*The Annals of Statistics*
1989, Vol. 17, No. 3, 1053–1069

# ADAPTIVE NONPARAMETRIC PEAK ESTIMATION[1]

BY HANS-GEORG MÜLLER

*University of California, Davis*

It is shown that consistent estimates of the optimal bandwidths for kernel estimators of location and size of a peak of a regression function are available. Such estimates yield the same joint asymptotic distribution of location and size of a peak as the optimal bandwidths themselves. Therefore data-adaptive efficient estimation of peaks is possible. In order to prove this result, the weak convergence of a two-dimensional stochastic process with appropriately scaled bandwidths as arguments to a Gaussian limiting process is shown. A practical method which leads to consistent estimates of the optimal bandwidths and is therefore asymptotically efficient is proposed and its finite sample properties are investigated by simulation.

**1. Introduction.** The nonparametric estimation of peaks of a regression function is a reasonable approach whenever we do not have much knowledge of the form of the regression function besides that there is a peak whose coordinates are of interest. Obtaining information about the location and size of such a peak is sometimes a central issue in practical curve estimation. In longitudinal studies we can use these "longitudinal parameters" to classify and compare the nonparametric curve estimates for different subjects; see Jørgensen, Nielsen, Keiding and Skakkeback (1985), where this idea is applied to longitudinal endocrinological data, Silverman (1985), where the estimation of peaks of bacterial growth curves is considered and Müller (1985), where results on consistency and the asymptotic distribution of estimated peaks are derived and further references can be found. Estimated peaks can replace ordinary parameters in parametric regression models for the purposes of classification and discrimination of samples of curves and often have the additional advantage of being interpretable from a subject-matter point of view.

We consider kernel estimates of a regression function in the fixed design regression model, i.e., the given measurements are $(x_i, y_i)$, $i = 1, \dots, n$, where the $(x_i)$ are fixed sites of measurement and the $(y_i)$ are the outcomes of noisy measurements. The following model is assumed:

$$y_{i,n} = g(x_{i,n}) + \varepsilon_{i,n}, \qquad i = 1, \dots, n, \tag{1.1}$$

where $g$ is the unknown regression function assumed to be in $\mathscr{C}^k([0,1])$ for some $k \geq 0$ and $(\varepsilon_{i,n})$ are the measurement errors which are assumed to form a triangular array of i.i.d. random variables with expectation 0 and variance $\sigma^2$.

Received September 1986; revised February 1988.
[1]Research supported in part by Deutsche Forschungsgemeinschaft.
*AMS* 1980 *subject classifications*. 62G05, 62G20.
*Key words and phrases*. Kernel estimator, choice of bandwidths, efficiency, weak convergence, tightness, Gaussian process, size and location of peaks, nonparametric regression.

Abbreviating, we write $\varepsilon_i$ instead of $\varepsilon_{i,n}$ (and analogously for $y_i$ and $x_i$). We require that $0 \le x_1 \le x_2 \le \cdots \le x_n \le 1$.

Given any nonparametric estimates $\hat{g}(\cdot)$ of $g(\cdot)$, we can find estimates of location and size of a peak by reading location $\hat{\theta}$ and size $\hat{g}(\hat{\theta})$ off the estimate $\hat{g}(\cdot)$. With $\hat{g}(\cdot)$ being the kernel estimate,

$$\hat{g}(x) = \frac{1}{b} \sum_{i=1}^{n} \int_{d_{i-1}}^{d_i} K\left(\frac{x-u}{b}\right) du\, y_i, \tag{1.2}$$

where $d_0 = 0$, $d_n = 1$ and $x_i \le d_i \le x_{i+1}$, $1 \le i \le n-1$, $K$ is the kernel function and $b$ is the smoothing parameter, called bandwidth. This approach was investigated in Müller (1985). There the question of bandwidth choice for peak estimation was left open. However, this is a crucial problem for practical applications, especially the estimated size of a peak depends strongly on the chosen bandwidth. Enlarging the bandwidth usually decreases the size of a peak (oversmoothing). Since a peak in a curve is a local phenomenon, global bandwidth choice by cross-validation or related criteria like the one proposed by Rice (1984a) [compare also the recent work of Burman (1985), Hall (1983), Marron (1985) and Stone (1984) for density estimation as well as Härdle and Marron (1985), Speckman (1985) and Wahba and Wold (1975) for nonparametric regression] may not be appropriate. It is then a natural approach to resort to local bandwidth choice specifically designed for the estimation of peaks.

It can be shown that the asymptotic distributions of $\hat{\theta}$ and $\hat{g}(\hat{\theta})$, appropriately scaled, are normal. The mean squared errors (MSE), derived from the asymptotic bias and variance of the limiting distributions, can be minimized w.r. to the bandwidths, separately for location and size of a peak. Inserting the resulting optimal local bandwidths, we obtain the optimal joint limiting distribution of location and size. The optimal bandwidths depend on unknown quantities like some higher derivative of the regression function $g$ and the variance $\sigma^2$. The question arises whether the optimal limiting distribution can be achieved by a fully data-adaptive method, i.e., if an efficient estimator in this sense exists. By showing that a two-dimensional stochastic process with appropriately scaled deviations of location and size, having the respective bandwidths as arguments, converges weakly to a continuous Gaussian limiting process, we conclude that it suffices to produce consistent estimates of the optimal bandwidths for location and size to achieve efficiency. Such estimators are plentiful and the optimal limiting distribution will then be in force.

The main results are compiled in Section 3, whereas Section 2 contains some preliminaries and definitions. Auxiliary results and proofs are in Section 4. One special data-adaptive method which produces consistent estimates for the optimal bandwidths is investigated in Section 5. We show by simulation that this method is advantageous compared to a global bandwidth choice for peak estimation. Application of weak convergence methods in curve estimation was introduced by Abramson (1982) and Krieger and Pickands (1981), who analysed local bandwidth choice in density estimation, and was also used by Bhattacharya and Mack (1987) and Müller and Stadtmüller (1987a) in nonparametric regression.

**2. Stochastic processes connected with peak estimation.** We need some notation. Denote weak convergence in a function space [compare Billingsley (1968)] by $\Rightarrow$ as $n \to \infty$. We write $\to_{\mathscr{D}}$ and $\to_p$ to denote convergence in distribution (resp. in probability) as $n \to \infty$; the bounds $o(\cdot)$ and $O(\cdot)$ and the bounds in probability $o_p(\cdot)$ and $O_p(\cdot)$ are always as $n \to \infty$. In the following, we apply kernel functions $K$ satisfying for some integer $k \geq 0$,

$$\text{Support}\,(K) \subset [-1,1],$$
$$(2.1) \qquad \int K(x)x^j\,dx = \begin{cases} 1 & j = 0, \\ 0 & 0 < j < k, \end{cases} \qquad K \in \mathscr{C}([-1,1]),$$

and denote the class of kernels satisfying (2.1) by $\mathscr{M}_k$. The following kernel estimates with special bandwidths are considered:

$$(2.2) \quad \hat{g}_j(x,s) = \frac{1}{sn^{-1/(2(k+j)+1)}} \sum_{i=1}^{n} \int_{d_{i-1}}^{d_i} K\left(\frac{x-u}{sn^{-1/(2(k+j)+1)}}\right) du\, y_i, \qquad j \geq 0.$$

For the sake of simplicity it is assumed that the interpolating sequence $(d_i)$ can be chosen equidistantly, i.e.,

$$(2.3) \qquad d_i - d_{i-1} = 1/n, \qquad i = 1,\ldots,n.$$

This assumption can be weakened considerably by introducing a design density which describes asymptotic nonequidistancy.

We assume that the curve $g$ has a unique global maximum at some $\theta \in I \subset (0,1)$, where $I$ is a compact interval chosen in order to avoid the discussion of end effects arising in nonparametric curve estimation near the endpoints 0 and 1 [see Rice (1984b) or Gasser and Müller (1984)], i.e., $g(x) < g(\theta)$ for all $x \neq \theta$, $x \in [0,1]$. As estimators of the location of the peak we consider

$$(2.4) \qquad \hat{\theta}(s) = \inf\left\{x \in I\colon \hat{g}_1(x,s) = \max_{0 \leq x \leq 1} \hat{g}_1(x,s)\right\}$$

for $s \in [s_a, s_b]$, where $0 < s_a < s_b < \infty$ are given [compare Müller (1985)]. As estimators of the size $g(\theta)$ of the peak, we use $\hat{g}_0(\hat{\theta}(s), t)$ for $t \in [t_a, t_b]$, where $0 < t_a < t_b < \infty$. Note that the scaling of the bandwidth for the estimation of the size of the peak is different from the scaling of the bandwidth for the estimation of the location. These scalings are reasonable by Lemmas 8 and 9 of Section 4 which show that under some mild regularity conditions, using twice continuously differentiable kernels of order $k$,

$$(2.5) \qquad n^{k/(2k+3)}\big(\hat{\theta}(s) - \theta\big) \to_{\mathscr{D}} \mathscr{N}\left(-\frac{s^k g^{(k+1)}(\theta) B_k}{g^{(2)}(\theta)}, \frac{\sigma^2 V^{(1)}}{s^3 g^{(2)}(\theta)^2}\right),$$

where

$$(2.6) \qquad B_k = (-1)^k \int K(v) v^k\,dv/k!, \qquad V^{(1)} = \int K^{(1)}(v)^2\,dv$$

and

$$n^{k/(2k+1)}\left(\hat{g}_0(\hat{\theta}(s), t) - g(\theta)\right) \to_{\mathscr{D}} \mathscr{N}\left(t^k g^{(k)}(\theta) B_k, \frac{\sigma^2 V}{t}\right), \tag{2.7}$$

where

$$V = \int K(v)^2 \, dv \tag{2.8}$$

[compare Parzen (1962) and Eddy (1980) for similar results in density estimation]. It should be noted that $s_a$, $s_b$, $t_a$, $t_b$ have to be chosen in such a way that the corresponding intervals contain certain "optimal" values $s^*$, $t^*$ [see (2.11) and (2.12) below].

From these limiting distributions we derive the asymptotic mean squared errors

$$\lim_{n\to\infty} n^{2k/(2k+3)} E\left(\hat{\theta}(s) - \theta\right)^2 = \frac{\sigma^2 V^{(1)}}{s^3 g^{(2)}(\theta)^2} + s^{2k} \frac{g^{(k+1)}(\theta)^2}{g^{(2)}(\theta)^2} B_k^2 \tag{2.9}$$

and

$$\lim_{n\to\infty} n^{2k/(2k+1)} E\left(\hat{g}_0(\hat{\theta}(s), t) - g(\theta)\right)^2 = \frac{\sigma^2 V}{t} + t^{2k} g^{(k)}(\theta)^2 B_k^2. \tag{2.10}$$

Assuming $g^{(k)}(\theta) \neq 0$, $g^{(k+1)}(\theta) \neq 0$ and minimizing w.r. to $s$ (resp. $t$) yields

$$s^* = \left(\frac{3\sigma^2 V^{(1)}}{2k g^{(k+1)}(\theta)^2 B_k^2}\right)^{1/(2k+3)}, \quad \text{respectively,} \tag{2.11}$$

$$t^* = \left(\frac{\sigma^2 V}{2k g^{(k)}(\theta)^2 B_k^2}\right)^{1/(2k+1)} \tag{2.12}$$

These are the optimal constants for the bandwidths, but $s^*$, $t^*$ contain the unknowns $\sigma^2$, $g^{(k)}(\theta)$ and $g^{(k+1)}(\theta)$ and are therefore not known either.

In order to investigate the behavior of the deviations $(\hat{\theta}(s) - \theta)$ and $\hat{g}_0(\hat{\theta}(s), t) - g(\theta)$ jointly, we define the two-dimensional random process

$$\begin{pmatrix} X_n(s) \\ Y_n(s,t) \end{pmatrix} := \begin{pmatrix} n^{k/(2k+3)}\left(\hat{\theta}(s) - \theta\right) \\ n^{k/(2k+1)}\left(\hat{g}_0(\hat{\theta}(s), t) - g(\theta)\right) \end{pmatrix} \tag{2.13}$$

on $[s_a, s_b] \times [t_a, t_b]$.

We are interested in the limiting distributions of $\begin{pmatrix} X_n(s) \\ Y_n(s,t) \end{pmatrix}$, especially for $s = s^*$ and $t = t^*$ and for consistent estimators $\hat{s} \to_p s^*$ and $\hat{t} \to_p t^*$. The theory developed in this paper also covers the case of the estimation of a minimum and can also be applied to the estimation of extrema of some derivative of $g$. In this case we would employ kernel estimators for the derivatives of $g$ [see Gasser and Müller (1984)].

**3. A functional limit theorem.** The following assumption will be needed repeatedly.

ASSUMPTION A. Let $g \in \mathscr{C}^{k+1}([0,1])$ for some $k \geq 1$, $E|\varepsilon_1|^r < \infty$ for some $r > 2 + 2/k$ and $K \in \mathscr{M}_k \cap \mathscr{C}^2$. Let $K^{(2)}$ be Lipschitz continuous on $\mathbb{R}$ and $g^{(2)}(\theta) < 0$.

In the following, we refer to the function space $D_q$ described by Bickel and Wichura [(1971), Section 3] for $q = 2$. For $q = 1$ this coincides with the space $D$ [see Billingsley (1968)].

THEOREM. *Under Assumption* A *and if the kernel K is symmetric,*

$$\begin{pmatrix} X_n(s) \\ Y_n(s,t) \end{pmatrix} \Rightarrow \begin{pmatrix} X(s) \\ Y(t) \end{pmatrix} \quad \text{on } D([s_a, s_b]) \times D_2([s_a, s_b] \times [t_a, t_b]),$$

*where the limiting process is a continuous Gaussian process with expectation*

$$EX(s) = -s^k g^{(k+1)}(\theta) B_k / g^{(2)}(\theta),$$

$$EY(t) = t^k g^{(k)}(\theta) B_k$$

[*for $B_k$ see* (2.6)] *and with covariance structure*

$$\operatorname{cov}(X(s_1), X(s_2)) = \frac{\sigma^2}{s_1^2 s_2^2 g^{(2)}(\theta)^2}\left(\int K^{(1)}\left(\frac{v}{s_1}\right) K^{(1)}\left(\frac{v}{s_2}\right) dv + o(1)\right),$$

$$\operatorname{cov}(Y(t_1), Y(t_2)) = \frac{\sigma^2}{t_1 t_2}\left(\int K\left(\frac{v}{t_1}\right) K\left(\frac{v}{t_2}\right) dv + o(1)\right),$$

$$\operatorname{cov}(X(s_1), Y(t_2)) = 0$$

*for any $s_1, s_2 \in [s_a, s_b]$ and $t_1, t_2 \in [t_a, t_b]$.*

*Especially, for any given $\alpha, \eta > 0$, there exist $\delta > 0$ and $n_0 \in \mathbb{N}$ so that for any $n > n_0$ we have simultaneously*

$$P\left(\sup_{\substack{|t_1 - t_2| < \delta,\ t_1, t_2 \in [t_a, t_b] \\ |s_1 - s_2| < \delta,\ s_1, s_2 \in [s_a, s_b]}} |Y_n(s_1, t_1) - Y_n(s_2, t_2)| > \alpha\right) < \eta \tag{3.1}$$

*and*

$$P\left(\sup_{|s_1 - s_2| < \delta,\ s_1, s_2 \in [s_a, s_b]} |X_n(s_1) - X_n(s_2)| > \alpha\right) < \eta. \tag{3.2}$$

The proof is in Section 4. Using Slutsky's theorem we obtain the following as an immediate consequence.

COROLLARY 1. *Under Assumption* A, *assuming K is symmetric,*

$$\begin{pmatrix} X_n(s) \\ Y_n(s,t) \end{pmatrix} \to_{\mathscr{D}} \mathscr{N}\left( \begin{bmatrix} -s^k g^{(k+1)}(\theta) B_k / g^{(2)}(\theta) \\ t^k g^{(k)}(\theta) B_k \end{bmatrix}, \sigma^2 \begin{bmatrix} V^{(1)}/g^{(2)}(\theta)^2 s^4 & 0 \\ 0 & V/t^2 \end{bmatrix} \right) \tag{3.3}$$

*for* $s \in [s_a, s_b]$ *and* $t \in [t_a, t_b]$, [*for* $B_k$, $V$ *and* $V^{(1)}$, *see* (2.6) *and* (2.8)]. *For consistent estimates* $\hat{\sigma} \to_p \sigma$ *and* $\hat{g}^{(j)}(\hat{\theta}) \to_p g^{(j)}(\theta)$ *for* $j = 2, k, (k+1)$, *we obtain*

$$\frac{1}{\hat{\sigma}} \begin{pmatrix} |\hat{g}^{(2)}(\hat{\theta})| s^2 / V^{(1)1/2} & 0 \\ 0 & t/V^{1/2} \end{pmatrix} \times \left( \begin{bmatrix} X_n(s) \\ Y_n(s,t) \end{bmatrix} - \begin{bmatrix} -s^k \hat{g}^{(k+1)}(\hat{\theta}) B_k / \hat{g}^{(2)}(\hat{\theta}) \\ t^k \hat{g}^{(k)}(\hat{\theta}) B_k \end{bmatrix} \right) \to_{\mathscr{D}} \mathscr{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, I \right), \tag{3.4}$$

*where I is the identity matrix.*

Applying this result we can construct asymptotic confidence regions for peaks, assuming that consistent estimates $\hat{\sigma} \to_p \sigma$ and $\hat{g}^{(2)} \to_p g^{(2)}(\theta)$ are available and that we either can neglect the bias in the asymptotic distribution since $B_k$ or $g^{(k+1)}(\theta)$ is small, or that further consistent estimates $\hat{g}^{(k)}(\hat{\theta}) \to_p g^{(k)}(\theta)$ and $\hat{g}^{(k+1)}(\hat{\theta}) \to g^{(k+1)}(\theta)$ are employed. For $\hat{\sigma}$, we adopt a proposal of Rice (1984a),

$$\hat{\sigma}_R^2 = \frac{1}{2(n-1)} \sum_{i=2}^{n} (y_i - y_{i-1})^2. \tag{3.5}$$

If the regression function $g$ is Lipschitz continuous, it follows immediately by the law of large numbers that

$$\hat{\sigma}_R \to_p \sigma. \tag{3.6}$$

The following lemma shows the existence of consistent estimators $\hat{g}^{(j)}(\hat{\theta}) \to g^{(j)}(\theta)$.

LEMMA 1. *Under Assumption* A, *setting* $b = tn^{-1/(2(k+\nu)+1)}$, *assuming for some* $0 \le \nu \le k+1$ *that* $K \in \mathscr{M}_{k+1} \cap \mathscr{C}^\nu$, $K^{(\nu)}$ *is Lipschitz continuous on* $\mathbb{R}$ *and defining*

$$\hat{g}_\nu^{(\nu)}(x) = \frac{1}{b^{\nu+1}} \sum_{i=1}^{n} \int_{d_{i-1}}^{d_i} K^{(\nu)}\left(\frac{x-u}{b}\right) du \, y_i,$$

*we obtain*

$$\hat{g}_\nu^{(\nu)}(\hat{\theta}(s), t) \to_p g^{(\nu)}(\theta) \quad \textit{uniformly in } s \in [s_a, s_b],\ t \in [t_a, t_b]. \tag{3.7}$$

The main application of the theorem is

COROLLARY 2. *Under Assumption* A, *assuming K is symmetric, for any consistent estimates* $\hat{s} \to_p s^*$ (2.11) *and* $\hat{t} \to_p t^*$ (2.12), *the asymptotic distribution of* $\begin{bmatrix} X_n(\hat{s}) \\ Y_n(\hat{s}, \hat{t}) \end{bmatrix}$ *is the same as that of* $\begin{bmatrix} X_n(s^*) \\ Y_n(s^*, t^*) \end{bmatrix}$, *i.e., the corresponding peak estimates are asymptotically efficient. The asymptotic distribution is given by* (3.3), *replacing s and t by* $s^*$ *and* $t^*$.

PROOF. It is sufficient to show that $X_n(\hat{s}) - X_n(s^*) \to_p 0$ and $Y_n(\hat{s}, \hat{t}) - Y_n(s^*, t^*) \to_p 0$. By (3.2), for any given $\alpha, \eta > 0$ there is a $\delta > 0$ s.t. for sufficiently large $n$,

$$P\big(|X_n(\hat{s}) - X_n(s^*)| > \alpha\big) \le P\big(|\hat{s} - s^*| < \delta, |X_n(\hat{s}) - X_n(s^*)| > \alpha\big) + P(|\hat{s} - s^*| > \delta) < 2\eta,$$

and an analogous result follows for $Y_n(\hat{s}, \hat{t})$ from (3.1). □

The estimate $\hat{\sigma}_R$ and the kernel estimates of the derivatives of $g$ at $\theta$ given in Lemma 1 can be employed for efficient peak estimation. A specific procedure and its finite sample properties are discussed in Section 5. It is interesting to note that location and size of a peak are asymptotically uncorrelated. The latter is due to the symmetry of the kernel $K$ which is used for the estimation of the size and is orthogonal in the $L^2$ sense to any scaled version of $K^{(1)}$, the kernel which is implicitly used for the location.

**4. Auxiliary results and proofs.** This section contains a sequence of auxiliary results which lead to the proof of the theorem and of Lemma 1. The starting point is an investigation of local bias and variance and of asymptotic normality of estimates $\hat{g}_j^{(j)}(x, s)$ (2.2).

LEMMA 2. *Let* $k \ge 1$, $g \in \mathscr{C}^{k+j}$ *and* $K \in \mathscr{M}_k \cap \mathscr{C}^j$ *for given j and k. Then*

$$\operatorname{var}\big(\hat{g}_j^{(j)}(x, s)\big) = n^{-2k/(2(k+j)+1)} s^{-(2j+1)} \sigma^2 V^{(j)}(1 + o(1)), \tag{4.1}$$

*where* $V^{(j)} = \int K^{(j)}(v)^2 \, dv$ *and*

$$E\hat{g}_j^{(j)}(x, s) - g^{(j)}(x) = n^{-k/(2(k+j)+1)} s^k g^{(k+j)}(x) B_k(1 + o(1)), \tag{4.2}$$

*where* $B_k$ *is defined in* (2.6).

PROOF. Similar considerations as in Gasser and Müller (1984) yield

$$E\hat{g}_j^{(j)}(x, s) - g^{(j)}(x) = \frac{1}{b^{j+1}} \int K^{(j)}\left(\frac{x - u}{b}\right) g(u) \, du + O\left(\frac{1}{nb^j}\right) - g^{(j)}(x), \tag{4.3}$$

and (4.2) follows by a Taylor expansion; compare (2.2). (4.1) follows by a direct integral approximation of the variance. □

LEMMA 3. *Under the assumptions of Lemma* 2,

$$n^{k/(2(k+j)+1)}\left(\hat{g}_j^{(j)}(x,s)-g^{(j)}(x)\right) \rightarrow_{\mathscr{D}} \mathscr{N}\left(s^k g^{(k+j)}(x)B_k,\, s^{-(2j+1)}\sigma^2 V^{(j)}\right). \tag{4.4}$$

PROOF. $(\hat{g}_j^{(j)}(x,s)-E\hat{g}_j^{(j)}(x,s))$ is a weighted average of a triangular array of independent random variables with weights

$$w_i(x)=\frac{1}{b^{j+1}}\int_{d_{i-1}}^{d_i} K^{(j)}\left(\frac{x-u}{b}\right)du, \tag{4.5}$$

where $b=sn^{-1/(2(k+j)+1)}$. By the assumptions, observing (4.1) and (4.2), Lindeberg's condition for the central limit theorem holds. □

Applying Lemmas 2 and 3 we obtain a result on the weak convergence of $\hat{g}_j^{(j)}(x,s)$ in $s$ for $x$ fixed.

LEMMA 4. *If $g\in\mathscr{C}^{k+j}(I)$ and $K\in\mathscr{M}_k\cap\mathscr{C}^{j+1}$, $k\ge 1$, we have for any $x\in I$, defining $H_n(s)=n^{k/(2(k+j)+1)}(\hat{g}_j^{(j)}(x,s)-g^{(j)}(x))$ that*

$$H_n(s)\Rightarrow H(s)\quad \text{on } \mathscr{C}([s_a,s_b]), \tag{4.6}$$

*where $H(s)$ is a Gaussian process defined by*

$$EH(s)=s^k g^{(k+j)}(x)B_k \tag{4.7}$$

*and*

$$\operatorname{cov}(H(s_1),H(s_2))=\frac{\sigma^2}{(s_1s_2)^{2j+1}}\int K^{(j)}\left(\frac{u}{s_1}\right)K^{(j)}\left(\frac{u}{s_2}\right)du \tag{4.8}$$

*for $s_1,s_2\in[s_a,s_b]$.*

PROOF. Writing

$$\begin{aligned}H_n(s)&=n^{k/(2(k+j)+1)}\left(\left[\hat{g}_j^{(j)}(x,s)-E\hat{g}_j^{(j)}(x,s)\right]+\left[E\hat{g}_j^{(j)}(x,s)-g^{(j)}(x)\right]\right)\\&=H_{n,1}(s)+H_{n,2}(s),\quad\text{say,}\end{aligned}$$

we show by direct calculation of variances and covariances similarly to (4.1), applying Lemma 3 and the Cramér–Wold device, that

$$(H_{n,1}(s_1),\dots,H_{n,1}(s_m))\rightarrow_{\mathscr{D}}\mathscr{N}(0,C)$$

for any $s_1,\dots,s_m\in[s_a,s_b]$, where the elements of the covariance matrix $C$ are as given by (4.8). Lemma 2 (4.2) implies that $EH_{n,2}(s)\to s^k g^{(k+j)}(x)B_k$ and therefore the proof is completed according to Theorem 8.1 of Billingsley (1968) if we show the tightness of $H_{n,1}(\cdot)$. The tightness of $H_{n,1}(\cdot)$ is shown by means of Theorem 12.3 of Billingsley. As in the proof of Lemma 3.1 of Müller and Stadtmüller (1987a) one shows by means of the mean value theorem that $E|H_{n,1}(s_1)-H_{n,1}(s_2)|^2\le c|s_1-s_2|^2$ for any $s_1,s_2\in[s_a,s_b]$ and some constant $c>0$, using the "intermediate kernel" $K_j(x)=(j+1)K^{(j)}(x)+xK^{(j+1)}(x)$. □

Next we discuss uniform convergence of $g_l^{(j)}(x, s)$ which will be employed to derive the asymptotic distribution of $\hat{\theta}(s)$. The deterministic and the stochastic parts of the maximal deviation are analysed separately.

LEMMA 5. *Let $K \in \mathscr{M}_k \cap \mathscr{C}^j$ for some $k \geq 1$, $0 \leq j \leq k$, and $g \in \mathscr{C}^j(I)$. For any $l \geq 0$ we have*

$$\sup_{x \in I} \left| E\hat{g}_l^{(j)}(x, s) - g^{(j)}(x) \right| = O\left(sn^{-1/(2(k+l)+1)} + s^{-j}n^{-1+j/(2(k+l)+1)}\right). \tag{4.9}$$

PROOF. Observe that (4.3) implies

$$\left| E\hat{g}_l^{(j)}(x, s) - g^{(j)}(x) \right| \leq \int |K(u)| \left| g^{(j)}(x - ub) - g^{(j)}(x) \right| du + O\left((nb^j)^{-1}\right)$$

uniformly in $x$. □

The following result is adapted from Müller and Stadtmüller [(1987b), Lemma 5.2] and the following remarks.

LEMMA 6. *Let $\hat{m}(\cdot)$ be a moving weighted average $\hat{m}(x) = \sum_{i=1}^n w_i(x)\varepsilon_i$ in the model* (1.1), *i.e., assume $(\varepsilon_i) = (\epsilon_{i,n})_{1 \leq i \leq n}$ form a triangular array, and let $E|\varepsilon_1|^r < \infty$ for some $r > 2$. Assume for some $\delta > 0$ and $\eta \in (0, r-2)$, that for constants $c$, $L$, $M > 0$ and a sequence $(\alpha_n)$, the following conditions are satisfied*:

(i) $\sup_{x_1, x_2 \in I} |w_i(x_1) - w_i(x_2)| \leq L|x_1 - x_2|^\delta$ *for* $x_1, x_2 \in I$.
(ii) $\max_{1 \leq i \leq n} |w_i(x)| \geq cn^{-1}$ *uniformly for* $x \in I$.
(iii) $n^{1/(r-\eta)} \max_{1 \leq i \leq n} |w_i(x)| \log n \leq \alpha_n / M$ *uniformly in* $x$.
(iv) $(\sum_{i=1}^n w_i(x)^2 \log n)^{1/2} \leq \alpha_n / M$ *uniformly in* $x$.

*Then*

$$\sup_{x \in I} |\hat{m}(x)| = O_p(\alpha_n). \tag{4.10}$$

Combining Lemmas 5 and 6, choosing $\alpha_n = O(((\log n)/nb^{2j+1})^{1/2})$, where the weights $w_i(x)$ appearing in $\hat{m}(x)$ are given by (4.5), one obtains the following.

LEMMA 7. *Under Assumption* A, *we have for* $j = 0, 1, 2$ *and for* $l \geq 0$,

$$\begin{aligned}\sup_{x \in I} \left| \hat{g}_l^{(j)}(x, s) - g^{(j)}(x) \right| &= O_p\Big(sn^{-1/[2(k+l)+1]} + n^{-1+j/[2(k+l)+1]}s^{-j} \\ &\qquad + \left((\log n)n^{-1+(2j+1)/[2(k+l)+1]}s^{-(2j+1)}\right)^{1/2}\Big) \\ &= o_p(1) \quad \text{uniformly in } s \in [s_a, s_b].\end{aligned} \tag{4.11}$$

Now we are ready to derive the asymptotic distributions of $\hat{\theta}(s)$ and of $\hat{g}_0(\hat{\theta}(s), t)$.

LEMMA 8. *Under Assumption* A, *it holds for any* $s \in [s_a, s_b]$ *that*

$$n^{k/(2k+3)}\big(\hat{\theta}(s) - \theta\big) \to_{\mathscr{D}} \mathscr{N}\left(\frac{-s^k g^{(k+1)}(\theta) B_k}{g^{(2)}(\theta)}, \frac{\sigma^2 V^{(1)}}{s^3 g^{(2)}(\theta)^2}\right),$$

*where* $V^{(1)}$ *and* $B_k$ *are defined in* (2.6).

PROOF. Lemma 7 for $l = 1$, $j = 0, 1, 2$ implies that

$$\sup_{x \in I} \big|\hat{g}_1^{(j)}(x, s) - g^{(j)}(x)\big| = o_p(1) \quad \text{uniformly in } s \in [s_a, s_b]. \tag{4.12}$$

Since $g^{(2)}(\theta) < 0$, $g^{(2)}(\cdot)$ is continuous and $g(\theta)$ is a unique maximum of $g(\cdot)$, it follows from (4.12) for $j = 0$ that

$$\sup_{s \in [s_a, s_b]} \big|\hat{\theta}(s) - \theta\big| = o_p(1). \tag{4.13}$$

By a Taylor expansion [compare Müller (1985)] we find [observing $\hat{g}_1^{(1)}(\hat{\theta}) = g^{(1)}(\theta) = 0$],

$$\hat{\theta}(s) - \theta = \big(g^{(1)}(\theta) - \hat{g}_1^{(1)}(\theta, s)\big)/\big(g^{(2)}(\theta)\big) + R_n(s), \tag{4.14}$$

where

$$R_n(s) = \frac{\big(g^{(1)}(\theta) - \hat{g}_1^{(1)}(\theta, s)\big)\big(g^{(2)}(\theta) - \hat{g}_1^{(2)}(\theta^*(s), s)\big)}{g^{(2)}(\theta)\hat{g}_1^{(2)}(\theta^*(s), s)} \tag{4.15}$$

and $\theta^*(s)$ is an intermediate value between $\theta$ and $\hat{\theta}(s)$. The result follows from Lemma 3 (choosing $j = 1$) via

$$\sup_{s \in [s_a, s_b]} n^{k/(2k+3)} |R_n(s)| = o_p(1) \tag{4.16}$$

and

$$\sup_{s \in [s_a, s_b]} \big|\big(g^{(2)}(\theta) - \hat{g}_1^{(2)}(\theta^*(s), s)\big)/\hat{g}_1^{(2)}(\theta^*(s), s)\big| = o_p(1),$$

which is a consequence of (4.12) for $j = 2$ and (4.13). □

LEMMA 9. *Under Assumption* A, *we have for any* $s \in [s_a, s_b]$ *and* $t \in [t_a, t_b]$ *that*

$$n^{k/(2k+1)}\big(\hat{g}_0(\hat{\theta}(s), t) - g(\theta)\big) \to_{\mathscr{D}} \mathscr{N}\left(t^k g^{(k)}(\theta) B_k, \frac{\sigma^2 V}{t}\right).$$

PROOF. By a Taylor expansion,

$$\hat{g}_0(\theta, t) - \hat{g}_0(\hat{\theta}(s), t) = \tfrac{1}{2}\hat{g}_0^{(2)}(\tilde{\theta}(s), t)\big(\tilde{\theta}(s) - \theta\big)^2 \quad \text{for some mean value } \tilde{\theta}(s).$$

We conclude by (4.11) for $l = 0$, $j = 2$ and by (4.13) and the continuity of $g^{(2)}(\cdot)$ that

$$|\hat{g}_0^{(2)}(\tilde{\theta}(s), t) - g^{(2)}(\theta)| = o_p(1) \quad \text{uniformly in } s \text{ and } t. \tag{4.17}$$

By Lemma 8, it follows that $n^{2k/(2k+3)}(\hat{g}_0(\theta, t) - \hat{g}_0(\hat{\theta}(s), t))$ has a nondegenerate limit distribution and therefore

$$n^{k/(2k+1)}|\hat{g}_0(\theta, t) - \hat{g}_0(\hat{\theta}(s), t)| = o_p(1), \tag{4.18}$$

so that Lemma 3 ($j = 0$) implies the result. □

In order to investigate the tightness of the process $\begin{bmatrix} X_n(\cdot) \\ Y_n(\cdot, \cdot) \end{bmatrix}$ [see (2.13)], we observe that owing to the continuity of $\hat{g}_l(\cdot, \cdot)$, $l = 0, 1$, on $I \times [t_a, t_b]$, the function $\hat{\theta}(\cdot)$ can be normalized so that $X_n(\cdot) \in D([s_a, s_b])$, $Y_n(\cdot, t) \in D([s_a, s_b])$ and $Y_n(s, \cdot) \in \mathscr{C}([t_a, t_b])$. For a sequence of processes $Z_n(\cdot, \cdot) \in D_2([s_a, s_b] \times [t_a, t_b])$ we apply the following tightness condition, which is a consequence of Billingsley (1968), Theorem 15.5 and of Bickel and Wichura (1971), Theorem 2.

TIGHTNESS CONDITION. The sequence $Z_n(\cdot, \cdot) \in D_2([s_a, s_b] \times [t_a, t_b])$ is tight if:

(i) $Z_n(s_0, t_0)$ is tight for some fixed $s_0 \in [s_a, s_b]$ and $t_0 \in [t_a, t_b]$.
(ii) For any given $\alpha, \eta > 0$ there exist a $\delta > 0$ and an integer $n_0$ s.t. for all $n > n_0$

$$P\left( \sup_{\substack{|s_1 - s_2| < \delta,\ s_1, s_2 \in [s_a, s_b] \\ |t_1 - t_2| < \delta,\ t_1, t_2 \in [t_a, t_b]}} |Z_n(s_1, t_1) - Z_n(s_2, t_2)| > \alpha \right) < \eta. \tag{4.19}$$

The limiting process is then continuous with probability 1.

We apply this criterion to the processes

$$Z_{n, \lambda, \mu}(s, t) = \lambda X_n(s) + \mu Y_n(s, t) \quad \text{for any } \lambda, \mu.$$

LEMMA 10. *Under Assumption* A, $Z_{n, \lambda, \mu}(\cdot, \cdot)$ *is tight.*

PROOF. By (4.14),

$$|X_n(s_1) - X_n(s_2)| \le \frac{n^{k/(2k+3)}}{g^{(2)}(\theta)} |\hat{g}_1^{(1)}(\theta, s_1) - \hat{g}_1^{(1)}(\theta, s_2)|$$
$$+ n^{k/(2k+3)} |R_n(s_1) - R_n(s_2)|.$$

Since the processes $H_n(\cdot)$ of Lemma 4 are in $\mathscr{C}([s_a, s_b])$, we can apply Theorem 8.2 of Billingsley (1968), Lemma 4 (choosing $x = \theta$) and (4.16) to obtain the existence of a $\delta' > 0$ s.t. for large $n$, given $\alpha$ and $\eta$,

$$P\left(\sup_{\substack{|s_1 - s_2| < \delta' \\ s_1, s_2 \in [s_a, s_b]}} \lambda |X_n(s_1) - X_n(s_2)| > \frac{\alpha}{2}\right) < \frac{\eta}{2}. \tag{4.20}$$

Therefore the processes $X_n(\cdot)$ satisfy (ii) of the tightness condition, and it follows by Lemma 8 that $X_n(\cdot)$ is tight. Since by Lemma 8 the finite-dimensional distributions of $X_n(\cdot)$ converge weakly, it follows that $X_n(\cdot)$ converges weakly and we conclude by the continuous mapping theorem (see Billingsley) that $\sup_{s_a \le s \le s_b} n^{2k/(2k+3)}(\hat{\theta}(s) - \theta)^2$ has a limiting distribution. Therefore,

$$\sup_{s_a \le s \le s_b} n^{k/(2k+1)}\left(\hat{\theta}(s) - \theta\right)^2 = o_p(1), \tag{4.21}$$

$$\begin{aligned} |Y_n(s_1, t_1) - Y_n(s_2, t_2)| &\le n^{k/(2k+1)}\Big(\left|\hat{g}_0\big(\hat{\theta}(s_1), t_1\big) - \hat{g}_0\big(\hat{\theta}(s_2), t_1\big)\right| \\ &\qquad + \left|\hat{g}_0\big(\hat{\theta}(s_2), t_1\big) - \hat{g}_0\big(\hat{\theta}(s_2), t_2\big)\right|\Big) \\ &= \mathrm{I} + \mathrm{II}, \quad \text{say}. \end{aligned}$$

$$\begin{aligned} \mathrm{I} &\le 2 \sup_{\substack{s_a \le s \le s_b \\ t_a \le t \le t_b}} n^{k/(2k+1)} \left|\hat{g}_0\big(\hat{\theta}(s), t\big) - \hat{g}_0(\theta, t)\right| \\ &\le \sup_{\substack{s_a \le s \le s_b \\ t_a \le t \le t_b}} \left|\hat{g}_0^{(2)}\big(\tilde{\theta}(s), t\big)\right| \sup_{s_a \le s \le s_b} \left(\tilde{\theta}(s) - \theta\right)^2 n^{k/(2k+1)} = o_p(1) \end{aligned}$$

for some mean value $\tilde{\theta}(s)$ between $\hat{\theta}(s)$ and $\theta$ by (4.21), (4.13) and Lemma 7 (choosing $l = 0$ and $j = 2$).

$$\mathrm{II} \le 2 \sup_{\substack{s_a \le s \le s_b \\ t_a \le t \le t_b}} n^{k/(2k+1)} \left|\hat{g}_0\big(\hat{\theta}(s), t\big) - \hat{g}_0(\theta, t)\right| + n^{k/(2k+1)} \left|\hat{g}_0(\theta, t_1) - \hat{g}_0(\theta, t_2)\right|.$$

The first term is treated as I above and seen to be $o_p(1)$. For the second term we use Lemma 4 and Theorem 8.2 of Billingsley (1968) to conclude that for given $\alpha, \eta$ there exists $\delta'' > 0$ s.t. for large $n$

$$P\left(\sup_{\substack{|t_1 - t_2| < \delta'',\ t_1, t_2 \in [t_a, t_b] \\ |s_1 - s_2| < \delta'',\ s_1, s_2 \in [s_a, s_b]}} \mu |Y_n(s_1, t_1) - Y_n(s_2, t_2)| > \frac{\alpha}{2}\right) < \frac{\eta}{2}. \tag{4.22}$$

(4.20) and (4.22) yield (ii) of the tightness condition; (i) follows from Lemmas 8 and 9. □

PROOF OF THE THEOREM. The weak convergence follows from Lemmas 8, 9 and 10. Lemma 10, (4.20) and (4.22), implies that the limiting process is continuous with probability 1. Lemmas 8 and 9 show that the limiting process is Gaussian and give the expectations. Applying the Cramér–Wold device and calculating the covariances directly [compare the approach in Müller (1985)], observing that $K^{(1)}$ is antisymmetric and applying (4.14), (4.16) and (4.18), yields the limiting process of the theorem. □

PROOF OF LEMMA 1. Apply (4.13) and an extension of Lemma 7 to the case $j = \nu$, $l = \nu$. □

**5. A practical procedure for adaptive peak estimation.** We assume here that Assumption A of Section 3 is satisfied for $k = 2$. As discussed in Section 3, the basic idea is to substitute consistent estimates of $\sigma$, $g^{(2)}(\theta)$ and $g^{(3)}(\theta)$ into the minimizers $s^*$ (2.11) and $t^*$ (2.9) of the MSE. Such an approach is asymptotically justified by (3.6) and Lemma 1. The aim of this section is to propose a practical algorithm for adaptive peak estimation and to compare its finite sample properties with a more standard procedure by simulation. We will use polynomial kernel functions $K_{k\mu} \in \mathscr{M}_k \cap \mathscr{C}^{\mu}$ [compiled, e.g., in Müller (1984)],

$$K_{2\mu}(x) \equiv \frac{3 \cdot 5 \cdot \cdots \cdot (2\mu + 1)}{2^{2\mu - [\mu/2]}}(1 - x^2)^{\mu}, \quad \mu = 1, \ldots, 4,$$

$$K_{41}(x) = \tfrac{15}{32}(3 - 10x^2 + 7x^4),$$

plus the asymmetric kernels

$$K_{31}(x) \equiv \tfrac{15}{32}(3 - 3x - 10x^2 + 10x^3 + 7x^4 - 7x^5)$$

and

$$K_{51}(x) \equiv 0.1367(15 - 105x^2 + 189x^4 - 99x^6) - 0.2099(-5x + 35x^3 - 63x^5 + 33x^7).$$

First we discuss an algorithm for the adaptive estimation of the location of the peak. From Lemma 2 we see that the optimal local bandwidth (w.r. to MSE) for the kernel estimate $\hat{g}_1^{(1)}(x, \cdot)$ of the first derivative at a fixed point $x$ is

$$b_l^*(x) = s^*(x) n^{-1/7}, \tag{5.1}$$

where

$$s^*(x) = \left(3\sigma^2 V^{(1)} / 4 g^{(3)}(x)^2 B_2^2\right)^{1/7}, \tag{5.2}$$

provided that $g^{(3)}(x) \neq 0$, so that by (2.11),

$$s^* = s^*(\theta). \tag{5.3}$$

The location of an extremum is the same as the location of a zero of the derivative, i.e., it holds that

$$\hat{\theta}(s) = \hat{\theta}_1(s), \tag{5.4}$$

where

$$\hat{\theta}_1(s) = \inf\{x \in I: \hat{g}_1^{(1)}(x, s) = 0\}. \tag{5.5}$$

Defining $\hat{\sigma}_R$ as in (3.5) and

$$\hat{s}(x) = \left(\left(3\hat{\sigma}_R^2\left(V^{(1)}/4\hat{g}_3^{(3)}\right)(x)^2 B_2^2\right)^{1/7} \wedge s_b\right) \vee s_a, \tag{5.6}$$

it follows (using Lemma 2) that

$$\hat{s}(x) \to_p s^*(x), \qquad \hat{s}(x) \in [s_a, s_b], \qquad \hat{s}(\cdot) \in \mathscr{C}([a, b])$$

for all $x \in I$ and all $n$. Defining

$$\begin{aligned} \hat{\theta}_2(s) &= \sup\{x \in I: \hat{g}_1^{(1)}(x, s) = 0\}, \\ \hat{\theta}_3 &= \inf\{x \in I: \hat{g}_1^{(1)}(x, \hat{s}(x)) = 0\}, \end{aligned} \tag{5.7}$$

it follows from $\hat{g}_1^{(1)}(\hat{\theta}_3, \hat{s}(\hat{\theta}_3)) = 0$ that $\hat{\theta}_3 \in [\hat{\theta}_1(\hat{s}(\hat{\theta}_3)), \hat{\theta}_2(\hat{s}(\hat{\theta}_3))]$ and since $\hat{\theta}_1(s)$ and $\hat{\theta}_2(s)$ behave identically w.r. to weak convergence and convergence in probability, $\hat{\theta}_3$ and $\hat{\theta}(\hat{s}(\hat{\theta}_3))$ behave identically according to (5.4). It follows from (3.6) and the uniformity of the convergence (3.7) that $\hat{s}(\hat{\theta}_3) \to_p s^*(\theta) = s^*$ and, therefore, $\hat{\theta}(\hat{s}(\hat{\theta}_3))$ and $\hat{\theta}_3$ are asymptotically efficient by Corollary 2.

This means that it is an efficient procedure to locate a zero in the kernel estimate of the first derivative of the regression curve using consistently estimated optimal local bandwidths. For the size of the peak, it follows that

$$\hat{t} = \left(\hat{\sigma}_R^2 V / 4\hat{g}_2^{(2)}(\hat{\theta}_3)^2 B_2^2\right)^{1/5} \tag{5.8}$$

satisfies $\hat{t} \to_p t^*$. Therefore, $(\hat{\theta}_3, \hat{g}_0(\hat{\theta}_3, \hat{t}))$ are efficient estimates of the peak coordinates.

Following these considerations, the practical procedure to be presented now consists of four steps:

1. Estimation of local optimal bandwidths for a kernel estimate $\hat{g}_1^{(1)}$ of $g^{(1)}$.
2. Location of a zero $\hat{\theta}_3$ of $\hat{g}_1^{(1)}$ using the local bandwidths of step 1.
3. Estimation of the local optimal bandwidth of the kernel estimate $\hat{g}_0$ of $g$ at $\hat{\theta}_3$.
4. Computation of $\hat{g}_0(\hat{\theta}_3)$ using the bandwidth of step 3.

For local bandwidth choice steps 1 and 3 a modified version of a procedure by Müller and Stadtmüller (1987a) is used. The asymptotically optimal global bandwidths $b_{j,k}^*$ w.r. to integrated MSE for estimating $g^{(j)}$, $j \geq 0$, using a kernel $K^{(j)}$ where $K \in \mathscr{M}_k \cap \mathscr{C}^j$, satisfy the asymptotic relations

$$b_{j,k}^* / b_{0,k+j}^* = d_{j,k} \tag{5.9}$$

with known constants $d_{j,k}$ depending only on the kernel functions used, and the relation between optimal local and global bandwidth for $j = 1$ and $k = 2$ is given by

$$b_l^*(x) = b_{1,2}^*\left(\int_0^1 g^{(3)}(u)^2\, du / g^{(3)}(x)^2\right)^{1/7} \tag{5.10}$$

[compare Müller and Stadtmüller (1987a)]. Therefore, we use the estimated local bandwidths

$$\hat{b}_l(x) = d_{1,2}\hat{b}_{0,3}\left[\left(\left(\frac{(1/N)\Sigma_{i=1}^N \hat{g}^{(3)}(i/N)^2}{\hat{g}^{(3)}(x)}\right)^{1/7} \wedge 1.4\right) \vee 0.6\right] \tag{5.11}$$

for the kernel $K_{22}^{(1)}$, where we choose $N = 200$ and $\hat{b}_{0,3}$ is the global bandwidth determined by the Rice criterion with the kernel $K_{31}$, i.e., the minimizer of

$$R(b) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{g}(x_i))^2 + \frac{2\hat{\sigma}_R^2 K_{31}(0)}{nb}, \tag{5.12}$$

w.r. to $b$, where $\hat{g}(\cdot)$ is defined in (1.2). The cutoff points 0.6 and 1.4 are chosen arbitrarily in order to avoid too extreme local bandwidths in accordance with experiences in Müller and Stadtmüller (1987a). The lower cutpoint is only very rarely attained in procedure (5.11), but making it too small would lead to outlying bandwidths and estimates. The upper cutpoint seems to be less important. For the given kernels we find $d_{1,2} = 0.7083$. For the estimation of $\hat{g}^{(3)}(\cdot)$ in (5.11) we use the kernel $K_{24}^{(3)}$ and the global bandwidth $\hat{b}_{3,2} = d_{2,3}\hat{b}_{0,5}$, where $\hat{b}_{0,5}$ is the minimizer of (5.12) using the kernel $K_{51}$ and $d_{2,3} = 0.6788$.

TABLE 1

*Performance of a practical procedure for adaptive peak estimation. Number of Monte Carlo runs* 200. *Number of observations* 50, *equidistantly in* [0, 1]. *Curve used is* $g \equiv 1 + 3\exp((x - 0.5)^2/0.01)$. *Coordinates of the peak are* (0.5, 4.0). *Small numbers denote powers of* 10 *by which to multiply.*

| | Variance | | |
|---|---|---|---|
| | $\sigma^2 = 0.25$ | $\sigma^2 = 0.5$ | $\sigma^2 = 1.0$ |
| Average bandwidth for location | $1.152^{-1}$ | $1.280^{-1}$ | $1.403^{-1}$ |
| Average bandwidth for size | $4.792^{-2}$ | $5.710^{-2}$ | $7.122^{-2}$ |
| Average estimated location | 0.5005 | 0.5012 | 0.5012 |
| Average squared error for location | $1.247^{-4}$ | $1.893^{-4}$ | $2.860^{-4}$ |
| Average estimated size | 3.909 | 3.887 | 3.861 |
| Average squared error for size | $8.341^{-2}$ | $1.550^{-1}$ | $3.332^{-1}$ |

TABLE 2

*Performance of a nonadaptive procedure for peak estimation choosing one global bandwidth by the Rice criterion. Minimization of* (5.12) *on the interval* [0.25, 0.75]. *Same design as in Table* 1 (*same random numbers*); *kernel* $K_{21}$.

| | Variance | | |
|---|---|---|---|
| | $\sigma^2 = 0.25$ | $\sigma^2 = 0.5$ | $\sigma^2 = 1.0$ |
| Average bandwidth chosen | $6.256^{-2}$ | $7.346^{-2}$ | $8.974^{-2}$ |
| Average estimated location | 0.5010 | 0.5010 | 0.5007 |
| Average squared error for location | $1.811^{-4}$ | $2.480^{-4}$ | $3.369^{-4}$ |
| Average estimated size | 3.813 | 3.769 | 3.705 |
| Average squared error for size | $1.006^{-1}$ | $1.910^{-1}$ | $3.758^{-1}$ |

The algorithm step 3 is completely analogous. Employing for the kernel $K_{21}$ the bandwidth

$$(5.13)\qquad \hat{b}_l(\hat{\theta}_3) = d_{2,2}\hat{b}_{0,4}\left[\left(\left(\frac{(1/N)\Sigma_{i=1}^{N}\hat{g}^{(2)}(i/N)^2}{\hat{g}^{(2)}(\hat{\theta}_3)}\right)^{1/9} \wedge 1.4\right) \vee 0.6\right],$$

where $\hat{\theta}_3$ is the estimated zero resulting from step 2, $\hat{b}_{0,4}$ is the minimizer of (5.12) using the kernel $K_{41}$ and $d_{2,2} = 1.028$. We obtain the adaptive estimate $\hat{g}(\hat{\theta}_3)$ of the size of the peak.

The algorithm was tested in a Monte Carlo study where the function $g(x) = 1 + 3\exp(-(x - 0.5)^2/0.01)$ [symmetric peak at (0.5, 4.0)] and the residual variances $\sigma^2 = 0.25, 0.5, 1.0$ were used. The results are given in Table 1.

The local bandwidths chosen become larger with increasing $\sigma^2$ and the squared errors for estimated location/size get worse with increasing $\sigma^2$ as is to be expected. The corresponding values for a locally nonadaptive procedure choosing one global bandwidth by the Rice criterion (5.12) and reading location and size of the peak off the estimated curve are given in Table 2.

A comparison of the finite sample behavior of the adaptive method with this nonadaptive method shows a clear advantage of local adaptation which usually leads to smaller squared errors.

**Acknowledgment.** I thank Thomas Schmitt for writing computer programs and carrying out the simulations.

## REFERENCES

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

BHATTACHARYA, P. K. and MACK, Y. P. (1987). Weak convergence of *k-NN* density and regression estimators with varying *k* and applications. *Ann. Statist.* **15** 976–994.

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670.

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

BURMAN, P. (1985). A data dependent approach to density estimation. *Z. Wahrsch. verw. Gebiete* **69** 609–628.

EDDY, W. F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.* **8** 870–882.

GASSER, TH. and MÜLLER, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.* **11** 171–184.

HALL, P. (1983). Large sample optimality of least squares cross-validation in density estimation. *Ann. Statist.* **11** 1156–1174.

HÄRDLE, W. and MARRON, J. S. (1985). Optimal bandwidth selection in nonparametric regression. *Ann. Statist.* **13** 1465–1481.

JORGENSEN, M., NIELSEN, C. T., KEIDING, N. and SKAKKEBACK, N. E. (1985). Parametrische und nichtparametrische Modelle für Wachstumsdaten. In *Med. Informatik und Statistik* (G. C. Pflug, ed.) **60** 74–87. Springer, Berlin.

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

MARRON, J. S. (1985). An asymptotically efficient solution to the bandwidth problem of kernel density estimation. *Ann. Statist.* **13** 1011–1023.

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

MÜLLER, H.-G. (1985). Kernel estimators of zeros and of location and size of extrema of regression functions. *Scand. J. Statist.* **12** 221–232.

MÜLLER, H.-G. and STADTMÜLLER, U. (1987a). Variable bandwidth kernel estimators of regression curves. *Ann. Statist.* **15** 182–201.

MÜLLER, H.-G. and STADTMÜLLER, U. (1987b). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15** 610–625.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

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

RICE, J. (1984b). Boundary modification for kernel regression. *Comm. Statist. A—Theory Methods* **13** 893–900.

RICE, J. (1985). Bandwidth choice for differentiation. *J. Multivariate Anal.* **19** 251–264.

SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *J. Roy. Statist. Soc. Ser. B* **47** 1–48.

SPECKMAN, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13** 970–983.

STONE, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.* **12** 1285–1297.

WAHBA, G. and WOLD, S. (1975). A completely automatic french curve: Fitting spline functions by cross-validation. *Comm. Statist. A—Theory Methods* **4** 1–17.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616