The Annals of Statistics
1992, Vol. 20, No. 4, 2008–2036

VARIABLE BANDWIDTH AND LOCAL LINEAR REGRESSION SMOOTHERS

BY JIANQING FAN[1] AND IRÈNE GIJBELS[2]

University of North Carolina and Limburgs Universitair Centrum

In this paper we introduce an appealing nonparametric method for estimating the mean regression function. The proposed method combines the ideas of local linear smoothers and variable bandwidth. Hence, it also inherits the advantages of both approaches. We give expressions for the conditional MSE and MISE of the estimator. Minimization of the MISE leads to an explicit formula for an optimal choice of the variable bandwidth. Moreover, the merits of considering a variable bandwidth are discussed. In addition, we show that the estimator does not have boundary effects, and hence does not require modifications at the boundary. The performance of a corresponding plug-in estimator is investigated. Simulations illustrate the proposed estimation method.

1. Introduction. In the case of bivariate observations, it is of common interest to explore the association between the covariate and the response. One possible way to describe such an association is via the mean regression function. A flexible estimation method does not make any assumption on the form of this function. This form should be determined completely by the data. In other words, a nonparametric approach is preferable.

In this paper, we will concentrate on nonparametric kernel-type estimation, a popular approach in curve estimation. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a random sample from a population (X, Y) and denote by

$$m(x) = E(Y|X = x)$$

the mean regression function of Y given X. Further, we use the notations $f_X(\cdot)$ and $\sigma^2(\cdot)$ for the marginal density of X and the conditional variance of Y given X, respectively.

Most regression estimators studied in the literature are of the form

$$\sum_{j=1}^{n} w_j(x; X_1, \dots, X_n) Y_j.$$

Such a kind of estimator is called a linear smoother, since it is linear in the response. In this paper we consider a linear smoother which is obtained via a local linear approximation to the mean regression function. More precisely, the

Received April 1991; revised January 1992.

[1]Supported in part by NSF Grant DMS-90-05905.

[2]Completed while visiting Department of Statistics, University of North Carolina, Chapel Hill. Senior Research Assistant at the National Fund for Scientific Research (Belgium).

AMS 1980 *subject classifications*. Primary 62G07; secondary 62G20, 62J99.

Key words and phrases. Boundary effects, local linear smoother, mean squared error, nonparametric regression, optimalities, variable bandwidth.

estimator is defined as $\hat{m}(x) = \hat{a}$, where $\hat{a}$ together with $\hat{b}$ minimizes

$$\sum_{j=1}^{n} \left(Y_j - a - b(x - X_j)\right)^2 K\left(\frac{x - X_j}{h_n}\right), \tag{1.1}$$

with $K(\cdot)$ a bounded (kernel) function and h_n a sequence of positive numbers tending to zero, called the smoothing parameter or bandwidth. It turns out that $\hat{m}(x)$ is the best linear smoother, in the sense that it is the asymptotic minimax linear smoother when the unknown regression function is in the class of functions having bounded second derivative. This property is established in Fan (1992a). The preceding idea is an extension of Stone (1977), who uses the kernel $K(x) = 1_{[|x| \le 1]}/2$, resulting in the running line smoother. For a further motivation and study of linear smoothers obtained via a local polynomial approximation to the regression function see Cleveland (1979), Lejeune (1985), Müller (1987), Cleveland and Devlin (1988) and Fan (1992a, b). We will refer to the estimator $\hat{m}(x)$ as a local linear smoother.

The smoothing parameter in (1.1) remains constant, that is, it depends on neither the location of x nor on that of the data X_j. Such an estimator does not fully incorporate the information provided by the density of the data points. Furthermore, a constant bandwidth is not flexible enough for estimating curves with a complicated shape. All these considerations lead to introducing a variable bandwidth $h_n/\alpha(X_j)$, where $\alpha(\cdot)$ is some nonnegative function reflecting the variable amount of smoothing at each data point. This concept of variable bandwidth was introduced by Breiman, Meisel and Purcell (1977) in the density estimation context. Further related studies can be found in Abramson (1982), Hall and Marron (1988), Hall (1990) and Jones (1990).

The estimation method considered in this paper combines the merits of the two preceding procedures. We will study a local linear smoother with variable bandwidth. It is expected that the proposed estimator has all the advantages of both the local linear smoothing method and the variable bandwidth idea. We now give a formal introduction of the estimator. Instead of (1.1), we minimize

$$\sum_{j=1}^{n} \left(Y_j - a - b(x - X_j)\right)^2 \alpha(X_j) K\left(\frac{x - X_j}{h_n}\alpha(X_j)\right) \tag{1.2}$$

with respect to a and b. Denote the solution to this problem by $\hat{a}, \hat{b}$. Then the regression estimator is defined as $\hat{a}$, which is given by

$$\hat{m}(x) = \hat{a} = \sum_{j=1}^{n} w_j Y_j \Big/ \sum_{j=1}^{n} w_j, \tag{1.3}$$

where

$$w_j \equiv \alpha(X_j) K\left(\frac{x - X_j}{h_n}\alpha(X_j)\right)\left[s_{n,2} - (x - X_j)s_{n,1}\right] \tag{1.4}$$

with

$$s_{n,l} = \sum_{j=1}^{n} \alpha(X_j) K\left(\frac{x - X_j}{h_n} \alpha(X_j)\right)(x - X_j)^l, \qquad l = 0, 1, 2. \tag{1.5}$$

It will become clear in the next sections that the estimator (1.3) has several important features. First of all, it shares the nice properties of the local linear smoother: it adapts to both random and fixed designs and to a variety of design densities $f_X(\cdot)$ [see Fan (1992a)]. Furthermore, it does not have the problem of "boundary effects." Also the implementation of a variable bandwidth leads to additional advantages. It gives a certain flexibility in smoothing various types of regression functions. By choosing $\alpha(x) = f_X^{1/4}(x)$ the estimator (1.3) is asymptotically equivalent to a smoothing spline [see Silverman (1984)]. With $\alpha(x) = f_X(x)$ the estimator $\hat{m}(x)$ corresponds approximately to a nearest-neighbor estimator [see Jennen-Steinmetz and Gasser (1988)]. The performance of the estimator can be studied via the mean integrated squared error (MISE). Optimization over all possible variable bandwidths leads to an optimal bandwidth and hence improves this performance. It will be seen that for an optimal choice of $\alpha(\cdot)$ this function is proportional to $f_X^{1/5}(\cdot)$, and this is precisely how an ideal variable kernel smoother should behave [see Silverman (1984)]. With a particular choice of the variable bandwidth, the estimator will have a homogeneous variance (i.e., independent of the location point x), and this is a desirable property. Other advantages of the proposed estimation method will show up in Sections 2–6.

The paper is organized as follows. In the next section, we study in detail the asymptotic properties of the proposed estimator and derive an optimal choice for the variable bandwidth. Section 3 focuses on boundary effects. In Section 4, we investigate the performance of the local linear smoother with estimated variable bandwidth. The finite sample properties of the estimator are illustrated via simulations in Section 5. Some further remarks and discussions are given in Section 6. The last section contains the proofs of the results.

2. Asymptotic properties and optimal variable bandwidth.

First of all, we study the asymptotic properties of the local linear smoother (1.3) introduced in Section 1. In the following theorem we give an expression for the conditional mean squared error (MSE) of the estimator.

THEOREM 1. *Assume that $f_X(\cdot)$, $\alpha(\cdot)$, $m''(\cdot)$ and $\sigma(\cdot)$ are bounded functions, continuous at the point x, where x is in the interior of the support of $f_X(\cdot)$. Suppose that $\min_z \alpha(z) > 0$, $\limsup_{|u| \to \infty} |K(u)u^5| < \infty$ and $nh_n \to \infty$. Then, the conditional MSE of the estimator (1.3) at the point x is given by*

$$E\left[(\hat{m}(x) - m(x))^2 \mid X_1, \ldots, X_n\right] = \left(b_n^2(x) + v_n^2(x)\right)(1 + o_P(1)), \tag{2.1}$$

where

$$b_n(x) = \frac{1}{2} m''(x) \frac{s_2^2 - s_1 s_3}{s_2 s_0 - s_1^2} \left(\frac{h_n}{\alpha(x)} \right)^2 \tag{2.2}$$

and

$$v_n^2(x) = \left(\frac{\int_{-\infty}^{+\infty} [s_2 - u s_1]^2 K^2(u)\, du}{[s_2 s_0 - s_1^2]^2} \right) \frac{\alpha(x) \sigma^2(x)}{f_X(x) n h_n}, \tag{2.3}$$

with $s_l = \int_{-\infty}^{+\infty} K(u) u^l \, du$, $l = 0, 1, 2, 3$.

If we take $\alpha(\cdot) = 1$, the preceding result slightly generalizes the known result for the estimator with a constant bandwidth [see Fan (1992a)]. Here, we do not require that the kernel function integrates out to 1 and has mean zero. Theorem 1 has some important implications: no matter what kernel is used, the bias for the local linear fit is always of the second order. This is in contrast with the local constant fit, resulting in the Nadaraya–Watson estimator, since in this case the asymptotic bias is not of the second order unless the kernel function has mean zero. When the kernel function is a density with mean zero (i.e., $s_0 = 1$ and $s_1 = 0$), expression (2.1) reduces to

$$\begin{aligned} &\frac{1}{4} \left[m''(x) s_2 \left(\frac{h_n}{\alpha(x)} \right)^2 \right]^2 + \frac{\alpha(x) \sigma^2(x)}{f_X(x) n h_n} \int_{-\infty}^{+\infty} K^2(u)\, du \\ &\quad + o_P\left(h_n^4 + \frac{1}{n h_n} \right). \end{aligned} \tag{2.4}$$

This result is similar to that for the estimator with constant bandwidth, but now with h_n replaced by $h_n / \alpha(x)$.

REMARK 1. The condition $\min_z \alpha(z) > 0$ in Theorem 1 is not an obligatory one. The result of the theorem remains valid if $\alpha(\cdot)$ is nonnegative and continuous with at most a finite number of roots, and $\liminf_{|z| \to \infty} \alpha(z) > 0$. Note that the function $\alpha_{\mathrm{opt}}(\cdot)$, defined in (2.9), possibly only satisfies this weaker condition. This remark also applies to Theorems 2 and 4.

Next, we investigate the global behavior of the estimator. A commonly used, simple measure of global loss is the mean integrated squared error (MISE), obtained by taking the expectation of a weighted integrated squared error. Theorem 2 provides an expression for the conditional MISE. Let $W(\cdot)$ be a nonnegative, bounded weight function with bounded support $[a, b]$ which is contained in the interior of the support of $f_X(\cdot)$. Assume that $f_X(\cdot)$ is bounded away from zero on $[a, b]$.

To avoid technicalities in the proof of Theorem 2, we slightly modify the estimator (1.3) as follows:

$$\hat{m}^*(x) = \sum_{j=1}^{n} w_j Y_j \bigg/ \left(\sum_{j=1}^{n} w_j + n^{-2} \right). \tag{2.5}$$

The reason for introducing the factor n^{-2} in the denominator of (2.5) is to assure that this denominator is bounded away from zero. We emphasize, however, that this technical modification has no impact on the forthcoming results in this paper. Moreover, it has no practical implications.

THEOREM 2. *Assume that* $\alpha(\cdot)$, $m''(\cdot)$ *and* $\sigma(\cdot)$ *are bounded and continuous functions on* $[a, b]$, *and that* $f_X(\cdot)$ *is uniformly Lipschitz continuous of order* $r > 0$. *Suppose that* $\min_z \alpha(z) > 0$ *and that* $\int_{-\infty}^{+\infty}|K(u)u^j|\,du < \infty$ *for all* $j \geq 0$. *Then, the conditional MISE of the estimator* (2.5) *is*

$$\left[\int_{-\infty}^{+\infty}\left[b_n^2(x) + v_n^2(x)\right]W(x)\,dx\right](1 + o_P(1)),$$

provided that $h_n = dn^{-\gamma}$, *with constants* $d > 0$ *and* $0 < \gamma < 1$.

In case the kernel function K is a density with mean zero, an asymptotic expression for the conditional MISE is defined by

$$\text{AMISE}(\hat{m}, m) = \int_{-\infty}^{+\infty}\left[\frac{1}{4}\left(m''(x)s_2\left(\frac{h_n}{\alpha(x)}\right)^2\right)^2 + \frac{\alpha(x)\sigma^2(x)}{f_X(x)nh_n}\int_{-\infty}^{+\infty}K^2(u)\,du\right]W(x)\,dx. \tag{2.6}$$

Note that this expression is justified by Theorem 2 and the remark about the modification preceding it. Throughout the rest of this section we will work with this simplified conditional AMISE.

We now discuss the optimal choice of the function $\alpha(\cdot)$. In order to find such an optimal function we proceed as follows. We first minimize the AMISE (2.6) with respect to h_n. This yields the optimal nonvariable bandwidth

$$h_{n,\alpha} = \left(\frac{\int_{-\infty}^{+\infty}\alpha(x)\sigma^2(x)W(x)/f_X(x)\,dx\int_{-\infty}^{+\infty}K^2(u)\,du}{s_2^2\int_{-\infty}^{+\infty}[m''(x)]^2W(x)/\alpha^4(x)\,dx}\right)^{1/5}n^{-1/5}. \tag{2.7}$$

Substituting this optimal choice into (2.6) leads to

$$\text{AMISE}(\hat{m}, m) = \frac{5C_K}{4n^{4/5}}\left(\int_{-\infty}^{+\infty}[m''(x)]^2\frac{W(x)}{\alpha^4(x)}\,dx\left[\int_{-\infty}^{\infty}\alpha(x)\sigma^2(x)\frac{W(x)}{f_X(x)}\,dx\right]^4\right)^{1/5}, \tag{2.8}$$

where $C_K = s_2^{2/5}[\int_{-\infty}^{+\infty}K^2(u)\,du]^{4/5}$. We now minimize (2.8) with respect to $\alpha(\cdot)$. The solution to this optimization problem is established in the following theorem.

THEOREM 3. *The optimal variable bandwidth is given by*

$$\alpha_{\text{opt}}(x) = \begin{cases} b\left(\dfrac{f_X(x)[m''(x)]^2}{\sigma^2(x)}\right)^{1/5}, & \text{if } W(x) > 0, \\ \alpha^*(x), & \text{if } W(x) = 0, \end{cases} \tag{2.9}$$

where b is any arbitrarily positive constant and $\alpha^*(x)$ can be taken to be any positive value.

Note that the optimal variable bandwidth $\alpha_{\text{opt}}(\cdot)$ does not depend on the weight function $W(\cdot)$, that is, $\alpha_{\text{opt}}(\cdot)$ is intrinsic to the problem.

With the preceding optimal choice of $\alpha(\cdot)$, the optimal nonvariable bandwidth $h_{n,\alpha}$ in (2.7) is equal to

$$h_{n,\text{opt}} = b\left(\frac{\int_{-\infty}^{+\infty} K^2(u)\,du}{s_2^2}\right)^{1/5} n^{-1/5}. \tag{2.10}$$

An important feature is that this optimal choice of h_n does not depend on unknown functions. With these optimal choices of the nonvariable and the variable bandwidth, the AMISE (2.8) is given by

$$\text{AMISE}_{v,\text{opt}} = \frac{5C_K}{4n^{4/5}} \int_{-\infty}^{+\infty} [m''(x)]^{2/5} \left[\frac{\sigma^2(x)}{f_X(x)}\right]^{4/5} W(x)\,dx. \tag{2.11}$$

On the other hand, the expression for the AMISE (2.6) with $\alpha(\cdot) = 1$ and an optimal choice of the constant bandwidth is

$$\begin{aligned} &\text{AMISE}_{c,\text{opt}} \\ &\quad = \frac{5C_K}{4n^{4/5}} \left(\int_{-\infty}^{+\infty} [m''(x)]^2 W(x)\,dx \left[\int_{-\infty}^{+\infty} \sigma^2(x) \frac{W(x)}{f_X(x)}\,dx \right]^4 \right)^{1/5}. \end{aligned} \tag{2.12}$$

Now, it is easy to see that

$$\text{AMISE}_{v,\text{opt}} \le \text{AMISE}_{c,\text{opt}},$$

and this fact reflects one of the advantages of using a variable bandwidth.

The concept of variable bandwidth is intuitively appealing: A different amount of smoothing is used at different data locations. Even in case of slight misspecification of the optimal variable bandwidth $\alpha_{\text{opt}}(\cdot)$, the proposed method [with $h_{n,\text{opt}}$ given in (2.10) as the nonvariable bandwidth] can still achieve the optimal rate of convergence. Finally, the optimal variable bandwidth $\alpha_{\text{opt}}(\cdot)$ depends on $f_X(\cdot)$, $\sigma^2(\cdot)$ and $[m''(\cdot)]^2$ only through a $1/5$ power function. This implies that even if the unknown quantity is misestimated by, say, a factor 2, the resulting α_{opt} would differ only by a factor 1.15. Therefore, we expect that substitution of reasonable estimators for the unknown functions into (2.9) will lead to a good estimator for the regression function. Denote by $\hat{h}$ the optimal bandwidth $[h_{\text{opt}}(\cdot) = h_{n,\text{opt}}/\alpha_{\text{opt}}(\cdot)]$ estimated from the data. The decisive question for the practical application concerns the speed of convergence for the relative rate $\mathrm{AMISE}(\hat{h})/\mathrm{AMISE}(h_{\mathrm{opt}})$. This interesting question is, however, beyond the scope of the present paper.

Another intuitive choice for the variable bandwidth is $\alpha(x) = (f_X(x)/\sigma^2(x))$. Indeed, this choice implies that a large bandwidth is used at low-density design points and also at locations with large conditional variance. With such a variable bandwidth, the regression smoother (1.3) has a homogeneous variance [see (2.3)]. Hence, this intuitive choice of $\alpha(\cdot)$ can be viewed as a rule comparable to the one introduced in Breiman, Meisel and Purcell (1977), but now in the regression setup. In contrast with $\alpha_{\mathrm{opt}}(\cdot)$, this choice of $\alpha(\cdot)$ is not optimal in the sense that it does not minimize the conditional AMISE.

3. Boundary effects. Let $X_1, \dots, X_n$ be i.i.d. random variables with a density $f_X(\cdot)$, having bounded support. Without loss of generality we consider this support to be the interval $[0, 1]$. Theorem 1 provides an expression for the conditional MSE for points in the interior of $[0, 1]$. In this section we study the behavior of the estimator (1.3) at boundary points. Such an investigation is necessary since it is not obvious that an estimator has the same behavior at the boundary as in the interior of the support. For example, the Nadaraya–Watson (1964) estimator and the Gasser–Müller (1979) estimator both have so-called boundary effects. In other words, the rate of convergence of these estimators at the boundary points is slower than that for points in the interior of the support. In practical curve estimation, both estimators require a modification at the boundary. For detailed discussions see Gasser and Müller (1979).

We now investigate the behavior of the estimator (1.3) at left-boundary points. Put $x_n = ch_n$, with $c > 0$. Assume that $nh_n \to \infty$ and denote $\alpha_0 = \alpha(0+)$.

THEOREM 4. *Assume that $f_X(\cdot)$, $\alpha(\cdot)$, $m''(\cdot)$ and $\sigma(\cdot)$ are bounded on $[0, 1]$ and right continuous at the point 0. Suppose that $\min_{z \in [0,1]} \alpha(z) > 0$ and that $\limsup_{u \to -\infty} |K(u)u^5| < \infty$. Then, the conditional MSE of the estimator (1.3) at the boundary point x_n is given by*

$$\begin{aligned}\left\{\frac{1}{4}\left[m''(0+)\frac{s_{2,c}^2 - s_{1,c}s_{3,c}}{s_{2,c}s_{0,c} - s_{1,c}^2}\right]^2\left(\frac{h_n}{\alpha_0}\right)^4\right.\\ \left.+\left(\frac{\int_{-\infty}^{\alpha_0 c}[s_{2,c} - us_{1,c}]^2 K^2(u)\,du}{[s_{2,c}s_{0,c} - s_{1,c}^2]^2}\right)\frac{\alpha_0\sigma^2(0+)}{f_X(0+)nh_n}\right\}(1 + o_P(1)),\end{aligned} \tag{3.1}$$

where $s_{l,c} = \int_{-\infty}^{\alpha_0 c} K(u)u^l\,du$, $l = 0, 1, 2, 3$.

REMARK 2. In an analogous way we obtain expressions for the conditional MSE of the estimator at right-boundary points which are of the form $x_n =$

$1 - ch_n$. More precisely, the conditional MSE at $x_n = 1 - ch_n$ is

$$\left\{ \frac{1}{4} \left[m''(1-) \frac{s_{2,c}^2 - s_{1,c}s_{3,c}}{s_{2,c}s_{0,c} - s_{1,c}^2} \right]^2 \left(\frac{h_n}{\alpha_1} \right)^4 + \left(\frac{\int_{-\alpha_1 c}^{\infty} [s_{2,c} - us_{1,c}]^2 K^2(u)\, du}{[s_{2,c}s_{0,c} - s_{1,c}^2]^2} \right) \frac{\alpha_1 \sigma^2(1-)}{f_X(1-)nh_n} \right\} (1 + o_P(1)), \tag{3.2}$$

where now $s_{l,c} = \int_{-\alpha_1 c}^{\infty} K(u)u^l\, du$, $l = 0, 1, 2, 3$, with $\alpha_1 = \alpha(1-)$. These expressions hold through under conditions comparable to those in Theorem 4, but now translated to the right-side of the boundary.

It follows from Theorem 4 and Remark 2 that the estimator (1.3) does have the right behavior at the boundary. Indeed, its rate of convergence is not influenced by the position of the point under consideration. Hence, the local linear smoother does not require modifications at the boundary. So, it turns out that the local linear smoother has an additional advantage over other kernel-type estimators. The intuition behind this fact goes back to the construction of the local linear smoother [see also Fan (1992a)]. The local linear approximation which was used results into a second order approximation of the underlying regression function. This holds through at all points of the support, including boundary points.

We now study how the constant factor

$$b^2(\alpha_0 c) \equiv \left[\frac{s_{2,c}^2 - s_{1,c}s_{3,c}}{s_{2,c}s_{0,c} - s_{1,c}^2} \right]^2$$

in the squared bias [see expression (3.1)] and the constant factor in the variance

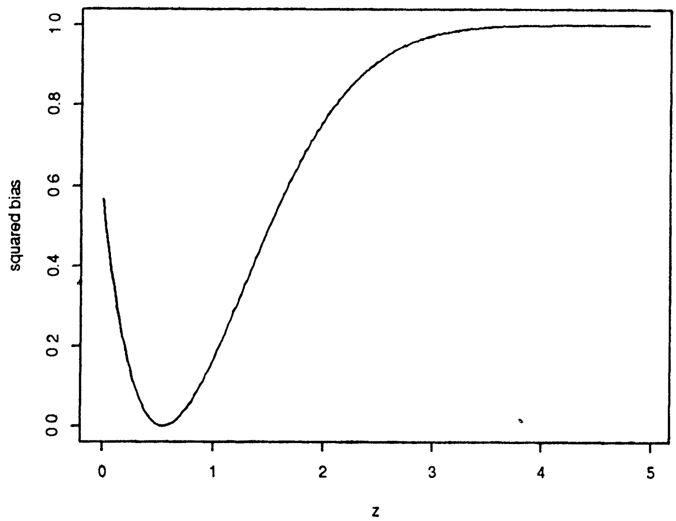
$$v(\alpha_0 c) \equiv \frac{\int_{-\infty}^{\alpha_0 c} [s_{2,c} - us_{1,c}]^2 K^2(u)\, du}{[s_{2,c}s_{0,c} - s_{1,c}^2]^2}$$

change with $\alpha_0 c$. Note that $\alpha_0 c$ measures how many effective bandwidths (i.e., h_n/α_0) the point $x_n = ch_n$ is away from the left boundary. We plot both functions $b^2(\cdot)$ and $v(\cdot)$ for three commonly used kernels:

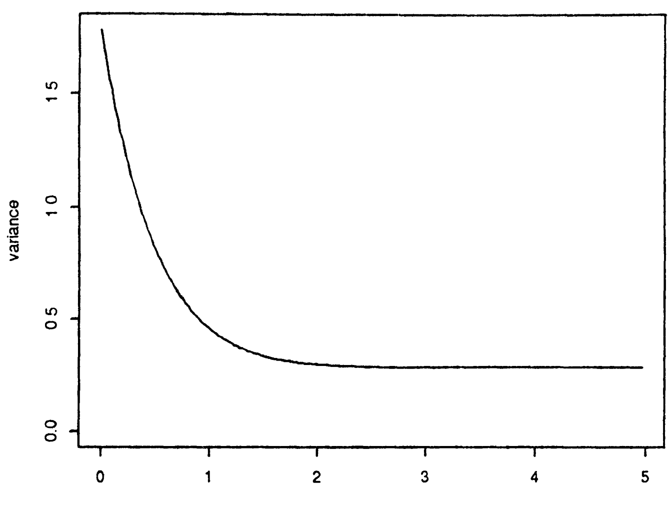
the standard normal kernel: $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2} \right)$,

the Epanechnikov kernel: $K(u) = \frac{3}{4}(1 - u^2)_+$,

the uniform kernel: $K(u) = 1_{u \in [-0.5, 0.5]}$.



(a)



(b)

FIG. 1. *Normal kernel.*

Note that Figures 1–3 show the same behavior. A first feature is that

$$\lim_{z\to\infty} b^2(z) = s_2^2 \quad \text{and} \quad \lim_{z\to\infty} v(z) = \int_{-\infty}^{+\infty} K^2(u)\, du,$$

and these limits are exactly the constant factors appearing, respectively, in the squared bias and the variance for an interior point. Further, it is clear from

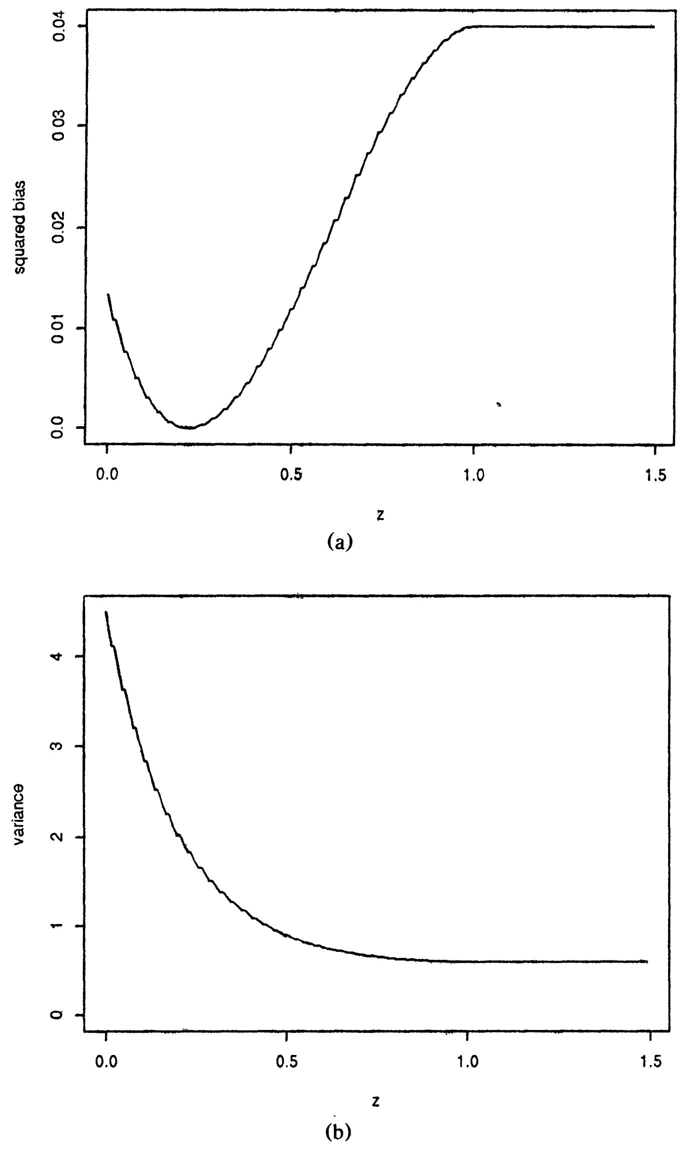
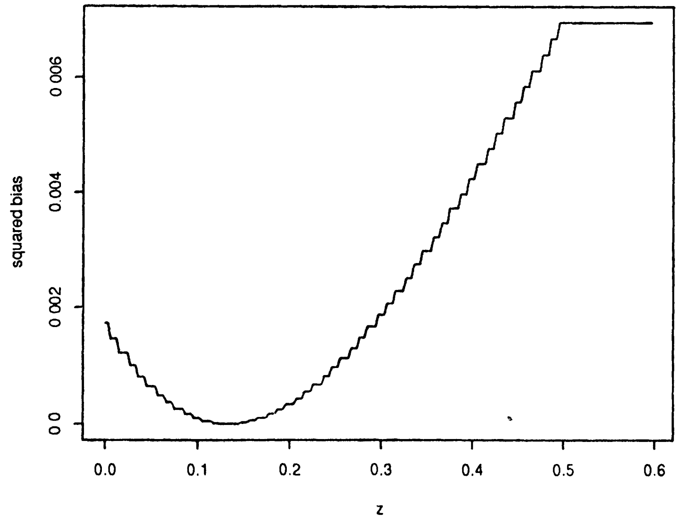
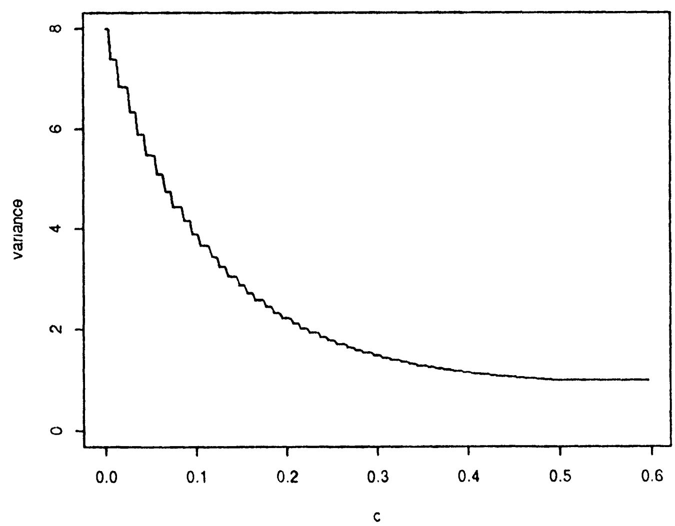


FIG. 2. *Epanechnikov kernel.*

Figures 1–3 that $b^2(z)$ is smaller than s_2^2 and that $v(z)$ is larger than $\int_{-\infty}^{+\infty} K^2(u)\,du$, for all values of z. This implies that the squared bias of the estimator (1.3) is smaller at a boundary point than at an interior point, at least if the value of m'' at each of these points is the same and the same amount of smoothing is used. On the other hand, the variance is larger at the boundary point. That the bias is smaller is due to the fact that one uses a linear



(a)



(b)

FIG. 3. *Uniform kernel.*

approximation on a smaller interval around the boundary point. The variance however tends to be larger, because on a smaller interval less observations contribute to computing the estimator.

4. Performance of the plug-in estimator. As already mentioned in Section 2, the optimal variable bandwidth $\alpha_{\text{opt}}(\cdot)$ depends on the unknown functions $f_X(\cdot)$, $m''(\cdot)$ and $\sigma^2(\cdot)$. Hence, practical implementation of the local linear smoother requires estimation of these unknown quantities. The estimated quantities are then substituted into the expression for $\alpha_{\mathrm{opt}}(\cdot)$. In this section, we justify such a "plug-in" procedure. This validates the applicability of the local linear smoother with variable bandwidth.

To emphasize the dependence of the local linear smoother (1.3) on the variable bandwidth $\alpha(\cdot)$, we denote, in this section, the estimator by $\hat{m}(x, \alpha)$. With $\hat{\alpha}_n(\cdot)$ an estimator of $\alpha(\cdot)$, we define the plug-in estimator as

$$\hat{m}(x, \hat{\alpha}_n) = \sum_{j=1}^{n} \hat{w}_j Y_j \Big/ \sum_{j=1}^{n} \hat{w}_j,$$

where

$$\hat{w}_j = \hat{\alpha}_n(X_j) K\left(\frac{x - X_j}{h_n} \hat{\alpha}_n(X_j)\right)\left[\hat{s}_{n,2} - (x - X_j)\hat{s}_{n,1}\right]$$

with

$$\hat{s}_{n,l} = \sum_{j=1}^{n} \hat{\alpha}_n(X_j) K\left(\frac{x - X_j}{h_n} \hat{\alpha}_n(X_j)\right)(x - X_j)^l, \qquad l = 0, 1, 2.$$

The following theorem shows that the plug-in estimator $\hat{m}(x, \hat{\alpha}_n)$ behaves asymptotically the same as $\hat{m}(x, \alpha)$.

THEOREM 5. *Suppose that the conditions of Theorem* 1 *hold. Let $\hat{\alpha}_n(\cdot)$ be a consistent estimator of $\alpha(\cdot)$ such that $\sup_z |\hat{\alpha}_n(z) - \alpha(z)| = o_P(a_n)$, where $a_n \to 0$. Assume that K is a uniformly Lipschitz continuous function such that $|u^3 K(u)| \le G(u)$ for all large $|u|$, where $G(u)$ is decreasing as $|u|$ increases and satisfies $G(a_n^{-1/5}) = o(h_n)$. Then,*

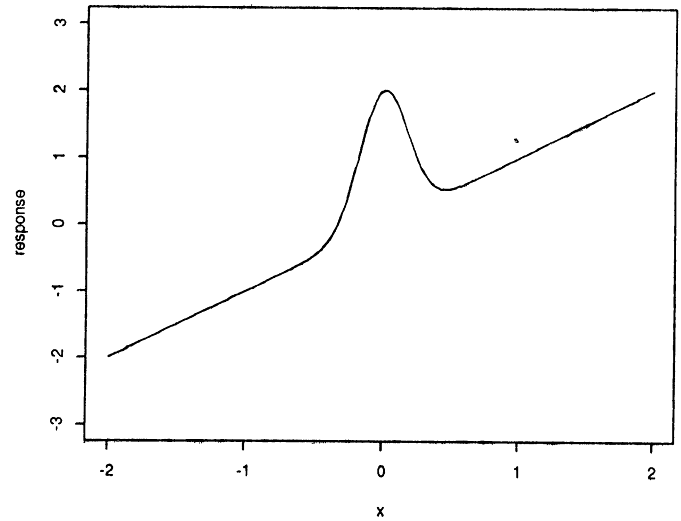
$$E\left[\left(\hat{m}(x, \hat{\alpha}_n) - \hat{m}(x, \alpha)\right)^2 \Big| X_1, \dots, X_n\right] = o_P\left(h_n^4 + \frac{1}{nh_n}\right).$$

Note that if K has bounded support, Theorem 1 states that any uniform consistent estimator of $\alpha(\cdot)$ will do the job.

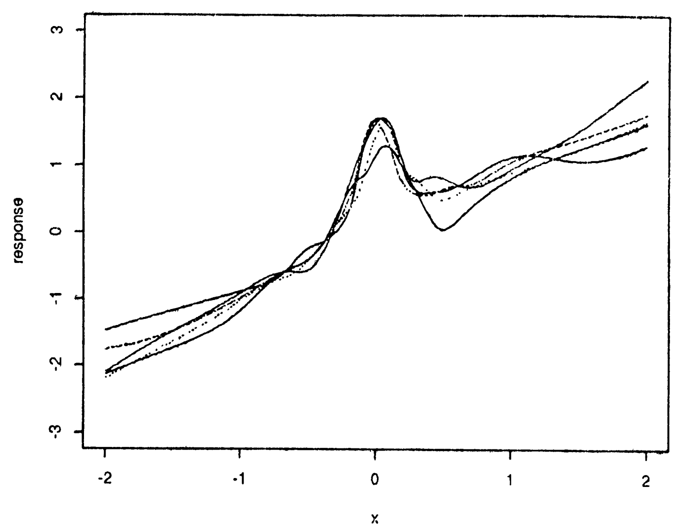
Furthermore, Theorem 5 provides a tool for obtaining an asymptotic normality result for the plug-in estimator via that for the local linear smoother. Proving asymptotic normality for $\hat{m}(x, \alpha)$ is, however, beyond the scope of this paper.

A simple way to estimate the unknown functions $f_X(\cdot)$, $m''(\cdot)$ and $\sigma^2(\cdot)$ is as follows. Starting with constant bandwidths, one can estimate $f_X(\cdot)$ and $m''(\cdot)$ using cross-validation techniques. Further, an estimator for $\sigma^2(\cdot)$ is based on the residuals $\hat{Y}_j = Y_j - \hat{m}(X_j)$. These preliminary estimators are then substituted into the expression for $\alpha_{\mathrm{opt}}(\cdot)$, and the resulting $\hat{\alpha}_{n,\mathrm{opt}}(\cdot)$ is used to calculate $\hat{m}(\cdot, \hat{\alpha}_{n,\mathrm{opt}})$.

5. Simulations. In this section, we illustrate the performance of the local linear smoother. It will be seen that the proposed method does a reasonable job for curve estimation, including the fact that it captures the shape of the theoretical curves. A finite sample comparison with other kernel-type estimation methods can be found in Fan (1992a). The simulation studies presented here are only a small part of larger sets of simulations which were carried out, but the simulations given in this section are typical. For each of the following examples we used the standard normal kernel function, simply for convenience. Further, we used the optimal nonvariable and variable bandwidths [see



(a)



(b)

FIG. 4. (a) *True regression curve.* (b) *Estimated regression functions.*

expressions (2.10) and (2.9)]. In a first example, we simulated 200 data points from a normal regression model

$$Y_j = m(X_j) + \varepsilon_j,$$

where $X_j \sim_{\text{i.i.d.}} N(0,1)$, $\varepsilon_j \sim_{\text{i.i.d.}} N(0, 0.7^2)$ and

$$m(x) = x + 2\exp(-16x^2).$$

Hence, here we have to detect linearity and a bump. Remark that only about

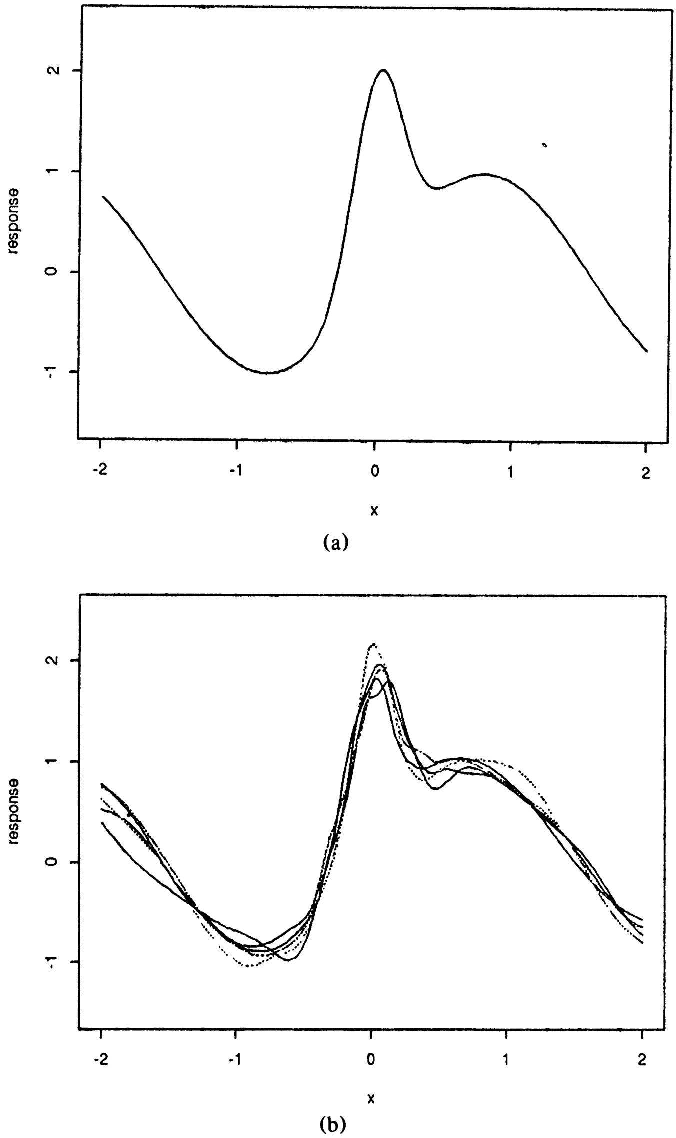
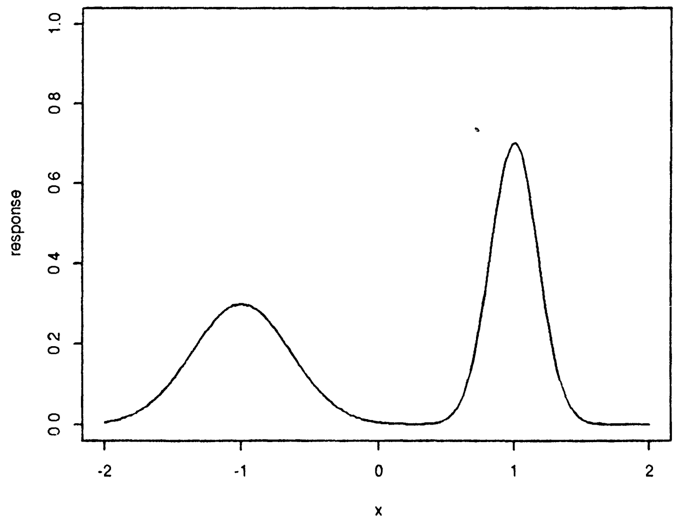


FIG. 5. (a) *True regression curve*. (b) *Estimated regression functions*.

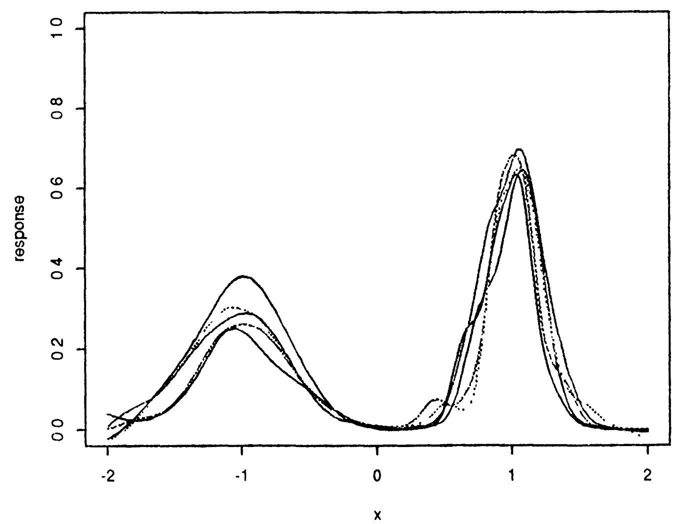
5% of the design points lie outside the interval $[-2, 2]$. Therefore, this interval can be viewed as a bounded support corresponding to the design density. Figure 4 presents the true regression curve and five estimated regression functions, each based on one simulation.

Next, we considered a normal regression model

$$Y_j = m(X_j) + \varepsilon_j,$$



(a)



(b)

FIG. 6. (a) *True regression curve*. (b) *Estimated regression functions*.

where $X_j \sim_{\text{i.i.d.}} N(0,1)$, $\varepsilon_j \sim_{\text{i.i.d.}} N(0, 0.5^2)$ and

$$m(x) = \sin(2x) + 2\exp(-16x^2).$$

The results of five simulations (sample size 200) can be found in Figure 5.

As a last example, we imposed the binary response model

$$P(Y_j = i | X_j) = \begin{cases} m(X_j), & \text{if } i = 1, \\ 1 - m(X_j), & \text{if } i = 0, \end{cases}$$

where $X_j \sim_{\text{i.i.d.}}$ uniform $(-2, 2)$ and

$$m(x) = 0.3\exp\big(-4(x+1)^2\big) + 0.7\exp\big(-16(x-1)^2\big).$$

Note that in this example $\sigma^2(x) = m(x)(1 - m(x))$. The true regression function and its simulated estimates (sample size 400) are shown in Figure 6.

A referee pointed out that the bias of the preceding estimates can be further reduced by using a local-quadratic approximation to the mean regression function instead of a local linear approximation. See, for example, Cleveland and Devlin (1988). Note also that the regression curves considered here are complicated, and that the local linear smoother with variable bandwidth is capable of capturing the various shapes.

6. Further discussions. In the context of density estimation, Abramson (1982) aims at choosing a variable bandwidth in order to reduce the order of the bias of the kernel estimator for f. This leads to the choice $h_n/f^{1/2}(X_j)$ for the smoothing parameter, which is known as the square root law. Also see Silverman (1986) for expressions of the asymptotic bias. Hall (1990) considers estimators of the form

$$\hat{\theta}(x) = (nh_n)^{-1} \sum_{j=1}^{n} Y_j \alpha(X_j) K^{(t)}\left(\frac{x - X_j}{h_n} \alpha(X_j) \right), \tag{6.1}$$

where $K^{(t)}$ denotes the tth derivative of the symmetric probability density function K. Examples include estimation of the tth derivative of a density function and estimation of the mean regression function. In his paper Hall provides a simple approach for calculating the bias of (6.1). In the special case of estimating the mean regression function, Hall (1990) proposes to use the estimator

$$\frac{(nh_{n,1})^{-1} \Sigma_{j=1}^{n} Y_j \alpha_1(X_j) K\big(((x - X_j)/h_{n,1})\alpha_1(X_j)\big)}{(nh_{n,2})^{-1} \Sigma_{j=1}^{n} \alpha_2(X_j) K\big(((x - X_j)/h_{n,2})\alpha_2(X_j)\big)}, \tag{6.2}$$

which is the ratio of two quantities having the form (6.1) with $t = 0$. The function $\alpha_1(\cdot)$ [respectively, $\alpha_2(\cdot)$] is chosen in such a way that it reduces the bias of the numerator (respectively, the denominator). Heuristically, the resulting estimator will have a reduced bias. Basically, this reduction is due to the fact that the estimator (6.2) does not have total weight 1, which already introduces a kind of bias correction. Note that the estimator (1.3), however, has total weight 1. In such a situation, there is no hope of finding a variable bandwidth which results in a reduction of the order of the bias. In other words, there will be no equivalent of the square root law in this regression setup. An attempt to reduce the order of the bias would be: Estimate the bias of the estimator, subtract it from the estimator and define this as the new one. But, this would result in a linear smoother which has total weight not equal to 1.

An alternative way of introducing the idea of *variable* bandwidth is to consider $h_n/\beta(x_0)$ as the smoothing parameter at the location point x_0. Knowing the value $\beta(x_0)$ suffices to estimate the regression function at this point. Hence, this type of variable bandwidth can be viewed as a *local* variable bandwidth. The variable bandwidth $h_n/\alpha(X_j)$, however, requires knowledge of the function α at each observation X_j. Therefore, one could refer to the latter bandwidth as a *global* variable bandwidth. For a location point x_0 such that $|m''(x_0)|$ is small, the optimal local variable bandwidth $h_n/\beta_{\text{opt}}(x_0)$ will be very large. The resulting estimator will misestimate the true value of $m(x_0)$. This illustrates that an estimator based on a local variable bandwidth relies too much on the particular value of $\beta(x_0)$. For Gasser–Müller type estimation of regression curves, using the idea of local variable bandwidth, see Müller and Stadtmüller (1987).

7. Proofs. Theorems 1 and 4 will be proved along the same lines. The proof of Theorem 4 is more involved and requires more details. For this reason we decide to prove Theorem 4 before Theorem 1 and hence postpone the proofs of Theorems 1 and 2.

PROOF OF THEOREM 3. We have to minimize (2.8) with respect to $\alpha(\cdot)$. First of all, note that

$$\begin{aligned}\min_{\alpha} \int_{-\infty}^{+\infty} [m''(x)]^2 W(x)/\alpha^4(x)\,dx \left[\int_{-\infty}^{+\infty} \alpha(x)\sigma^2(x)W(x)/f_X(x)\,dx\right]^4 \\ = \min_{a} a^4 \min_{\alpha\in\mathscr{F}_a} \int_{-\infty}^{+\infty} [m''(x)]^2 W(x)/\alpha^4(x)\,dx,\end{aligned} \tag{7.1}$$

where $\mathscr{F}_a = \{\alpha(\cdot) \geq 0\colon \int_{-\infty}^{+\infty}\alpha(x)\sigma^2(x)W(x)/f_X(x)\,dx = a\}$. In order to solve the second minimization problem in (7.1), we use the method of Lagrange multipliers. Hence, we search for the minimum of

$$\int_{-\infty}^{+\infty} [m''(x)]^2 W(x)/\alpha^4(x)\,dx + 4\lambda \int_{-\infty}^{+\infty} \alpha(x)\sigma^2(x)W(x)/f_X(x)\,dx$$

with respect to α. This translates into minimizing

$$[m''(x)]^2 W(x)/\alpha^4(x) + 4\lambda\alpha(x)\sigma^2(x)W(x)/f_X(x) \tag{7.2}$$

for each x. The solution to problem (7.2) is given by

$$\alpha_\lambda(x) = \begin{cases} \left(\dfrac{f_X(x)[m''(x)]^2}{\lambda\sigma^2(x)}\right)^{1/5}, & \text{if } W(x) > 0, \\ \alpha^*(x), & \text{if } W(x) = 0, \end{cases} \tag{7.3}$$

where $\alpha^*(x)$ can be taken to be any nonnegative value and λ is chosen so that $\alpha_\lambda \in \mathscr{F}_a$. Denote this choice of λ by λ_a. Substituting the solution (7.3) into (7.1), we find that the objective function for the first minimization (that in terms of a) does not depend on a. Hence, any choice of a and, therefore, of λ_a is appropriate. This completes the proof. □

In the sequel we prove Theorem 4. This will involve the following two lemmas.

LEMMA 1. *Assume that* $f_X(\cdot)$, $\alpha(\cdot)$ *and* $m''(\cdot)$ *are bounded on* $[0,1]$ *and right continuous at the point* 0. *Suppose that* $\min_{z\in[0,1]}\alpha(z) > 0$ *and that* $\limsup_{u\to-\infty}|K(u)u^{l+4}| < \infty$ *for a nonnegative integer* l. *Then,*

$$\sum_{j=1}^n \alpha(X_j)K\left(\frac{x_n - X_j}{h_n}\alpha(X_j)\right)R(X_j)(x_n - X_j)^l$$

$$= \frac{n}{2}\alpha_0\left(\frac{h_n}{\alpha_0}\right)^{l+3} m''(0+)f_X(0+)\int_{-\infty}^{\alpha_0 c} K(u)u^{l+2}\,du(1 + o_P(1)),$$

where $R(X_j) = m(X_j) - m(x_n) + m'(x_n)(x_n - X_j)$.

PROOF. Throughout this proof, we use the notations d_j, $j = 1,\dots,8$, for arbitrarily positive constants. Let

$$Z_{n,j} = \alpha(X_j)K\big(((x_n - X_j)/h_n)\alpha(X_j)\big)R(X_j)(x_n - X_j)^l$$

and note that

$$\sum_{j=1}^n Z_{n,j} = E\sum_{j=1}^n Z_{n,j} + O_P\left(\sqrt{\mathrm{var}\left(\sum_{j=1}^n Z_{n,j}\right)}\right) \tag{7.4}$$

$$= nEZ_{n,1} + O_P\left(\sqrt{nEZ_{n,1}^2}\right).$$

In the sequel, we will calculate the first two moments of $Z_{n,1}$. By a change of variable, we obtain

$$EZ_{n,1} = \int_0^1 \alpha(y)K\left(\frac{x_n - y}{h_n}\alpha(y)\right)R(y)(x_n - y)^l f_X(y)\,dy$$

$$= h_n^{l+1}\int_{c-1/h_n}^{c} \alpha(x_n - zh_n)K(z\alpha(x_n - zh_n)) \tag{7.5}$$

$$\times R(x_n - zh_n)z^l f_X(x_n - zh_n)\,dz.$$

A two-term Taylor expansion gives that

$$R(x_n - zh_n) = \tfrac{1}{2}m''(\xi_n)(zh_n)^2, \tag{7.6}$$

where ξ_n is between x_n and $x_n - zh_n$. We will now approximate (7.5) by $h_n^{l+3}A_n/2$ with

$$A_n = \int_{c-1/h_n}^{c} \alpha_0 K(z\alpha_0)m''(0+)z^{l+2}f_X(0+)\,dz. \tag{7.7}$$

Therefore, we study the following difference:

$$\left|\int_{c-1/h_n}^{c}\alpha(x_n - zh_n)K(z\alpha(x_n - zh_n))m''(\xi_n)z^{l+2}f_X(x_n - zh_n)\,dz - A_n\right|$$

$$\begin{aligned}(7.8)\quad &\le \left[\int_{c-1/h_n}^{-M} + \int_{-M}^{c}\right]|\alpha(x_n - zh_n)K(z\alpha(x_n - zh_n))m''(\xi_n)\\ &\qquad\times f_X(x_n - zh_n) - \alpha_0 K(z\alpha_0)m''(0+)f_X(0+)|\,|z|^{l+2}\,dz\\ &\equiv I_{n,1} + I_{n,2},\end{aligned}$$

with M a fixed positive number large enough such that

$$|K(u)u^{l+4}| < d_1 \quad \text{for all } u < -M\alpha^*, \tag{7.9}$$

where $\alpha^* = \min_{z\in[0,1]}\alpha(z)$. Note that the tail condition on K guarantees the existence of such a number. Applying the dominated convergence theorem, together with the continuity assumptions, we obtain that $\lim_{n\to\infty} I_{n,2} = 0$. The term $I_{n,1}$ is bounded by

$$\begin{aligned}(7.10)\quad &\int_{c-1/h_n}^{-M}|\alpha(x_n - zh_n)K(z\alpha(x_n - zh_n))m''(\xi_n)f_X(x_n - zh_n)|\,|z|^{l+2}\,dz\\ &\qquad + d_2\int_{-\infty}^{-M}|\alpha_0 K(z\alpha_0)|\,|z|^{l+2}\,dz \equiv J_{n,1} + J_{n,2}.\end{aligned}$$

First note that

$$J_{n,2} = d_3\int_{-\infty}^{-M\alpha_0}|K(u)u^{l+2}|\,du,$$

and this tends to zero as $M \to \infty$. Using (7.9), the boundedness of $\alpha(\cdot)$, $m''(\cdot)$ and $f_X(\cdot)$ and the definition of α_0, we find

$$\begin{aligned}J_{n,1} &\le d_4\int_{c-1/h_n}^{-M}|K(z\alpha(x_n - zh_n))|\,|z|^{l+2}\,dz\\ &\le d_4\int_{c-1/h_n}^{-M} d_1|z\alpha(x_n - zh_n)|^{-l-4}|z|^{l+2}\,dz\\ &\le d_1 d_4(\alpha^*)^{-l-4}\int_{-\infty}^{-M}|z|^{-2}\,dz,\end{aligned}$$

which tends to zero as $M \to \infty$. By (7.10), we conclude that $\lim_{M\to\infty}\lim_{n\to\infty} I_{n,1} = 0$. Hence (7.8) leads to

$$\begin{aligned}&\int_{c-1/h_n}^{c}\alpha(x_n - zh_n)K(z\alpha(x_n - zh_n))m''(\xi_n)z^{l+2}f_X(x_n - zh_n)\,dz\\ (7.11)\quad &\qquad = A_n + o(1)\\ &\qquad = \int_{-\infty}^{c}\alpha_0 K(z\alpha_0)m''(0+)z^{l+2}f_X(0+)\,dz(1 + o(1))\\ &\qquad = m''(0+)f_X(0+)\alpha_0^{-l-2}\int_{-\infty}^{\alpha_0 c}K(u)u^{l+2}\,du(1 + o(1)).\end{aligned}$$

Finally, by (7.5), (7.6) and (7.11) we get

$$EZ_{n,1} = \frac{\alpha_0}{2}\left(\frac{h_n}{\alpha_0}\right)^{l+3} m''(0+) f_X(0+) \int_{-\infty}^{\alpha_0 c} K(u) u^{l+2}\, du (1 + o(1)). \tag{7.12}$$

For the second moment, we proceed as follows. Using (7.6) and (7.9), we obtain

$$\begin{aligned} EZ_{n,1}^2 &= \frac{1}{4} h_n^{2l+1} \int_{c-1/h_n}^{c} \alpha^2(x_n - zh_n) K^2(z\alpha(x_n - zh_n)) \\ &\qquad \times [m''(\xi_n)]^2 (zh_n)^4 z^{2l} f_X(x_n - zh_n)\, dz \\ &\le h_n^{2l+5} d_5 \int_{c-1/h_n}^{c} K^2(z\alpha(x_n - zh_n)) z^{2l+4}\, dz \\ &= h_n^{2l+5} d_5 \left[\int_{-M}^{c} + \int_{c-1/h_n}^{-M}\right] K^2(z\alpha(x_n - zh_n)) z^{2l+4}\, dz \\ &\le h_n^{2l+5} d_5 \left[d_6 + d_7 \int_{c-1/h_n}^{-M} [z\alpha(x_n - zh_n)]^{-2l-8} z^{2l+4}\, dz\right] \\ &\le d_8 h_n^{2l+5}. \end{aligned} \tag{7.13}$$

Combining (7.4), (7.12) and (7.13), along with $nh_n \to \infty$, completes the proof. □

LEMMA 2. *Assume that* $f_X(\cdot)$, $\alpha(\cdot)$, $L(\cdot)$ *and* $S(\cdot)$ *are bounded on* $[0, 1]$ *and right continuous at the point* 0. *Suppose that* $\min_{z \in [0,1]} \alpha(z) > 0$, *and that* $\limsup_{u \to -\infty} |L(u) u^{l+2}| < \infty$ *for a nonnegative integer* l. *Then,*

$$\sum_{j=1}^{n} \alpha(X_j) L\left(\frac{x_n - X_j}{h_n} \alpha(X_j)\right) S(X_j)(x_n - X_j)^l$$
$$= n\alpha_0 \left(\frac{h_n}{\alpha_0}\right)^{l+1} S(0+) f_X(0+) \int_{-\infty}^{\alpha_0 c} L(u) u^l\, du (1 + o_P(1)).$$

The proof follows the same lines as that of Lemma 1.

PROOF OF THEOREM 4. The conditional MSE of the estimator (1.3) is given by

$$\frac{\left[\sum_1^n w_j (m(X_j) - m(x_n))\right]^2}{\left(\sum_1^n w_j\right)^2} + \frac{\sum_1^n w_j^2 \sigma^2(X_j)}{\left(\sum_1^n w_j\right)^2}. \tag{7.14}$$

Recall the definition of $s_{n,l}$ [see (1.5)]. Applying Lemma 2, with $L = K$ and $S = 1$, we obtain

$$s_{n,l} = n\alpha_0 (h_n/\alpha_0)^{l+1} s_{l,c} f_X(0+)(1 + o_P(1)), \qquad l = 0, 1, 2, \tag{7.15}$$

and hence,

$$\sum_1^n w_j = s_{n,2}s_{n,0} - s_{n,1}^2 \tag{7.16}$$
$$= n^2\alpha_0^2(h_n/\alpha_0)^4(s_{2,c}s_{0,c} - s_{1,c}^2)f_X^2(0+)(1 + o_P(1)).$$

Further, Lemma 1 with $l = 0$ yields

$$\sum_{j=1}^n \alpha(X_j)K\left(\frac{x_n - X_j}{h_n}\alpha(X_j)\right)R(X_j) \tag{7.17}$$
$$= \frac{n}{2}\alpha_0\left(\frac{h_n}{\alpha_0}\right)^3 m''(0+)f_X(0+)s_{2,c}(1 + o_P(1)).$$

Similarly, Lemma 1 with $l = 1$ leads to

$$\sum_{j=1}^n \alpha(X_j)K\left(\frac{x_n - X_j}{h_n}\alpha(X_j)\right)R(X_j)(x_n - X_j) \tag{7.18}$$
$$= \frac{n}{2}\alpha_0\left(\frac{h_n}{\alpha_0}\right)^4 m''(0+)f_X(0+)s_{3,c}(1 + o_P(1)).$$

Now, since $\sum_{j=1}^n w_j(x_n - X_j) = 0$, we obtain from (7.15), (7.17) and (7.18) that

$$\sum_{j=1}^n w_j[m(X_j) - m(x_n)]$$
$$= s_{n,2}\sum_{j=1}^n \alpha(X_j)K\left(\frac{x_n - X_j}{h_n}\alpha(X_j)\right)R(X_j) \tag{7.19}$$
$$- s_{n,1}\sum_{j=1}^n \alpha(X_j)K\left(\frac{x_n - X_j}{h_n}\alpha(X_j)\right)R(X_j)(x_n - X_j)$$
$$= \frac{n^2}{2}\alpha_0^2\left(\frac{h_n}{\alpha_0}\right)^6 m''(0+)f_X^2(0+)[s_{2,c}^2 - s_{1,c}s_{3,c}](1 + o_P(1)).$$

Next, we write

$$\sum_1^n w_j^2\sigma^2(X_j) = s_{n,2}^2\sum_1^n \alpha^2(X_j)K^2\left(\frac{x_n - X_j}{h_n}\alpha(X_j)\right)\sigma^2(X_j)$$
$$- 2s_{n,1}s_{n,2}\sum_1^n \alpha^2(X_j)K^2\left(\frac{x_n - X_j}{h_n}\alpha(X_j)\right)(x_n - X_j)\sigma^2(X_j)$$
$$+ s_{n,1}^2\sum_1^n \alpha^2(X_j)K^2\left(\frac{x_n - X_j}{h_n}\alpha(X_j)\right)(x_n - X_j)^2\sigma^2(X_j)$$

and apply Lemma 2 to each of the three terms. This yields

$$\sum_1^n w_j^2\sigma^2(X_j) = n^3\alpha_0^4(h_n/\alpha_0)^7\sigma^2(0+)f_X^3(0+) \times\int_{-\infty}^{\alpha_0 c}[s_{2,c} - us_{1,c}]^2K^2(u)\,du(1 + o_P(1)). \tag{7.20}$$

The result now follows from (7.14), (7.16), (7.19) and (7.20). □

In order to prove Theorem 1, we need the following lemmas. Note that they are comparable with Lemmas 1 and 2.

LEMMA 3. *Assume that* $f_X(\cdot)$, $\alpha(\cdot)$ *and* $m''(\cdot)$ *are bounded functions, continuous at the point* x, *which is in the interior of the support of* $f_X(\cdot)$. *Suppose that* $\min_z \alpha(z) > 0$ *and that* $\limsup_{|u|\to\infty}|K(u)u^{l+4}| < \infty$, *for a nonnegative integer* l. *Then*,

$$\sum_{j=1}^n \alpha(X_j)K\left(\frac{x - X_j}{h_n}\alpha(X_j)\right)R(X_j)(x - X_j)^l = \frac{n}{2}\alpha(x)\left(\frac{h_n}{\alpha(x)}\right)^{l+3} m''(x)f_X(x)\int_{-\infty}^{+\infty}K(u)u^{l+2}\,du(1 + o_P(1)),$$

where $R(X_j) = m(X_j) - m(x) + m'(x)(x - X_j)$.

The basic ideas of the proof are similar to those in the proof of Lemma 1. We omit the proof.

LEMMA 4. *Assume that* $f_X(\cdot)$, $\alpha(\cdot)$, $L(\cdot)$ *and* $S(\cdot)$ *are bounded functions, continuous at the point* x, *which is in the interior of the support of* $f_X(\cdot)$. *Suppose that* $\min_z \alpha(z) > 0$ *and that* $\limsup_{|u|\to\infty}|L(u)u^{l+2}| < \infty$ *for a nonnegative integer* l. *Then*,

$$\sum_{j=1}^n \alpha(X_j)L\left(\frac{x - X_j}{h_n}\alpha(X_j)\right)S(X_j)(x - X_j)^l = n\alpha(x)\left(\frac{h_n}{\alpha(x)}\right)^{l+1} S(x)f_X(x)\int_{-\infty}^{+\infty}L(u)u^l\,du(1 + o_P(1)).$$

The proof is similar to the proof of Lemma 3.

PROOF OF THEOREM 1. The proof follows the same lines as that of Theorem 4, using Lemmas 3 and 4 instead of Lemmas 1 and 2. □

PROOF OF THEOREM 2. Denote $d_n(x) = E[(\hat{m}^*(x) - m(x))^2|X_1,\dots,X_n] - b_n^2(x) - v_n^2(x)$. The proof of Theorem 1 in Fan (1992b) yields that

$$\frac{Ed_n^2(x)}{h_n^4 + (nh_n)^{-1}} = o(1), \qquad \forall\, x\in[a,b],$$

and moreover $(Ed_n^2(x))/(h_n^4 + (nh_n)^{-1})$ is bounded uniformly in x and n. Using the Cauchy–Schwarz inequality, it follows that

$$E\int_a^b |d_n(x)W(x)|\,dx = \int_a^b E|d_n(x)W(x)|\,dx$$
$$\leq (b-a)^{1/2}\left(\int_a^b E|d_n(x)W(x)|^2\,dx\right)^{1/2}$$
$$= o\left(h_n^4 + (nh_n)^{-1}\right).$$

Since L_1-convergence implies convergence in probability, we conclude that

$$\int_a^b d_n(x)W(x)\,dx = o_P\left(h_n^4 + (nh_n)^{-1}\right),$$

which proves the theorem. □

In what follows we will prove Theorem 5. The proof will rely on the next two lemmas.

LEMMA 5. *Assume that $f_X(\cdot)$, $\alpha(\cdot)$ and $m''(\cdot)$ are bounded functions. Let $\hat{\alpha}_n(\cdot)$ be a consistent estimator of $\alpha(\cdot)$ such that $\sup_z|\hat{\alpha}_n(z) - \alpha(z)| = o_P(a_n)$, where $a_n \to 0$. Assume that K is a uniformly Lipschitz continuous function such that $|u^{l+2}K(u)| \leq G(u)$ for all large $|u|$, where $G(u)$ is decreasing as $|u|$ increases and satisfies $G(a_n^{-1/(l+4)}) = o(h_n)$ for a nonnegative integer l. Further, suppose that $\min_z \alpha(z) > 0$. Then,*

$$\sum_{j=1}^n \left[\hat{\alpha}_n(X_j)K\left(\frac{x-X_j}{h_n}\hat{\alpha}_n(X_j)\right) - \alpha(X_j)K\left(\frac{x-X_j}{h_n}\alpha(X_j)\right)\right] \times R(X_j)(x-X_j)^l = o_P(nh_n^{l+3}), \tag{7.21}$$

where $R(X_j) = m(X_j) - m(x) + m'(x)(x - X_j)$.

PROOF. In this proof d_j, $j = 1, 2, 3, 4$, denote positive constants. Let

$$I = \left\{j: \left|\frac{x-X_j}{h_n}\right| \leq 2(\alpha^*)^{-1}a_n^{-1/(l+4)}\right\},$$

where $\alpha^* = \min_z \alpha(z)$. Denote the left-hand side of (7.21) by $D_n(x)$ and write

$$D_n(x) = \left[\sum_{j\in I} + \sum_{j\notin I}\right]\left[\hat{\alpha}_n(X_j)K\left(\frac{x-X_j}{h_n}\hat{\alpha}_n(X_j)\right) - \alpha(X_j)K\left(\frac{x-X_j}{h_n}\alpha(X_j)\right)\right]R(X_j)(x-X_j)^l \tag{7.22}$$
$$\equiv D_{n,1}(x) + D_{n,2}(x).$$

First of all, note that

$$\#(I) = n\left[\hat{F}_n\left(x + 2(\alpha^*)^{-1}a_n^{-1/(l+4)}h_n\right) - \hat{F}_n\left(x - 2(\alpha^*)^{-1}a_n^{-1/(l+4)}h_n\right)\right],$$

where $\hat{F}_n(x)$ is the empirical distribution function of $X_1, \dots, X_n$. Let $F_X(\cdot)$ denote the corresponding distribution function of the X_j's. It is clear that

$$\begin{aligned} E\#(I) &= n\left[F_X\left(x + 2(\alpha^*)^{-1}a_n^{-1/(l+4)}h_n\right) - F_X\left(x - 2(\alpha^*)^{-1}a_n^{-1/(l+4)}h_n\right)\right] \\ &= O\left(na_n^{-1/(l+4)}h_n\right), \end{aligned}$$

which implies that

$$\#(I) = O_P\left(na_n^{-1/(l+4)}h_n\right). \tag{7.23}$$

We will now deal with each of the two terms in (7.22), starting with the first one. Using the conditions on $K(\cdot)$, $\alpha(\cdot)$, $\hat{\alpha}_n(\cdot)$ and $m''(\cdot)$, and incorporating the definition of I and (7.23), we obtain

$$\begin{aligned} &|D_{n,1}(x)| \\ &\le h_n^l \sum_{j\in I}\left|\left[\hat{\alpha}_n(X_j)K\left(\frac{x-X_j}{h_n}\hat{\alpha}_n(X_j)\right)\right.\right. \\ &\qquad \left.\left. -\alpha(X_j)K\left(\frac{x-X_j}{h_n}\alpha(X_j)\right)\right]R(X_j)\left(\frac{x-X_j}{h_n}\right)^l\right| \\ &\le h_{nn}^l \sum_{j\in I}\left|\alpha(X_j)\left[K\left(\frac{x-X_j}{h_n}\hat{\alpha}_n(X_j)\right)\right.\right. \\ &\qquad \left.\left. -K\left(\frac{x-X_j}{h_n}\alpha(X_j)\right)\right]R(X_j)\left(\frac{x-X_j}{h_n}\right)^l\right| \qquad (7.24) \\ &\quad + h_n^l \sum_{j\in I}\left|\left(\hat{\alpha}_n(X_j) - \alpha(X_j)\right)K\left(\frac{x-X_j}{h_n}\hat{\alpha}_n(X_j)\right)R(X_j)\left(\frac{x-X_j}{h_n}\right)^l\right| \\ &\le d_1 h_n^l \sum_{j\in I}\left[\left|\frac{x-X_j}{h_n}\right|\left|\hat{\alpha}_n(X_j) - \alpha(X_j)\right|\right]\frac{1}{2}|m''(\xi_j)|(x-X_j)^2\left|\frac{x-X_j}{h_n}\right|^l \\ &\quad + d_2 h_n^l \sum_{j\in I}\left|\left(\hat{\alpha}_n(X_j) - \alpha(X_j)\right)\frac{1}{2}|m''(\xi_j)|(x-X_j)^2\left(\frac{x-X_j}{h_n}\right)^l\right| \\ &\le d_3 \sup_z|\hat{\alpha}_n(z) - \alpha(z)|h_n^{l+2}\left[a_n^{-(l+3)/(l+4)} + a_n^{-(l+2)/(l+4)}\right]\#(I) \\ &= o_P\left(nh_n^{l+3}\right), \end{aligned}$$

where ξ_j lies between x and X_j. For the second term in (7.22) we rely on the conditions on $\alpha(\cdot)$ and $m''(\cdot)$, and the tail condition on K, and find

$$\begin{aligned}
|D_{n,2}(x)| & \\
\le h_n^{l+2} \sum_{j\notin I} & \left| K\left(\frac{x-X_j}{h_n}\hat\alpha_n(X_j)\right)\left(\frac{x-X_j}{h_n}\hat\alpha_n(X_j)\right)^{l+2} \right. \\
& \left. \times \frac{R(X_j)}{(x-X_j)^2}\left(\hat\alpha_n(X_j)\right)^{-(l+1)} \right| \\
+ h_n^{l+2} \sum_{j\notin I} & \left| K\left(\frac{x-X_j}{h_n}\alpha(X_j)\right)\left(\frac{x-X_j}{h_n}\alpha(X_j)\right)^{l+2} \right. \\
& \left. \times \frac{R(X_j)}{(x-X_j)^2}\left(\alpha(X_j)\right)^{-(l+1)} \right| \\
= h_n^{l+2} \sum_{j\notin I} & \left| K\left(\frac{x-X_j}{h_n}\hat\alpha_n(X_j)\right)\left(\frac{x-X_j}{h_n}\hat\alpha_n(X_j)\right)^{l+2} \right. \\
& \left. \times \frac{1}{2} m''(\xi_j)\left(\hat\alpha_n(X_j)\right)^{-(l+1)} \right| \\
+ h_n^{l+2} \sum_{j\notin I} & \left| K\left(\frac{x-X_j}{h_n}\alpha(X_j)\right)\left(\frac{x-X_j}{h_n}\alpha(X_j)\right)^{l+2} \right. \\
& \left. \times \frac{1}{2} m''(\xi_j)\left(\alpha(X_j)\right)^{-(l+1)} \right| \\
\le & \left[d_4 G(a_n^{-1/(l+4)})(\alpha^*)^{-(l+1)} h_n^{l+2} \sum_{j\notin I} m''(\xi_j)\right] O_P(1) \\
= & \, o_P(n h_n^{l+3}),
\end{aligned} \tag{7.25}$$

where we used the fact that

$$P\left\{\min_{1\le j\le n} \hat\alpha_n(X_j) \ge \frac{\alpha^*}{2}\right\} \to 1 \quad \text{as } n \to \infty.$$

The result now follows from (7.22), (7.24) and (7.25). □

LEMMA 6. *Assume that $f_X(\cdot)$, $\alpha(\cdot)$ and $S(\cdot)$ are bounded functions. Let $\hat\alpha_n(\cdot)$ be a consistent estimator of $\alpha(\cdot)$ such that $\sup_z |\hat\alpha_n(z) - \alpha(z)| = o_P(a_n)$, where $a_n \to 0$. Assume that L is a uniformly Lipschitz continuous function such that $|u^l L(u)| \le G(u)$ for all large $|u|$, where $G(u)$ is decreasing as $|u|$ increases and satisfies $G(a_n^{-1/(l+2)}) = o(h_n)$, for a nonnegative integer l.*

Further, suppose that $\min_z \alpha(z) > 0$. *Then,*

$$\sum_{j=1}^{n} \left[\hat{\alpha}_n(X_j) L\left(\frac{x - X_j}{h_n} \hat{\alpha}_n(X_j) \right) - \alpha(X_j) L\left(\frac{x - X_j}{h_n} \alpha(X_j) \right) \right] S(X_j)(x - X_j)^l$$
$$= o_P\left(n h_n^{l+1} \right).$$

The proof uses similar arguments as that of Lemma 5 and is omitted.
We now are in the position to prove Theorem 5.

PROOF OF THEOREM 5. By the definitions of the estimators and a mean-variance decomposition, we get

$$\begin{aligned}
&E\left[\left(\hat{m}(x, \hat{\alpha}_n) - \hat{m}(x, \alpha) \right)^2 \middle| X_1, \dots, X_n \right] \\
&\quad = E\left\{ \left[\sum_{j=1}^{n} \left(\frac{\hat{w}_j}{\sum_{i=1}^{n} \hat{w}_i} - \frac{w_j}{\sum_{i=1}^{n} w_i} \right) Y_j \right]^2 \middle| X_1, \dots, X_n \right\} \\
&\quad = \left[\sum_{j=1}^{n} \left(\frac{\hat{w}_j}{\sum_{i=1}^{n} \hat{w}_i} - \frac{w_j}{\sum_{i=1}^{n} w_i} \right) m(X_j) \right]^2 \qquad (7.26) \\
&\qquad + \sum_{j=1}^{n} \left(\frac{\hat{w}_j}{\sum_{i=1}^{n} \hat{w}_i} - \frac{w_j}{\sum_{i=1}^{n} w_i} \right)^2 \sigma^2(X_j) \\
&\quad \equiv B_n^2(x) + V_n(x).
\end{aligned}$$

We first handle the term $B_n(x)$. Using that $\sum_{j=1}^{n} \hat{w}_j (x - X_j) = 0$ and $\sum_{j=1}^{n} w_j (x - X_j) = 0$, we rewrite $B_n(x)$ as

$$\begin{aligned}
B_n(x) &= \frac{\sum_{j=1}^{n} \hat{w}_j \left(m(X_j) - m(x) \right)}{\sum_{i=1}^{n} \hat{w}_i} - \frac{\sum_{j=1}^{n} w_j \left(m(X_j) - m(x) \right)}{\sum_{i=1}^{n} w_i} \qquad (7.27) \\
&= \frac{\sum_{j=1}^{n} \hat{w}_j R(X_j)}{\sum_{i=1}^{n} \hat{w}_i} - \frac{\sum_{j=1}^{n} w_j R(X_j)}{\sum_{i=1}^{n} w_i}.
\end{aligned}$$

For the numerator of the first term in (7.27), we have

$$\begin{aligned}
\sum_{j=1}^{n} \hat{w}_j R(X_j) &= \hat{s}_{n,2} \sum_{j=1}^{n} \hat{\alpha}_n(X_j) K\left(\frac{x - X_j}{h_n} \hat{\alpha}_n(X_j) \right) R(X_j) \qquad (7.28) \\
&\quad - \hat{s}_{n,1} \sum_{j=1}^{n} \hat{\alpha}_n(X_j) K\left(\frac{x - X_j}{h_n} \hat{\alpha}_n(X_j) \right) R(X_j)(x - X_j).
\end{aligned}$$

Applying Lemmas 5 and 6, we find that

$$\begin{aligned}
&\sum_{j=1}^{n} \hat{w}_j R(X_j) \\
(7.29)\quad &= \left[s_{n,2} + o_P(nh_n^3)\right]\left[\sum_{j=1}^{n} \alpha(X_j) K\left(\frac{x - X_j}{h_n}\alpha(X_j)\right) R(X_j) + o_P(nh_n^3)\right] \\
&\quad - \left[s_{n,1} + o_P(nh_n^2)\right]\left[\sum_{j=1}^{n} \alpha(X_j) K\left(\frac{x - X_j}{h_n}\alpha(X_j)\right)\right. \\
&\qquad \left. \times R(X_j)(x - X_j) + o_P(nh_n^4)\right].
\end{aligned}$$

Hence, Lemmas 5 and 6 allow us to replace $\hat{\alpha}_n(\cdot)$ by $\alpha(\cdot)$ in expression (7.28). Further, using Lemmas 3 and 4 we can simplify (7.29) to

$$(7.30)\quad \begin{aligned}\sum_{j=1}^{n} \hat{w}_j R(X_j) = \frac{n^2}{2}\alpha^2(x)\left(\frac{h_n}{\alpha(x)}\right)^6 m''(x) f_X^2(x) \\ \times\left[s_2^2 - s_1 s_3\right](1 + o_P(1)).\end{aligned}$$

For the denominator of the first term in (7.27), we use similar arguments and obtain

$$(7.31)\quad \sum_{i=1}^{n} \hat{w}_i = n^2 \alpha^2(x)(h_n/\alpha(x))^4 f_X^2(x)\left[s_2 s_0 - s_1^2\right](1 + o_P(1)).$$

Combination of (7.30) and (7.31) leads to

$$(7.32)\quad \frac{\sum_{j=1}^{n} \hat{w}_j R(X_j)}{\sum_{i=1}^{n} \hat{w}_i} = \frac{1}{2} m''(x)\left(\frac{h_n}{\alpha(x)}\right)^2 \left[\frac{s_2^2 - s_1 s_3}{s_2 s_0 - s_1^2}\right](1 + o_P(1)).$$

From the proof of Theorem 1 (which refers to that of Theorem 4), it can be seen that

$$(7.33)\quad \frac{\sum_{j=1}^{n} w_j R(X_j)}{\sum_{i=1}^{n} w_i} = \frac{1}{2} m''(x)\left(\frac{h_n}{\alpha(x)}\right)^2 \left[\frac{s_2^2 - s_1 s_3}{s_2 s_0 - s_1^2}\right](1 + o_P(1)).$$

Expressions (7.27), (7.32) and (7.33) assure that

$$(7.34)\quad B_n(x) = o_P(h_n^2).$$

We now deal with the variance term in (7.26). First use the boundedness of $\sigma^2(\cdot)$ to obtain

$$\begin{aligned}
V_n(x) &\le \sup_z \sigma^2(z) \sum_{j=1}^{n}\left(\frac{\hat{w}_j}{\sum_{i=1}^{n}\hat{w}_i} - \frac{w_j}{\sum_{i=1}^{n} w_i}\right)^2 \\
(7.35)\quad &= \sup_z \sigma^2(z)\left[\frac{\sum_{j=1}^{n}\hat{w}_j^2}{\left(\sum_{i=1}^{n}\hat{w}_i\right)^2} - 2\frac{\sum_{j=1}^{n}\hat{w}_j w_j}{\sum_{i=1}^{n}\hat{w}_i \sum_{i=1}^{n} w_i} + \frac{\sum_{j=1}^{n} w_j^2}{\left(\sum_{i=1}^{n} w_i\right)^2}\right].
\end{aligned}$$

Note that it suffices to evaluate the numerators of the first and the second term in (7.35). Indeed, all the other factors appearing in that expression have been discussed previously. Again relying on Lemmas 3–6 and the proof of Theorem 1, we find

$$
\begin{aligned}
\sum_{j=1}^{n} \hat{w}_j^2 &= \hat{s}_{n,2}^2 \sum_{j=1}^{n} \hat{\alpha}_n^2(X_j) K^2\left(\frac{x-X_j}{h_n}\hat{\alpha}_n(X_j)\right) \\
&\quad - 2\hat{s}_{n,1}\hat{s}_{n,2} \sum_{j=1}^{n} \hat{\alpha}_n^2(X_j) K^2\left(\frac{x-X_j}{h_n}\hat{\alpha}_n(X_j)\right)(x-X_j) \\
&\quad + \hat{s}_{n,1}^2 \sum_{j=1}^{n} \hat{\alpha}_n^2(X_j) K^2\left(\frac{x-X_j}{h_n}\hat{\alpha}_n(X_j)\right)(x-X_j)^2 \\
&= n^3\alpha^4(x)\left(\frac{h_n}{\alpha(x)}\right)^7 f_X^3(x)\int_{-\infty}^{+\infty}[s_2 - us_1]^2 K^2(u)\,du(1+o_P(1))
\end{aligned} \tag{7.36}
$$

and

$$
\begin{aligned}
\sum_{j=1}^{n} \hat{w}_j w_j &= \hat{s}_{n,2} s_{n,2} \sum_{j=1}^{n} \hat{\alpha}_n(X_j) K\left(\frac{x-X_j}{h_n}\hat{\alpha}_n(X_j)\right)\alpha(X_j) K\left(\frac{x-X_j}{h_n}\alpha(X_j)\right) \\
&\quad - \left[s_{n,2}\hat{s}_{n,1} + \hat{s}_{n,2}s_{n,1}\right] \sum_{j=1}^{n} \hat{\alpha}_n(X_j) K\left(\frac{x-X_j}{h_n}\hat{\alpha}_n(X_j)\right)\alpha(X_j) \\
&\qquad \times K\left(\frac{x-X_j}{h_n}\alpha(X_j)\right)(x-X_j) \\
&\quad + \hat{s}_{n,1}s_{n,1} \sum_{j=1}^{n} \hat{\alpha}_n(X_j) K\left(\frac{x-X_j}{h_n}\hat{\alpha}_n(X_j)\right)\alpha(X_j) \\
&\qquad \times K\left(\frac{x-X_j}{h_n}\alpha(X_j)\right)(x-X_j)^2 \\
&= n^3\alpha^4(x)\left(\frac{h_n}{\alpha(x)}\right)^7 f_X^3(x)\int_{-\infty}^{+\infty}[s_2 - us_1]^2 K^2(u)\,du(1+o_P(1)).
\end{aligned} \tag{7.37}
$$

Substituting the expressions we have evaluated so far, including those proved in Theorem 1, into (7.35) we get

$$
V_n(x) = o_P\left(\frac{1}{nh_n}\right). \tag{7.38}
$$

The result now follows from (7.26), (7.34) and (7.38). □

REFERENCES

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

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

CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74** 829–836.

CLEVELAND, W. S. and DEVLIN, S. J. (1988). Locally-weighted regression: an approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.* **83** 597–610.

FAN, J. (1992a). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* To appear.

FAN, J. (1992b). Local linear regression smoothers and their minimax efficiency. *Ann. Statist.* To appear.

GASSER, T. and MÜLLER, H. G. (1979). Kernel estimation of regression functions. *Smoothing Techniques for Curve Estimation. Lectures Notes in Math.* **757** 23–68. Springer, New York.

HALL, P. (1990). On the bias of variable bandwidth curve estimators. *Biometrika* **77** 529–535.

HALL, P. and MARRON, J. S. (1988). Variable window width kernel estimates of probability densities. *Probab. Theory Related Fields* **80** 37–49.

JENNEN-STEINMETZ, C. and GASSER, T. (1988). A unifying approach to nonparametric regression estimation. *J. Amer. Statist. Assoc.* **83** 1084–1089.

JONES, M. C. (1990). Variable kernel density estimates and variable kernel density estimates. *Austral. J. Statist.* **32** 361–371.

LEJEUNE, M. (1985). Estimation non-paramètrique par noyaux: règression polynomiale mobile. *Rev. Statist. Appl.* **33** 43–68.

MÜLLER, H. G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Amer. Statist. Assoc.* **82** 231–238.

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

NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

SILVERMAN, B. W. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.* **12** 898–916.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

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

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599-3260

DEPARTMENT OF MATHEMATICS
LIMBURGS UNIVERSITAIR CENTRUM
B-3590 DIEPENBEEK
BELGIUM