_n(\tau), \quad \text{say.}$$

According to (2.4)

$$B_n(\tau) \to \tau^k \beta_{0,k} g^{(k)}(t), \quad \text{as } n \to \infty, \text{ uniformly on } [r,s].$$

By Theorem 8.1 and Theorem 12.3 of Billingsley (1968) it is sufficient to show for the first part:

(i) Appropriate convergence of the finite-dimensional limit distributions of $Z_n(\tau)$.

(ii) $Z_n(\tau^*)$ is tight.

(iii) There exist constants $\gamma \geq 0$, $\alpha > 1$, and a continuous function F nondecreasing on $[r,s]$ s.t.

$$(3.2)\quad E\left(|Z_n(\tau_2) - Z_n(\tau_1)|^\gamma\right) \leq |F(\tau_2) - F(\tau_1)|^\alpha,$$

for all $r \leq \tau_1 \leq \tau_2 \leq s$ and all n.

(ii) and (iii) imply tightness of $Z_n(\tau)$.

(i) Under our assumptions,

$$Z_n(\tau) \to_{\mathscr{D}} \mathscr{N}\left(0, \frac{\sigma^2 v_{0,k}}{\tau}\right), \quad \text{for any fixed } \tau \in [r,s]$$

[compare, e.g., Stadtmüller (1982, 1986)]. Furthermore, an application of the Cramér–Wold device (see Billingsley) yields for $\tau_1 < \tau_2 < \cdots < \tau_m \in [r, s]$

$$(Z_n(\tau_1), \dots, Z_n(\tau_m)) \to_{\mathscr{D}} \mathscr{N}(0, C),$$

where the elements of the covariance matrix C are

$$c_{ij} = \frac{\sigma^2}{\tau_i \tau_j} \int K\left(\frac{x}{\tau_i}\right) K\left(\frac{x}{\tau_j}\right) dx, \qquad 1 \le i, j \le m.$$

The derivation is based on similar algebra as in (2.5). Hence the limit distribution coincides with the distribution of $(Z(\tau_1), \dots, Z(\tau_m))$, where

$$Z(\tau) = \frac{\sigma}{\tau} \int K\left(\frac{x}{\tau}\right) dW(x).$$

(ii) Follows from $E(Z_n^2(\tau^*)) \le c < \infty$ for all n [compare (2.5)].

(iii) We show that there exists a constant $c > 0$ s.t.

$$E\left(|Z_n(\tau_2) - Z_n(\tau_1)|^2\right) \le c(\tau_2 - \tau_1)^2, \text{ for all } \tau_1, \tau_2 \in [r, s], \tag{3.3}$$

which implies (3.2) with $F(\tau) = \sqrt{c}\,\tau$.

For (3.3) we proceed as follows, setting $b_1 = \tau_1 n^{-1/(2k+1)}$, $b_2 = \tau_2 n^{-1/(2k+1)}$:

$$\begin{aligned} &E\left((Z_n(\tau_2) - Z_n(\tau_1))^2\right) \\ &= n^{2k/2k+1} E\left(\left[\sum_{i=1}^n \left(\frac{1}{nb_1} K\left(\frac{t - t_i}{b_1}\right) - \frac{1}{nb_2} K\left(\frac{t - t_i}{b_2}\right)\right)\varepsilon_i\right]^2\right) \\ &= n^{2k/(2k+1)-2} \sigma^2 \sum_{n=1}^n \left(\frac{\partial}{\partial b}\left\{\frac{1}{b} K\left(\frac{t - t_i}{b}\right)\right\}\Bigg|_{b=b_3} (b_2 - b_1)\right)^2, \end{aligned}$$

with $b_3 \in [b_1, b_2]$.

We conclude with $\tilde{K}(x) = K(x) + xK'(x)$ and $\tau_3 = b_3 n^{1/2k+1}$ such that

$$\begin{aligned} E(Z_n(\tau_2) - Z_n(\tau_1))^2 &\le (\tau_2 - \tau_1)^2 \frac{n^{-3/(2k+1)} \sigma^2}{b_3^3} \frac{1}{nb_3} \sum_{i=1}^n \tilde{K}^2\left(\frac{t - t_i}{b_3}\right) \\ &\le c(\tau_2 - \tau_1)^2, \quad \text{for any } \tau_i \in [r, s],\ i = 1, 2. \end{aligned}$$

Observe that $\sum_{i=1}^n \tilde{K}^2((t - t_i)/b_3) = O(nb_3)$, since $\operatorname{supp}(\tilde{K}) = [-1, 1]$. □

As a consequence we obtain

Theorem 3.1. *Let the assumptions of Lemma* 3.1 *hold; in addition suppose that there is a sequence of r.v.'s* $\hat{\tau} = \hat{\tau}(n)$ *s.t.* $\hat{\tau} \to_p \tau^*$ *as* $n \to \infty$. *Then*

$$n^{k/(2k+1)}(\hat{g}(t, b(\hat{\tau})) - \hat{g}(t, b(\tau^*))) \to_p 0, \quad \text{as } n \to \infty. \tag{3.4}$$

Remarks. (i) Since

$$n^{k/(2k+1)}(\hat{g}(t, b(\tau^*)) - g(t)) \to_{\mathscr{D}} \mathscr{N}\left(\tau^{*k} \beta_{0,k} g^{(k)}(t), \frac{\sigma^2 v_{0,k}}{\tau^*}\right),$$

(3.4) implies that a kernel estimator supplied with a consistent estimate of the optimal local bandwidth behaves asymptotically as well as the optimal estimator in terms of the limiting distribution.

(ii) An analogous result can be proven for $\hat{g}_\nu(\cdot,\cdot)$.

PROOF. Given $\alpha, \eta > 0$, by Lemma 3.1 we can choose $\delta > 0$ s.t.

$$P\Big(\max_{|\tau-\tau^*|\le\delta} n^{k/(2k+1)}|\hat{g}(t, b(\tau)) - \hat{g}(t, b(\tau^*))| > \alpha\Big) < \eta/2,$$

for n large enough. Hence

$$P\big(n^{k/(2k+1)}|\hat{g}(t, b(\hat{\tau})) - \hat{g}(t, b(\tau^*))| > \alpha\big) \le \frac{n}{2} + P(|\hat{\tau} - \tau^*| > \delta) \le \eta,$$

for n large enough. □

In practice it is more useful to consider a truncated bandwidth estimator like

$$(3.5)\qquad \hat{\rho}_t := \begin{cases} \hat{\tau}_t, & \text{if } r \le \hat{\tau}_t \le s,\\ r, & \text{if } \hat{\tau}_t < r,\\ s, & \text{elsewhere.}\end{cases}$$

For the choice of s compare the discussion following (2.8). In the following we want to analyze the behavior of $\hat{g}(t, b(\hat{\rho}_t))$ w.r.t. IMSE. Given a consistent estimator $\hat{\tau}_t$ s.t. $\hat{\tau}_t \to_p \tau_t^*$ (as $n \to \infty$), where $b_t^* = \tau_t^* n^{-1/(2k+1)}$ is the optimal local bandwidth, there is some $\delta > 0$ depending on s and the constants in (2.6) s.t.

$$\hat{\rho}_t \to_p \begin{cases} \tau_t^*, & \text{if } |g^{(k)}(t)| > \delta,\\ s, & \text{elsewhere,}\end{cases} \quad \text{as } n \to \infty.$$

According to (2.7), the optimal global bandwidth is $b^* = \tilde{\tau} n^{-1/(2k+1)}$ with some $\tilde{\tau} > 0$, provided that $\int_0^1 [g^{(k)}(t)]^2\,dt > 0$. Using this notation we state

THEOREM 3.2. *Assume that $\int_0^1 (g^{(k)}(t))^2\,dt > 0$ and $r < \tilde{\tau} < s$. Then*

$$\sup_{a>0} \limsup_{n\to\infty} \int_0^1 E\Big[n^{2k/(2k+1)}\big(\hat{g}(t, b(\hat{\rho}_t)) - g(t)\big)^2 \wedge a\Big]\,dt$$
$$\le \lim_{n\to\infty} n^{2k/(2k+1)} \mathrm{IMSE}\big(\hat{G}(t, b^*)\big).$$

If in addition $E(\varepsilon_1^4) < \infty$, the right-hand side of the inequality equals

$$\sup_{a>0} \limsup_{n\to\infty} \int_0^1 E\Big[n^{2k/(2k+1)}\big(\hat{G}(t, b^*) - g(t)\big)^2 \wedge a\Big]\,dt.$$

PROOF. We split the interval $[0,1]$ into the sets

$$T_1 := \{t \in [0,1] : |g^{(k)}(t)| \ge \delta\} \quad \text{and} \quad T_2 := [0,1] \setminus T_1.$$

Observing that $X_n - Y_n \to_p 0$ implies $(X_n \wedge a) - (Y_n \wedge a) \to_p 0$, we obtain by

Lemma 3.1 and the dominated convergence theorem for any $a > 0$,

$$\limsup_{n\to\infty} \int_{T_1} E\left[n^{2k/(2k+1)}(\hat{g}(t, b(\hat{\rho}_t)) - g(t))^2 \wedge a\right] dt$$
$$= \limsup_{n\to\infty} \int_{T_1} E\left[n^{2k/(2k+1)}(\hat{g}(t, b_t^*) - g(t))^2 \wedge a\right] dt$$
$$\leq \limsup_{n\to\infty} \int_{T_1} E\left[n^{2k/(2k+1)}(\hat{g}(t, b_t^*) - g(t))^2\right] dt,$$

and by the considerations following (2.8) in Section 2,

$$\sup_{a>0} \limsup_{n\to\infty} \int_{T_1} E\left[n^{2k/(2k+1)}(\hat{g}(t, b(\hat{\rho}_t)) - g(t))^2 \wedge a\right] dt$$
$$\leq \limsup_{n\to\infty} \int_{T_1} E\left[n^{2k/(2k+1)}(\hat{G}(t, b^*) - g(t))^2\right] dt.$$

If $E(\varepsilon_1^4) < \infty$, we get $E(n^{k/(2k+1)}(\hat{G}(t, b^*) - g(t)))^4 \leq c < \infty$ for all n, and hence

$$\lim_{a\to\infty} \limsup_{n\to\infty} E\left(n^{2k/(2k+1)}(\hat{G}(t, b^*) - g(t))^2 \vee a\right) = 0$$

[see Loève (1977), page 164ff.]. Therefore

$$\limsup_{n\to\infty} \int_{T_1} E\left[n^{2k/(2k+1)}(\hat{G}(t, b^*) - g(t))^2\right] dt$$
$$= \lim_{a\to\infty} \limsup_{n\to\infty} \int_{T_1} E\left[n^{2k/(2k+1)}(\hat{G}(t, b^*) - g(t))^2 \wedge a\right] dt.$$

For $t \in T_2$ observe that

$$\lim_{n\to\infty} n^{2k/(2k+1)} \mathrm{MSE}(\hat{g}(t, b(\tau_t))), \qquad \tau_t \in [r, s],$$

is smallest if $\tau_t = s$ and that $\hat{\rho}_t \to_p s$, as $n \to \infty$ on T_2. □

In case the kernel employed is the rectangular function (hence $k = 2$) the following stronger result can be derived. The restriction on the rectangular kernel is only a technical point.

THEOREM 3.3. *With the notation of this section we assume that* $E(\varepsilon_1^4) < \infty$, $K(t) = \chi_{[-1,1]}(t)/2$, $\int_0^1 (g^{(2)}(t))^2\, dt > 0$ *and* $r < \tilde{\tau} < s$.

(i) *Then*

$$\lim_{n\to\infty} \left\{ \int_0^1 \mathrm{MSE}(\hat{g}(t, b(\hat{\rho}_t)))\, dt / \mathrm{IMSE}(\hat{G}(t, b^*)) \right\} \leq 1.$$

(ii) *If, in addition,* $|g^{(k)}(t)| \geq \delta$ *on* $[0,1]$ *for some* $\delta > 0$, *then*

$$\lim_{n\to\infty} \left\{ \int_0^1 \mathrm{MSE}(\hat{g}(t, b(\hat{\rho}_t)))\, dt \Big/ \int_0^1 \mathrm{MSE}(\hat{g}(t, b_t^*))\, dt \right\} = 1.$$

The proof is based on

LEMMA 3.2. *Let (X_i) be a sequence of independent r.v.'s with $EX_i = 0$, $EX_i^2 = \sigma^2 < \infty$ and $E(|X_i|^\gamma) \le M < \infty$ for some $\gamma \ge 2$. If there exists a constant $c_\gamma > 0$ s.t.*

$$E\left(\left|\sum_{i=1}^{m} X_i\right|^\gamma\right) \le c_\gamma E\left(\sum_{i=1}^{m} X_i^2\right)^{\gamma/2}, \quad \text{for all } m,$$

then there exists a constant $d_\gamma > 0$ (independent of m) s.t.

$$E\left(\left|\max_{1\le j\le m} \sum_{i=1}^{j} X_i\right|^\gamma\right) \le d_\gamma E\left(\sum_{i=1}^{m} X_i^2\right)^{\gamma/2}; \quad \text{for all } m.$$

PROOF. Use the generalized Kolmogorov inequality, Loève (1977), page 275. □

LEMMA 3.3. *Under the assumptions of Theorem 3.3,*

$$E\left(n^{6/5} \max_{r\le\tau\le s}\left|\frac{1}{nb(\tau)}\sum_{i=1}^{n}\frac{1}{2}\chi_{[-1,1]}\left(\frac{t-t_i}{b(\tau)}\right)Y_i - g(t)\right|^3\right) = O(1), \tag{3.6}$$

as $n \to \infty$ for all $t \in [0,1]$.

PROOF. Let $I_n := \{i\colon tn - \tau n^{4/5} \le i \le tn + \tau n^{4/5}\}$ and define

$$Q_n(\tau) := \frac{1}{2nb(\tau)}\sum_{i\in I_n}\varepsilon_i, \qquad R_n(\tau) := \frac{1}{2nb(\tau)}\sum_{i\in I_n} g(t_i) - g(t).$$

The left-hand side of (3.6) is bounded by

$$\begin{aligned} &n^{6/5}E\left(\max_{r\le\tau\le s}|Q_n(\tau)+R_n(\tau)|^3\right)\\ &\quad\le n^{6/5}\Big[E^{3/4}\left(\max_\tau|Q_n(\tau)|^4\right) + 3\max_\tau|R_n(\tau)|E\left(\max_\tau|Q_n(\tau)|^2\right)\\ &\qquad + 3\max_\tau|R_n(\tau)|^2E^{1/2}\left(\max_\tau|Q_n(\tau)|^2\right) + \max_\tau|R_n(\tau)|^3\Big]. \end{aligned}$$

By (2.4),

$$\max_{r\le\tau\le s}\left|\frac{1}{nb(\tau)}\sum_{i\in I_n} g\left(\frac{i}{n}\right) - g(t)\right| = O(n^{-2/5}). \tag{3.7}$$

Furthermore, since the ε_i are i.i.d.,

$$\max_{r\le\tau\le s}|Q_n(\tau)| \equiv_D \max_{r\le\tau\le s}\left|\frac{1}{2\tau n^{4/5}}\sum_{i=1}^{2\tau n^{4/5}}\varepsilon_i\right| \le \frac{1}{2rn^{4/5}}\max_{2rn^{4/5}\le j\le 2sn^{4/5}}\left|\sum_{i=1}^{j}\varepsilon_i\right|.$$

Observing the independence of the ε_i, an application of Lemma 3.2 yields

$$E\Big(\max_{r \le \tau \le s} |Q_n(\tau)|^\gamma\Big) \le cn^{-2\gamma/5}, \quad \text{for } \gamma = 2,3,4. \tag{3.8}$$

(3.6) follows from (3.7) and (3.8). □

PROOF OF THEOREM 3.3. (3.6) implies that the sequence

$$S_n^2 = \big[n^{2/5}\max_{r \le \tau \le s}|\hat{g}(t, b(\tau)) - g(t)|\big]^2$$

is uniformly integrable. Hence it is possible to interchange limit $n \to \infty$ and expectation. Similar considerations as in Theorem 3.2 complete the proof. For part (ii) observe that the set T_2 defined in the proof of Theorem 3.2 is empty, and because of the tightness we do not need the truncation at a. □

4. Consistent local bandwidth selection. By the results of the preceding section, kernel estimators with data-based local bandwidth selection behave asymptotically well provided that the selection procedure is consistent. Here we assume that $g \in \mathscr{C}^k([0,1])$ and that a kernel $K \in \mathscr{M}_{0,k}$ of order k is used, on which we want to do bandwidth variation.

Assume that $\nu = 0$ and $k \ge 2$. The basic idea is to apply the formula

$$b_t^* = b^*\left(\frac{\int_0^1 g^{(k)}(x)^2\,dx}{g^{(k)}(t)^2}\right)^{1/(2k+1)}, \tag{4.1}$$

which follows from (2.6) and (2.7).

Assume that there are known constants $0 < r < s < \infty$ s.t. for $b^* = \tilde{\tau}n^{-1/(2k+1)}$ and $b_t^* = \tau_t^* n^{-1/(2k+1)}$ we have $r < \tilde{\tau} < s$ and $r < \inf_t \tau_t^*$.

LEMMA 4.1. *Suppose that*

(i) $\hat{b}$ *is a consistent estimator of* b^* *s.t.*

$$n^{1/(2k+1)}(\hat{b} - b^*) \to_p 0, \quad \textit{as } n \to \infty;$$

(ii) $\hat{g}^{(k)}$ *is a consistent estimate of* $g^{(k)}$ *s.t.*

$$n^{-1}\sum_{i=1}^n \{\hat{g}^{(k)}(t_i)\}^2 \to_p \int_0^1 g^{(k)}(t)^2\,dt, \quad \textit{as } n \to \infty.$$

Then, if

$$\tilde{\tau}_t := n^{1/(2k+1)}\hat{b}\left(\frac{1/n\sum_{i=1}^n \hat{g}^{(k)}(t_i)^2}{\hat{g}^{(k)}(t)^2}\right)^{1/(2k+1)}, \tag{4.2}$$

the estimator $\hat{\tau}_t := \tilde{\tau}_t \wedge s$ *satisfies*

$$\hat{\tau}_t \to_p \begin{cases} \tau_t^*, & \textit{if } \tau_t^* \le s, \\ s, & \textit{else}, \end{cases} \quad \textit{as } n \to \infty. \tag{4.3}$$

REMARK. (4.3) is just the requirement for the results of Section 3 which then apply to $\hat{g}(t, \hat{\tau}_t n^{-1/(2k+1)})$.

PROOF. If $|g^{(k)}(t)| > 0$,

$$\tau_t^* - \tilde{\tau}_t = n^{1/(2k+1)}(b^* - \hat{b})\left(\frac{\int_0^1 g^{(k)}(x)^2\,dx}{g^{(k)}(t)^2}\right)^{1/(2k+1)}$$
$$+ n^{1/(2k+1)}\hat{b}\left[\left(\frac{\int_0^1 g^{(k)}(x)^2\,dx}{g^{(k)}(t)^2}\right)^{1/(2k+1)} - \left(\frac{1/n\sum_{i=1}^n \hat{g}^{(k)}(t_i)^2}{\hat{g}^{(k)}(t)^2}\right)^{1/(2k+1)}\right]$$
$$\to_p 0,$$

by the assumptions. If $g^{(k)}(t) = 0$, $\tilde{\tau}_t \wedge s \to_p s$. The result follows. □

Lemma 4.1 can serve as the basis of many different consistent local bandwidth selection procedures. Condition (ii) is satisfied if, e.g.,

$$\sup_{t\in[0,1]} \left|\hat{g}^{(k)}(t) - g^{(k)}(t)\right| \to_p 0, \quad \text{as } n \to \infty. \tag{4.4}$$

The following result gives sufficient conditions.

THEOREM 4.1. *Assume that $\hat{b}$ is a consistent estimator of b^* s.t.*

$$n^{1/(2k+1)}(\hat{b} - b^*) \to_p 0, \qquad n \to \infty,$$

further that $g \in \mathscr{C}^k([0,1])$, the kernel $K_k \in \mathscr{M}_{k,k}$ used in $\hat{G}_k$ is Lipschitz continuous on $\mathbb{R}$ and that $nb^{2k+2}/(\log n)^2 \to \infty$ for a sequence $b \to 0$. Let

$$\hat{\tau} = n^{1/(2k+1)}\hat{b}\left[\frac{1/n\sum_{i=1}^n \hat{G}_k(t_i, b)^2}{\hat{G}_k(t, b)^2}\right]^{1/(2k+1)} \wedge s$$

and assume that the estimate $\hat{g}(t, b(\hat{\tau}))$ employs a kernel $K \in \mathscr{M}_{0,k}$. Then Theorems 3.1 and 3.2 apply to $\hat{g}(t, b(\hat{\tau}))$, and Theorem 3.3, if $K \equiv \frac{1}{2}\chi_{[-1,1]}$ (or is of simple structure).

PROOF. By Lemma 2.1 (4.4) is satisfied and therefore Lemma 4.1 applies. □

5. A specific procedure and simulation results. In the following, we describe a procedure which allows bandwidth variation for small-to-medium sample sizes as they are usually encountered in practical applications. We assume $g \in \mathscr{C}^{k+2}([0,1])$ and estimate curve and second derivative with kernels of the orders k, respectively, $(k+2)$. We worked out an example for $k = 2$ and found that the performance of the procedure depends on the choice of kernels of the correct orders and of sufficient smoothness. The kernels chosen have been described in Müller (1984) and satisfy the smoothness requirements of Rice (1984).

Step 1. Compute the minimizer $\hat{b}$ of the Rice criterion, using the kernel $P_{0,2,3} \equiv \frac{35}{32}(1 - 3x^2 + 3x^4 - x^6)\chi_{[-1,1]}$ which is in $\mathscr{M}_{0,2}$ [according to (2.2), Lemma 4.1(i) is satisfied].

Step 2. Compute the minimizer $\tilde{b}$ of the Rice criterion, using the kernel $P_{0,4,3} \equiv \frac{315}{512}(3 - 20x^2 + 42x^4 - 36x^6 + 11x^8)\chi_{[-1,1]}$ which is in $\mathscr{M}_{0,4}$.

Step 3. Compute $b(\hat{\tau}_t) = \hat{\tau}_t n^{-1/5}$ according to (4.2), substituting $\hat{g}^{(2)}(t) = \hat{G}_2(t, \tilde{b}d_{2,4})$, where $d_{2,4} = 0.9$ is determined according to (2.11). Choose a sufficiently large cut-off to determine $b(\hat{\tau}_t)$. Hence $\tilde{b}d_{2,4}n^{1/9} \to_p b^{(2)*}n^{1/9}$ [any choice $b = \tau n^{-1/9}$, $\tau > 0$, implies (4.4) and hence (ii) in Lemma 4.1].

Step 4. Estimate $g(t)$ by $\hat{g}(t, b(\hat{\tau}_t))$.

Simulations for this procedure were based on the following seven curves of different type:

$$g_1(x) = h(0.5, 0.1) + 16x(1 - x),$$

where $h(\mu, \sigma)$ is a Gaussian density with mean μ and variance σ^2;

$$g_2(x) = h(0.25, 0.05) + h(0.5, 0.1);$$

$$g_3(x) = g_2(x) + (1 - 2x - \log(x + 0.6));$$

$$g_4(x) = 6x - 2;$$

$$g_5(x) = \begin{cases} 16x, & 0 \le x \le 0.5, \\ 16(1 - x), & 0.5 \le x \le 1; \end{cases}$$

$$g_6(x) = \begin{cases} 8x + 4, & 0 \le x \le 0.25, \\ -20x + 11, & 0.25 \le x \le 0.5, \\ 12x - 5, & 0.5 \le x \le 0.75, \\ -24x + 22, & 0.75 \le x \le 1; \end{cases}$$

$$g_7(x) = 4\sin(2\pi x).$$

These curves were computed on $[-0.5, 1.5]$ and observations were extended accordingly in order to avoid boundary effects. Curves g_5, g_6 were included as examples of nondifferentiable functions. Curve g_4 was included since $g_4'' \equiv 0$ which is an interesting case in view of (2.6)–(2.10). No gain for local bandwidth choice can be expected in this case. We allowed for $n = 50$ observations per curve and the results were averaged over $N = 125$ random samples. The noise variables generated were normal with standard deviation $\sigma = 0.1(\max_t g(t) - \min_t g(t))$.

We found that the estimation of residual variance necessary for the Rice method works sufficiently well if constants are fitted to successive triples of points. In general, the Rice method for global bandwidth choice produced good results for all curves. An impression is given in Table 1, where we used the kernel $P_{0,4,1} \equiv \frac{15}{32}(3 - 10x^2 + 7x^4)\chi_{[-1,1]}$.

The last column of Table 1 displays the finitely optimal global bandwidth (w.r.t. IMSE). This bandwidth was computed by the finite evaluation technique described in Gasser et al. (1984): If the true curve g and σ are given, for any linear estimate $\hat{g}(t) = \sum_1^n W_i Y_i$ of $g(t)$ we compute

$$\operatorname{var} \hat{g}(t) = \sigma^2 \sum_1^n W_i^2 \quad \text{and} \quad \operatorname{Bias}(\hat{g}(t)) = \sum_1^n W_i g(t_i) - g(t),$$

TABLE 1
Quality of estimated global bandwidth $\hat{b}$ by the Rice method(kernel $P_{0,4,1}$), $N = 125$

$g_i(x)$	Mean of $\hat{b}$	Variance of $\hat{b}$	Optimal value w.r.t. IMSE (finite evaluation)
g_1	0.203	0.003	0.21
g_2	0.107	0.00036	0.11
g_3	0.107	0.00036	0.11
g_4	0.365	0.012	0.45
g_5	0.264	0.0028	0.27
g_6	0.20	0.0036	0.22
g_7	0.28	0.006	0.27

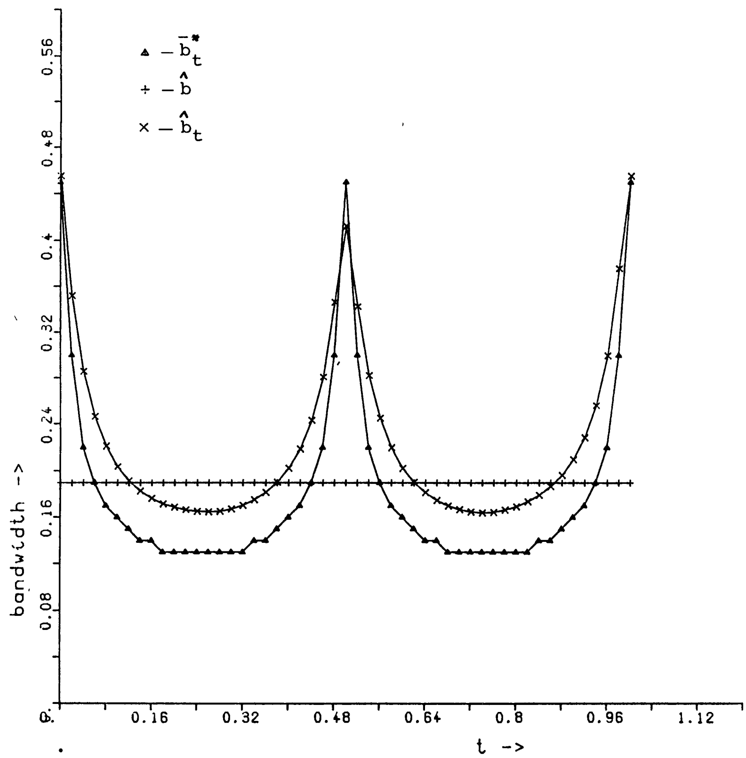

FIG. 1. *Estimated local bandwidths ($\hat{b}_t$), finitely optimal local bandwidths ($\bar{b}_t^*$) and estimated global bandwidth ($\hat{b}$). Results for $\hat{b}_t$, $\hat{b}$ for one sample; curve is g_7.*

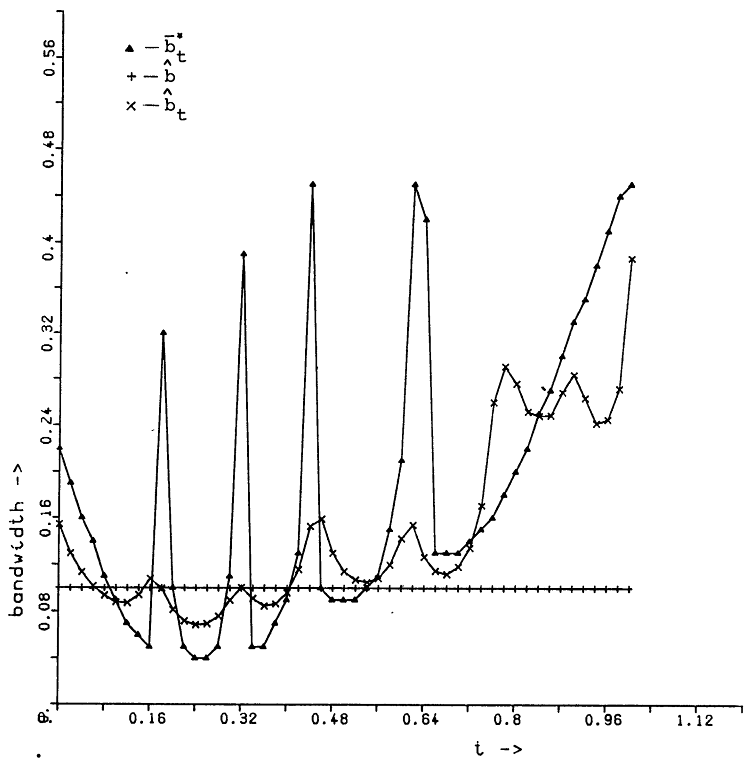

FIG. 2. *Same as Figure* 1, *but for* g_2 (*one sample*).

TABLE 2

Relative gains of IMSE of variable bandwidth estimator versus global bandwidth estimator, $N = 125$

$g_i(x)$	Optimal bandwidths[a]	Estimated bandwidths[b]	Estimated bandwidths, comparison with higher order kernel[c]
g_1	29.6	11.1	3.8
g_2	42.2	23.4	17.3
g_3	42.3	23.3	17.4
g_4	0.0	−0.3	31.0
g_5	49.0	28.3	26.9
g_6	32.7	14.9	10.7
g_7	9.5	7.7	−42.6

[a]Possible gain in percent of IMSE($g(t, b_t^*)$) against IMSE($G(\cdot, b^*)$) using kernel $P_{0,2,3}$.
[b]Average gain in percent of IMSE($\hat{g}(t, \hat{b}_t)$) against IMSE($G(\cdot, \hat{b})$) using kernel $P_{0,2,3}$.
[c]Average gain in percent of IMSE($g(t, \hat{b}_t)$) using kernel $P_{0,2,3}$ against IMSE($\hat{G}(\cdot, \hat{b})$) using kernel $P_{0,4,1}$.

and thus we can evaluate IMSE $(\hat{G}(\cdot, b))$. The minimizers w.r.t. b, b_t are denoted by $\bar{b}^*$, $\bar{b}_t^*$, called the finitely optimal bandwidths.

The results for g_4 must be interpreted with care, since bandwidths were cut off at 0.45 in order to avoid boundary effects (as data were available in $[-0.5, 1.5]$). It turned out that for the present application it is slightly better to oversmooth $\hat{G}_2(t)$ somewhat. Therefore we chose $d_{2,4} = 1.1$ instead of the correct value $d_{2,4} = 0.9$ (Step 3).

We investigated the behavior of estimated local bandwidths $\hat{b}_t$ compared to $\bar{b}_t^*$ for $t \in [0,1]$. Figures 1 and 2 show these bandwidths and the global bandwidth estimate $\hat{b}$ for the functions g_2 and g_7.

The gains (in percent) of our local bandwidth choice as compared to the Rice method with the comparable kernel $P_{0,2,3} \in \mathcal{M}_{0,2}$ and with the higher-order kernel $P_{0,4,1} \in \mathcal{M}_{0,4}$ are given in Table 2. The first column contains the gain using finitely optimal local, respectively, global bandwidths (computed by the finite evaluation method), using the kernel $P_{0,2,3}$. These gains are close to the asymptotic values given by (2.8) and (2.9). Results for g_4 again must be interpreted with caution, since for the kernel $P_{0,4,1}$ sometimes larger bandwidths than 0.45 would have been optimal. The second column shows that for these

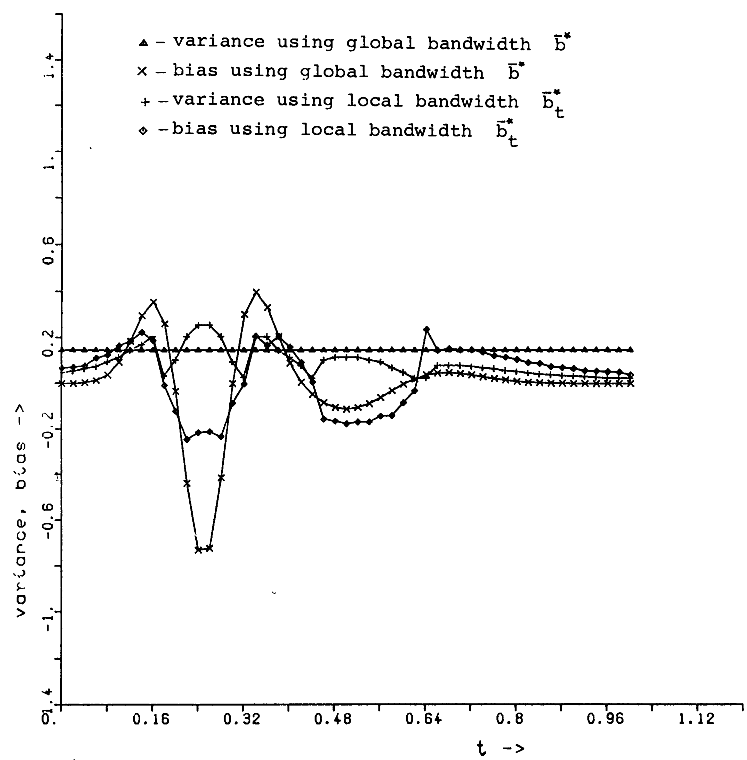

FIG. 3. *Local variance and bias using finitely optimal local bandwidths $\bar{b}_t^*$, respectively, finitely optimal global bandwidth $\bar{b}^*$.*

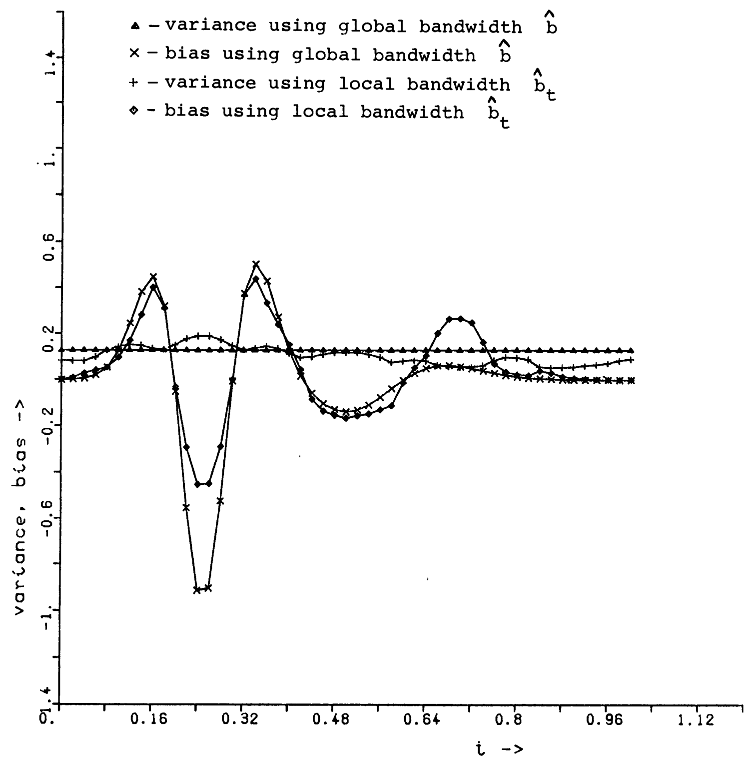

FIG. 4. *Local variance and bias using estimated local bandwidths $\hat{b}_t$, respectively, estimated global bandwidth $\hat{b}$. Same sample as in Figure* 2.

examples, data-based local bandwidth choice is always better than global bandwidth choice and generally better than using higher-order kernels. An exception are sinus-type functions, where $|g^{(4)}|$ is of the same order of magnitude as $|g^{(2)}|$, which implies by (2.4) that bias is decreased appreciably if higher-order kernels are used, since usually $|\beta_{0,4}| \ll |\beta_{0,2}|$. In such cases it is preferable to use a higher-order kernel supplied with a global bandwidth estimate.

In Figure 4 we compare local variance and bias of kernel estimators with global bandwidth choice (Rice) and our local bandwidth selection procedure for g_2 for one sample. Figure 3 shows the same quantities, if finitely optimal bandwidths are used. As is to be expected, the local procedure reduces bias near peaks and variance in flat parts of the curve. For the same sample as in Figure 4, Figure 5 displays observations, true curve and estimated curves with global and local bandwidth selection.

6. Discussion. We assumed that the regression model (1.1) is homoscedastic. A referee pointed out that in view of (2.6), local bandwidth variation would be of particular interest for heteroscedastic models [i.e., $E\varepsilon_i^2 = \sigma^2(t_i)$ for some smooth function $\sigma^2(\cdot)$]. A closer analysis of (2.8) and (2.9) and application of

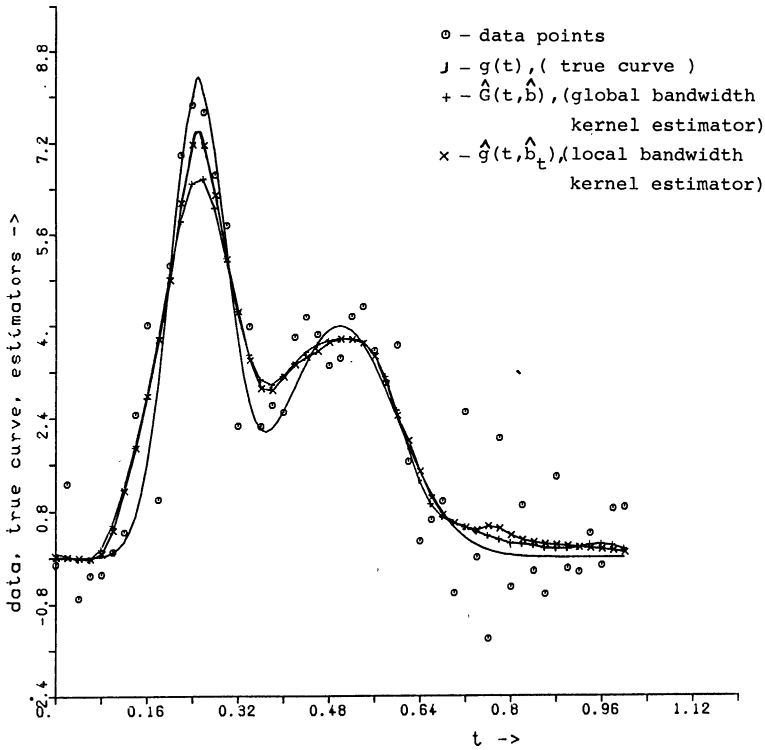

FIG. 5. *True curve g_2, data points and estimated curves by local/global bandwidth kernel estimators. Same sample as in Figure 2.*

Hölder's inequality shows indeed that in this case we obtain

$$\text{IMSE (local optimal bandwidths)/IMSE (global optimal bandwidth)}$$
$$= \frac{\int_0^1 \left(\left[\sigma^2(t)\right]^{2(k-\nu)} \left[g^{(k)}(t)^2\right]^{(2\nu+1)} \right)^{1/(2k+1)} dt}{\left(\left[\int_0^1 \sigma(t)^2\,dt\right]^{2(k-\nu)} \left[\int_0^1 g^{(k)}(t)^2\,dt\right]^{(2\nu+1)} \right)^{1/(2k+1)}}$$
$$\le 1, \quad \text{provided } g^{(k)}(t)^2 > 0 \text{ on } [0,1].$$

Therefore adaption to heteroscedasticity is expected to yield similar gains as those by adaption to local curvature. Carroll (1982) proposes to estimate $\sigma(t)^2$ by means of Priestley–Chao kernel estimators applied to squared residuals in the context of estimating weights for linear regression. A similar procedure could be thought of in the present situation. Any uniformly consistent estimate of local variance allows to extend Lemma 4.1 and Theorem 4.1 to the heteroscedastic case [applying an extended version of (4.1)].

Another important point concerns choice of the orders of the kernels used. Theorem 4.1 allows to perform bandwidth variation on a kth-order kernel if

$g \in \mathscr{C}^k([0,1])$ and therefore no loss is incurred in the asymptotic rate of convergence and constants are improved. But this requires estimating $g^{(k)}$ by a kth-order kernel which yields a poorly behaved estimate for small sample sizes. In order to overcome this difficulty, we employed in the simulations and the finite sample procedure of Section 5 a $(k+2)$th-order estimate for $g^{(k)}$ and a kth-order estimate for g. As the referee remarked, there might be two problems with this procedure: First, it is not clear, whether the $(k+2)$th-order estimate of $g^{(k)}$ suffices to make bandwidth variation competitive against the possibilities of kth- as well as $(k+2)$th-order global bandwidth estimation; and second, this procedure is suboptimal from an asymptotic point of view since the rate of convergence is better for the ordinary $(k+2)$th-order estimate of g than for the kth-order variable bandwidth estimate.

For the first point, we refer to the simulation results displayed in Table 2, columns 2 and 3, which show that for rather different curves the method is competitive. To address the second point, we want to emphasize that the procedure of Section 5 was specifically devised for the small sample situation. For large samples, bandwidth variation can be based on Theorem 4.1 (see above). A realistic asymptotic theory for our specific procedure (and of kernel estimation in general) would have to take into account varying the orders of kernels with increasing number of observations. A definite evaluation of our specific procedure is not possible until such a theory exists.

Acknowledgments. We wish to thank Friedrich Löffler for help with the simulation and John Rice for providing us with pre-publication versions of his papers. Further thanks are due to a referee for his detailed comments and his criticisms, which are discussed in Section 6.

REFERENCES

ABRAMSON, I. (1982a). Arbitrariness of the pilot estimator in adaptive kernel methods. *J. Multivariate Anal.* **12** 562–567.

ABRAMSON, I. (1982b). On bandwidth variation in kernel estimates—a square root law. *Ann. Statist.* **10** 1217–1223.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BREIMAN, L., MEISEL, W. and PURCELL, E. (1977). Variable kernel estimates of probability densities. *Technometrics* **19** 135–144.

CARROLL, R. J. (1982). Adapting for heteroscedasticity in linear models. *Ann. Statist.* **10** 1224–1233.

CHENG, K. F. and LIN, P. E. (1981). Nonparametric estimation of a regression function. *Z. Wahrsch. verw. Gebiete* **57** 223–233.

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.* **31** 377–403.

GASSER, TH. and MÜLLER, H. G. (1984). Nonparametric estimation of regression functions and their derivatives. *Scand. J. Statist.* **11** 171–185.

GASSER, TH., MÜLLER, H. G., KÖHLER, W., MOLINARI, L. and PRADER, A. (1984). Nonparametric regression analysis of growth curves. *Ann. Statist.* **12** 210–229.

KRIEGER, A. M. and PICKANDS, J. (1981). Weak convergence and efficient density estimation at a point. *Ann. Statist.* **9** 1066–1078.

LOÈVE, M. (1977). *Probability Theory* **1**, 4th ed. Springer, New York.

MÜLLER, H. G. (1983). Consistency of fixed design nonparametric regression. Manuscript. (See also Beiträge zur nichtparametrischen Kurvenschätzung. Dissertation, Univ. Ulm.)

MÜLLER, H. G. (1984). Smooth optimum kernel estimators of regression curves, densities and modes. *Ann. Statist.* **12** 766–774.

PRIESTLEY, M. B. and CHAO, M. T. (1972). Nonparametric function fitting. *J. Roy. Statist. Soc. Ser. B* **34** 385–392.

RICE, J. (1983). Methods for bandwidth choice in nonparametric kernel regression. In *Computer Science and Statistics: The Interface* (J. E. Gentle, ed.) 186–190. North-Holland, Amsterdam.

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230.

RICE, J. (1986). Bandwidth choice for differentiation. *J. Multivariate Anal.* **19** 251–264.

SILVERMAN, B. W. (1978). Choosing a window width when estimating a density. *Biometrika* **65** 1–11.

STADTMÜLLER, U. (1982). Nichtparametrische Schätzung einer Regressionsfunktion in einem Modell mit festem Meßdesign. Habilitationsschrift, Univ. Ulm.

STADTMÜLLER, U. (1986). Asymptotic properties of nonparametric curve estimates. *Period. Math. Hungar.* **17** 83–108.

STONE, C. J. (1977). Consistent nonparametric regression (with discussion). *Ann. Statist.* **5** 595–645.

WOODROOFE, M. (1970). On choosing a delta sequence. *Ann. Math. Statist.* **41** 1665–1671.

INSTITUT FÜR MEDIZINISCH-BIOLOGISCHE STATISTIK
UNIVERSITÄT MARBURG
ERNST-GILLER-STRASSE 20
D-3550 MARBURG
FEDERAL REPUBLIC OF GERMANY

ABTEILUNG FÜR MATHEMATIK I
UNIVERSITÄT ULM
OBERER ESELSBERG
D-7900 ULM
FEDERAL REPUBLIC OF GERMANY