

ADAPTIVE BANDWIDTH CHOICE FOR KERNEL REGRESSION

by

R.L. Eubank and W.R. Schucany

Technical Report No. SMU/DS/TR/241

Department of Statistical Science

July 1990

Adaptive Bandwidth Choice for Kernel Regression

R. L. Eubank and William R. Schucany*

SUMMARY

A data-based procedure is introduced for local bandwidth selection for kernel estimation of a regression function at a point. The estimated bandwidth is shown to be consistent and asymptotically normal as an estimator of the (asymptotic) optimal value for minimum mean square estimation. The rate of convergence is identical to that of plug-in bandwidth estimators. The proposed method has the practical advantage that it reduces the need for *a priori* values and does not require pilot estimates of the regression function, optimization of estimated objective functions or resampling. A small Monte Carlo study is used to examine the behavior of the new bandwidth estimator in a variety of situations. The resulting finite-sample mean square errors of the corresponding curve estimates are generally found to be less than or equal to those of an idealized plug-in estimator.

KEY WORDS: curve estimation; local bandwidth selection; nonparametric regression; smoothing.

* R. L. Eubank is Professor of Statistics at Texas A&M University, College Station, Texas 77843-3143 and William R. Schucany is Professor of Statistical Science at Southern Methodist University, Dallas, Texas 75275-0332. The research of the first author is supported by NSF Grant No. DMS 8902576 and of the second by DARPA Contract F19628-88-K-0042. Computational assistance by Weihang Bao is gratefully acknowledged.

1. INTRODUCTION

Bandwidth selection occupies an important role in the literature of nonparametric regression (cf. Marron, 1989, or Eubank, 1988, for references). With few exceptions, the primary emphasis of this work has been on the selection of globally optimal bandwidths. However, it is known (see, e.g., Müller, 1988, and Staniswalis, 1989) that gains in estimator performance can be realized by optimizing the bandwidth locally rather than on a global basis. Thus, in this paper we present a simple, effective method for selecting local bandwidths in kernel regression.

Consider the nonparametric regression model where responses Y_1, \dots, Y_n are observed following the model

$$Y_i = m(t_i) + \epsilon_i, \quad i = 1, \dots, n. \quad (1.1)$$

Here the ϵ_i are independent, identically distributed, random variables with zero mean and finite variance σ^2 , the t_i satisfy $0 \leq t_1 < \dots < t_n \leq 1$ and m is an unknown function. Without having to assume more about m than certain smoothness conditions, we wish to estimate m at some fixed argument t .

There are many good estimators for $m(t)$. Examples of these can be found in Eubank(1988) and Müller(1988). In particular, the Priestley-Chao kernel estimator of m at t is

$$\hat{m}_h(t) = \frac{1}{h} \sum_{i=1}^n (t_i - t_{i-1}) K\left(\frac{t-t_i}{h}\right) Y_i, \quad (1.2)$$

where K is a kernel function, $h > 0$ is the bandwidth or smoothing parameter and $t_0 = 0$. The kernel K is assumed to be continuously differentiable, symmetric with support on $[-1, 1]$ and of order p in the sense that

$$\int_{-1}^1 z^j K(z) dz = \begin{cases} 1, & j = 0, \\ 0, & j = 1, \dots, p-1, \\ k_p \neq 0, & j = p. \end{cases} \quad (1.3)$$

To use \hat{m}_h in practice one requires choices for both h and K . Discussions of methods for selecting K can be found in Müller(1988). We will concentrate here on the problem of selecting h . The value used for h will be allowed to depend on the point of estimation t . Our goal is to find a good choice of h for each value of t in the sense of making the mean squared error (mse) of estimation as small as possible.

There are several data adaptive local bandwidth selection techniques that have been proposed in the literature. Modifications of squared-error cross validation for consistent estimation of optimal local smoothing have been introduced by Hall and Schucany (1989) and Vieu (1990). An alternative resampling approach that uses the bootstrap to estimate the mse of $\hat{m}_h(t)$ is described by Härdle and Bowman (1988). Two other approaches to estimating the mse that use pilot estimates of $m(t)$ have been studied by Müller (1985) and Staniswalis (1989). All of these algorithms involve a search for a local minimum of an estimated mse and require the specification of some other tuning parameter, e.g., a global bandwidth for a pilot estimate of m . In contrast, the technique that we propose essentially does not require such initial values and there is no search required for the minima of a cross-validation or estimated mse function.

Our approach to local bandwidth selection stems from some simple asymptotic analysis. Let $Var(h) = Var(\hat{m}_h(t))$ and $Bias(h) = E\hat{m}_h(t) - m(t)$. Then, standard Taylor expansions reveal that if $m \in C^p[0, 1]$, the mean squared error $mse[\hat{m}_h(t)] = E(\hat{m}_h(t) - m(t))^2$ can be written as

$$\begin{aligned} mse[\hat{m}_h(t)] &= Var(h) + Bias^2(h) \\ &= \frac{\sigma^2 Q}{nh} + [h^p k_p m^{(p)}(t)/p!]^2 + o\left(\frac{1}{nh}\right) + o(h^{2p}), \end{aligned} \quad (1.4)$$

where $Q = \int_{-1}^1 K^2(z) dz$ and $k_p = \int_{-1}^1 z^p K(z) dz$. Minimization of (1.4) with respect to h yields

$$h_t^* = \left[\frac{\sigma^2 Q}{2pn(k_p m^{(p)}(t)/p!)^2} \right]^{1/(2p+1)}, \quad (1.5)$$

if we ignore higher order terms. By substituting (1.5) into (1.4) we then obtain

$$\text{Var}(h_t^*) = 2p \text{Bias}^2(h_t^*), \quad (1.6)$$

again neglecting higher order terms. More general results for integrated mse and derivative estimation can be found in Gasser, Müller, Köhler, Molinari and Prader (1984) and Müller(1988).

The basic proposal here is to capitalize on the balance between variance and bias present in (1.6). We first estimate both the variance and bias over a grid of fixed h values. For large n we should have for any fixed h that $\text{Var}(h) \sim A/nh$ and $\text{Bias}^2(h) \sim Bh^{2p}$ for constants A and B . Thus, given several estimated values of the variance and bias one can obtain estimates \hat{A} and \hat{B} of A and B (e.g., by least squares) and then solve (1.6) to find the adaptive bandwidth choice

$$\hat{h}_t = \left[\frac{\hat{A}}{2pn\hat{B}} \right]^{1/(2p+1)}. \quad (1.7)$$

In Section 3 we will show that \hat{h}_t is consistent and asymptotically normal as an estimator of h_t^* and attains the same convergence rate as plug-in estimators.

The remainder of the paper is organized as follows. In the next section we provide details concerning the computation of our bandwidth estimator. Then, in Section 3, asymptotic properties of \hat{h}_t are described. The findings of a small simulation experiment are presented in Section 4. Finally, some concluding remarks are collected in Section 5.

2. ADAPTIVE BANDWIDTHS

In this section we give a detailed description of our method for local bandwidth selection. Throughout the remainder of the paper we assume that the design is equally spaced, i.e., $t_i = i/n$. It should be emphasized however that this is merely for simplicity and the approach extends directly to more general designs.

Two essential ingredients of the proposed method are estimators of $Var(h)$ and $Bias(h)$. The exact variance of $\hat{m}_h(t)$ is

$$Var(h) = \sigma^2 \sum_{i=1}^n \frac{1}{n^2 h^2} K^2\left(\frac{t-t_i}{h}\right). \quad (2.1)$$

Thus, to estimate the variance we need only estimate σ^2 . For this purpose we use the estimator proposed by Gasser, Sroka and Jennen-Steinmetz(1986) which has the form $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=2}^{n-1} [Y_{i-1} - 2Y_i + Y_{i+1}]^2/6$ for an equispaced design. Consequently, $Var(h)$ can be easily estimated for any given value of h by replacing σ^2 by $\hat{\sigma}^2$ in (2.1). We denote the result by $\hat{Var}(h)$.

It remains to estimate $Bias(h)$ for which purpose we use

$$\hat{Bias}(h) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K_G\left(\frac{t-t_i}{h}\right) Y_i, \quad (2.2)$$

where $K_G(z) = K(z) - K_{p+2}(z)$ and K_{p+2} is any $(p+2)$ th order kernel. In Section 4 we specialize to $p = 2$ and use the 4th order kernel studied by Schucany(1989): $K_4(z) = [K(z) - c^3 K(cz)]/(1-c^2)$, with $c = .671$ and $K(z) = \frac{3}{4}(1-z^2)$, $|z| \leq 1$.

A heuristic motivation for the use of (2.2) as an estimator of $Bias(h)$ can be derived from the asymptotic form (1.4). For large n and a kernel of order ν we will have $E\hat{m}_h(t) \sim m(t) + h^\nu k_\nu m^{(\nu)}(t)/\nu!$. Taking the difference of such asymptotic expressions for $\nu = p$ and $p+2$ leaves the

lead term $h^p k_p m^{(p)}(t)/p!$ as required. In actuality m need not have $p + 2$ derivatives for \hat{Bias} to be effective but requires only slightly more smoothness than membership in $C^p[0, 1]$ (see Section 3 for more details). Our approach is closely related to twicing for estimation of bias (Stuetzle and Mittal, 1979) and can, in fact, be shown to include twicing as a special case in an asymptotic sense.

To obtain the estimators \hat{A} and \hat{B} in (1.7) we evaluate (2.1) and (2.2) over a grid of predetermined bandwidths, h_1, \dots, h_k . In practice the number in this grid may be reasonably small and we have used $k = 7$ successfully. Thus, at a fixed t one obtains k estimates from (2.1), which we denote by $(v_1, \dots, v_k) = (\hat{Var}(h_1), \dots, \hat{Var}(h_k))$ and, by squaring the values from (2.2), $(b_1^2, \dots, b_k^2) = (\hat{Bias}^2(h_1), \dots, \hat{Bias}^2(h_k))$. These two sets of estimators are then fit to their simple asymptotic expressions as functions of h , namely A/nh and Bh^4 , via ordinary least squares. The resulting estimates of A and B are

$$\hat{A} = n \left(\sum_{j=1}^k v_j / h_j \right) \left(\sum_{j=1}^k h_j^2 \right)^{-1} \quad (2.3)$$

and

$$\hat{B} = \left(\sum_{j=1}^k b_j^2 h_j^{2p} \right) \left(\sum_{j=1}^k h_j^4 \right)^{-1}. \quad (2.4)$$

By substituting (2.3) and (2.4) into (1.7) we obtain our estimator \hat{h}_t of h_t^* .

Figure 1 illustrates the idea. It displays fits to the v_j and b_j^2 for a specific simulated example with $p = 2$, $m(t) = \sin(t)$ at $t = .50$, $\sigma = .015$ and $n = 100$. Actually the \hat{Bias}^2 curve is multiplied by 4 so that the intersection of the two curves occurs at the desired value, \hat{h}_t , given by (1.7). The maximum bandwidth in the grid of values has been chosen to be the largest permissible without encountering boundary bias. This implies $\max_i h_i = .671 \max\{t - t_1, t_n - t\}$. The minimum bandwidth is large enough for a sufficient number of points to be in the window. We have set $\min_i h_i = 6 \min\{t_i -$

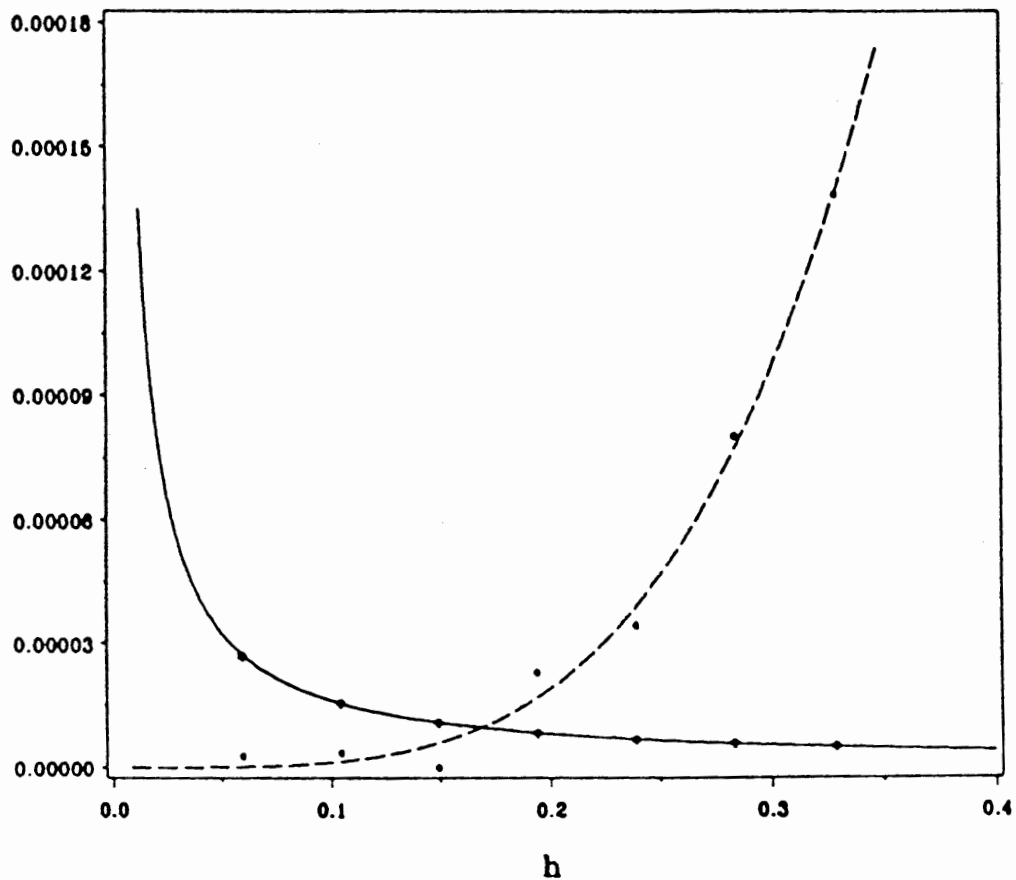


Figure 1.
 Least Squares Fits to Estimated Variance and Bias² Values
 for a Grid of Fixed Bandwidths.

t_{i-1}). This lower endpoint is not critical to the stability of the algorithm; the variance estimates in (2.1) use all of the data regardless of the magnitude of h . The bias estimates do become erratic when too few points get nonzero weights, but the curve Bh^4 is forced through the origin and thus an errant positive value at one small h has no noticeable impact on the fit. Design considerations for efficient estimation of A and B produce two designs that are skewed in opposite directions. To balance these and have stable estimates a reasonable compromise appears to be equally spaced values of h_j . On the other hand, estimating B is more difficult than estimating A . Consequently, a grid of values more concentrated toward the right might prove beneficial, although we have not experimented with this to any great extent.

3. ASYMPTOTIC PROPERTIES

In this section we state and prove our principal asymptotic results. Recall that the data values are equally spaced and the errors are independent with common variance σ^2 . We will require an assumption of Lipschitz continuity for $m^{(p)}$. By this we mean the following: a function f is said to be Lipschitz continuous of order $0 < \gamma < 1$ if there exists a finite constant M such that $\sup_{s \neq t} |f(s) - f(t)| \leq M|s - t|^\gamma$.

Theorem. Assume that $m^{(p)}$ is Lipschitz continuous of order $0 < \gamma < 1$ and the h_i satisfy $h_i \sim C_i n^{-\alpha}$ for $0 < \alpha < 1$ and $0 < C_1 < \dots < C_k < \infty$. Then if $1/(2(p + \gamma) + 1) < \alpha < 1/2$, $n^{(1-(2p+1)\alpha)/2} [\hat{h}_i/h_i^* - 1] \xrightarrow{d} N(0, 4\theta^2/B(2p+1)^2)$, where $\theta^2 = \sigma^2 \sum_{i=1}^k \sum_{j=1}^k (C_i C_j)^{3p-1} \int_{-1}^1 K_G(u/C_i) K_G(u/C_j) du / (\sum_{i=1}^k C_i^{4p})^2$ and B is the coefficient for the dominant term in the squared bias that is estimated by (2.5).

The Theorem states, among other things, that $\hat{h}_i/h_i^* - 1 = O_p(n^{-(1-(2p+1)\alpha)/2})$. Thus, \hat{h}_i/h_i^*

converges to 1 in probability provided that $\alpha < 1/(2p + 1)$. The rate of convergence is quite slow for α close to $1/(2p + 1)$ but can be much faster if γ is large and α is selected to be small. These rates may be inherent to the local bandwidth selection problem (cf. Staniswalis, 1989, and the Corollary below).

It is interesting to compare the rates given by the Theorem with those for global bandwidth estimators. In the case of global bandwidth estimates obtained by cross-validation, it is known that when $p = 2$ the ratio of the bandwidth estimate to the optimal bandwidth will converge to one at the rate $n^{-1/10}$, regardless of the value of γ (cf. Härdle, Hall and Marron, 1988, and Park and Marron, 1990). In contrast, by choosing α appropriately, we can always make \hat{h}_i/h_i^* converge faster than $n^{-1/10}$ if γ is sufficiently close to one.

Of course the Theorem establishes much more than just rates of convergence. The asymptotic normality result may be useful for constructing confidence intervals for h_i^* . Discussion of this point in a related setting can be found in Härdle, Hall and Marron(1988).

The conditions on the range of α restrict the h_i from being either too small or too large. The lower bound is needed to insure that $n^{(1-(2p+1)\alpha)/2}[\hat{B} - B]$ is asymptotically normal. This property of \hat{B} is what drives the proof of the Theorem. An essentially identical condition is needed by Staniswalis(1989) for the bandwidth of a pilot estimator of $m^{(p)}$ that was used to construct her estimator of the optimal local bandwidth.

Proof of Theorem. To establish the Theorem two lemmas are required that we now state and prove.

Lemma 1. $nh\hat{Var}(h) - A = O_P((nh)^{-1}) + O_P(n^{-1/2})$ if $nh \rightarrow \infty$ as $n \rightarrow \infty$.

Proof. The proof follows from standard arguments (cf. Eubank, 1988, Lemma 4.1) and the fact that $\hat{\sigma}^2 - \sigma^2 = O_P(n^{-1/2})$ (cf. Gasser, et al, 1986). □

Lemma 2. Let $B = D^2$, $\hat{D}_{h_i} = \text{Bias}(h_i)/h_i^p$ and set $\hat{D} = (\hat{D}_{h_1}, \dots, \hat{D}_{h_k})'$. Then, if $m^{(p)}$ is Lipschitz continuous of order γ , $h_i \sim C_i n^{-\alpha}$ for $0 < C_1 < \dots < C_k < \infty$ and $\alpha > 1/(2(p + \gamma) + 1)$, $n^{(1-(2p+1)\alpha)/2}[\hat{D} - D1] \xrightarrow{d} N_k(0, \sigma^2 \Sigma)$ with 1 a k -vector of all unit elements and Σ a $k \times k$ matrix having typical element $\sigma_{ij} = \int_{-1}^1 K_G(u/C_i)K_G(u/C_j)du/(C_i C_j)^{p+1}$.

Proof. Using Theorem 4.2 of Müller(1988) and standard arguments one can show that $n^{(1-(2p+1)\alpha)/2} \mathbf{a}'[\hat{D} - E\hat{D}]$ converges in distribution to a zero mean normal random variable with variance $\mathbf{a}'\Sigma\mathbf{a}$ for any vector \mathbf{a} such that this quantity is nonzero. Since $E\hat{D} - D1$ is $O(n^{-\gamma\alpha})$, as a consequence of our assumptions about m , the lemma now follows from our conditions on α and the definition of asymptotic normality (cf. Serfling, 1980, pg. 21). \square

We are now ready to establish the Theorem. Our estimator of A is $\hat{A} = n[\sum_{j=1}^k \text{Var}(h_j)/C_j]/n^\alpha \sum_{j=1}^k C_j^{-2} = A + O_P(n^{\alpha-1/2})$, by Lemma 1 and our assumptions on the h_j . Also, from Lemma 2 and results on transformations of asymptotically normal random variables we have $n^{(1-(2p+1)\alpha)/2}[\hat{B} - B] \xrightarrow{d} N(0, 4B\theta^2)$.

The Theorem can now be established by writing

$$\begin{aligned} n^{(1-(2p+1)\alpha)/2} \{ \hat{h}_i/h_i^* - 1 \} &= n^{(1-(2p+1)\alpha)/2} \{ [\hat{A}B/\hat{B}A]^{1/(2p+1)} - 1 \} \\ &= n^{(1-(2p+1)\alpha)/2} \{ [(1 + O_P(n^{\alpha-1/2}))]/(1 + (\hat{B} - B)/B) \}^{1/(2p+1)} - 1 \} \\ &= - n^{(1-(2p+1)\alpha)/2} (\hat{B} - B)/B(2p + 1) + O_P(n^{\alpha-1/2}) + O_P(n^{-(2p-1)\alpha/2}). \end{aligned}$$

Applying our asymptotic normality result for \hat{B} completes the proof. \square

As a final comment we note that similar techniques to those used in proving the Theorem can be used to establish rates of convergence for a plug-in estimator of h_i^* such as $\hat{h}_{t,PI} =$

$[\hat{\sigma}^2 Q/nk_p \hat{m}_b^{(p)}(t)^2]^{1/(2p+1)}$, where $\hat{m}_b^{(p)}(t)$ is a kernel estimator of $m^{(p)}(t)$ with bandwidth b and kernel K_D . Specifically, we have the following.

Corollary. Assume that $m^{(p)}$ is Lipschitz continuous of order $0 < \gamma < 1$ and that b satisfies $b \sim n^{-\alpha}$ for $1/(2(p + \gamma) + 1) < \alpha < 1/2$. Then, $n^{(1-(2p+1)\alpha)/2} [\hat{h}_{t,PI}/h_t^* - 1] \xrightarrow{d} N(0, 4\sigma^2 \int_{-1}^1 K_D^2(u) du / m^{(p)}(t)^2)$.

4. FINITE SAMPLE PERFORMANCE

To demonstrate the implementation and investigate the stability of the algorithm for \hat{h}_t , several example problems were generated on the IBM 3081D computer at Southern Methodist University. A Fortran program using IMSL subroutines computed observations Y_1, \dots, Y_n for $n = 50, 100, 200, 400$ and 1000 from (1.1) with $m(t) = \sin(t)$. The disturbances were obtained by generating standard normal random deviates that were then rescaled to have standard deviations $\sigma = .005$ and $.05$. At $t = .5$ the true value of interest is $m(.5) = .479$.

The estimator \hat{h}_t was computed for each sample using $k = 7$ and $K_4(z) = [K(z) - c^3 K(cz)] / (1 - c^2)$, with $c = .671$ and $K(z) = \frac{3}{4}(1 - z^2)$, $|z| \leq 1$. Also computed were two "competing" bandwidths: namely, the true asymptotically optimal bandwidth h_t^* from (1.5) and an optimal plug-in type estimator, $\hat{h}_{t,PI}$. The estimator $\hat{h}_{t,PI}$ is obtained by using the estimator $\hat{\sigma}^2$ in place of σ^2 and a kernel estimator for $m''(t)$ in (1.5). Although this estimator is data dependent, the bandwidth, b , that is used for $\hat{m}_b''(t)$ has been set at its asymptotically optimal value. More specifically, $\hat{m}_b''(t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{b^3} K_*\left(\frac{t-t_i}{b}\right) Y_i$, where $K_*(z) = 105(-5z^4 + 6z^2 - 1)/16$, the optimal kernel of order (2, 4) from Gasser, Müller and Mammitzsch (1985), and $b = \left[5\sigma^2 Q^* / 4n(k_4^* m^{(4)}(t)/4!)^2 \right]^{1/9}$ with $Q^* = \int_{-1}^1 K_*^2(z) dz$ and $k_4^* = \int_{-1}^1 z^4 K_*(z) dz$.

Average values of the estimated bandwidths, \hat{h}_t and $\hat{h}_{t,PI}$, along with their Monte Carlo standard deviations were calculated for $M = 1000$ replications. Table 1 summarizes the results. Examination of

Table 1
 Summary of Average Bandwidths over $M = 1000$
 Monte Carlo Repetitions (Standard Deviations in parentheses)

σ	n	Asymptotically Optimal, h_t^*	Adaptive Choice, \hat{h}_t	Ideal Plug-In, \hat{h}_{tPI}	Correlation $r(\hat{h}_t, \hat{h}_{tPI})$
.005	50	.1267	.1258 (.0080)	.1206 (.0095)	.88
	100	.1103	.1104 (.0048)	.1095 (.0062)	.87
	200	.0960	.0962 (.0029)	.0967 (.0037)	.85
	400	.0836	.0839 (.0018)	.0838 (.0022)	.87
	1000	.0696	.0699 (.0009)	.0698 (.0011)	.86
.05	50	.3182	.2914 (.035)	.2719 (.049)	.67
	100	.2770	.2772 (.037)	.2715 (.053)	.67
	200	.2412	.2467 (.032)	.2551 (.054)	.66
	400	.2099	.2125 (.020)	.2220 (.046)	.70
	1000	.1748	.1764 (.010)	.1815 (.028)	.71

the values in the table reveals that the proposed bandwidth estimator is comparable to the ideal plug-in estimator for estimation of h_t^* but tends to be more stable. The greater variability of $\hat{h}_{t,PI}$ is evidently due to estimating $m''(t)$ using only a single bandwidth. Even though b is set at the "ideal" asymptotically optimal value for $\hat{m}_b''(t)$, \hat{h}_t appears to gain stability from the smoothing afforded by the fit across several bandwidths used for \hat{B} and \hat{A} .

The rate of convergence of \hat{h}_t , as measured by the decrease in its standard deviation as n increases, appears to be better than that predicted for cross validation. The asymptotics begin to take effect more slowly when the noise is greater. In other words, at $\sigma = .05$, the standard deviation of \hat{h}_t does not begin its characteristic decline until after $n = 200$. This delay is still more pronounced for the plug-in estimator.

The final column of the table contains the Monte Carlo correlation coefficient between \hat{h}_t and $\hat{h}_{t,PI}$ over the 1000 pairs of estimated bandwidths. The strong correlation should not be surprising since both techniques are trying to estimate the same unknown ingredients of h_t^* in (1.5). The correlation appears to be very weakly dependent upon n and a decreasing function of σ .

It is important to note that the primary interest is not in estimating h_t^* . The main objective is to have some practical method for local bandwidth selection that leads to small finite-sample mse for $\hat{m}_{h_t}(t)$. Table 2 presents the average of $M = 1000$ squared errors for the same example and the three bandwidths covered in Table 1. In the low noise case $\sigma = .005$ the two estimated bandwidths yield about the same results. However, when σ is larger the adaptive bandwidth \hat{h}_t is 5-9% more efficient than the ideal plug-in rule. The adaptive rule is not competitive with the fixed bandwidth, h_t^* , but, of course, such a quantity is not available in practice.

It is also of interest to examine the rate of decay of the sample mse sequence as a function of n . We expect this sequence to decay like $n^{-4/5}$ if h_t^* is being estimated correctly. Figure 2 displays the mse's on a log-log scale for the asymptotically optimal and adaptive bandwidths when $\sigma = .05$. The slope of $-4/5$ is apparent and the relative efficiency of the adaptive procedure is improving with

Table 2
 Comparison of the Effects of Data-Based Bandwidths
 on Mean-Square Errors of Estimation

σ	n	Asymptotically Optimal, h_t^*	Adaptive Choice, \hat{h}_t	Ideal Plug-In, \hat{h}_{tPI}
.005	50	.0247	.0467	.0556
	100	.0344	.0196	.0203
	200	.0097	.0101	.0101
	400	.0052	.0056	.0057
	1000	.0028	.0028	.0029
.05	50	1.2523	1.4233	1.5502
	100	.6675	.7324	.7721
	200	.3716	.4002	.4232
	400	.2150	.2341	.2513
	1000	.1102	.1165	.1222

Entries are $10^4 \times mse [\hat{m}(t; h)]$ with three different rules for h .

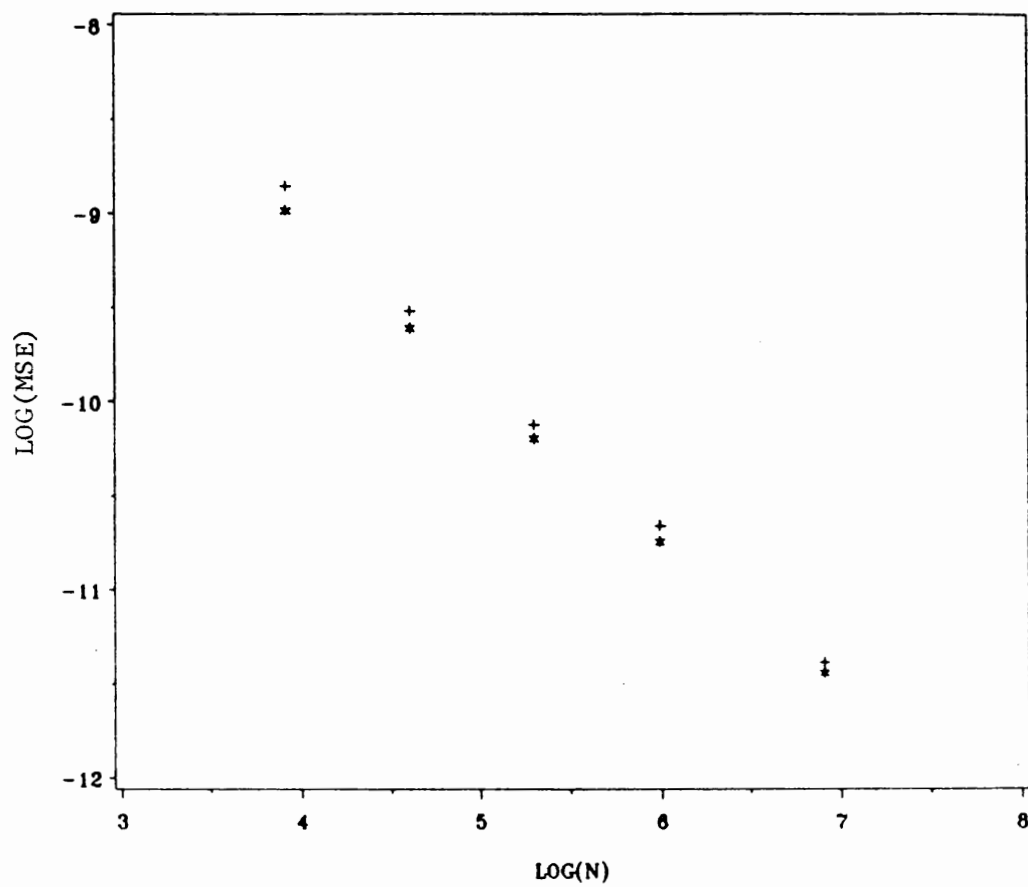


Figure 2.
Mean Square Errors versus Sample Size for Asymptotically
Optimal(*) and Adaptive(+) Bandwidths

increasing sample size.

To further illustrate the utility of our proposed rule, another simulated data has been used to produce a curve estimate in Figure 3. In this example $m(t)$ is estimated at each t using $\hat{m}_{\hat{h}_t}(t)$. The function being estimated is $m(t) = 4.26(e^{-3.25t} - 4e^{-6.5t} + 3e^{-9.75t})$ which has been used in numerous studies, e.g., Staniswalis (1989). To eliminate the complication of the boundary effects realizations of Y_i for t_i in the interval $(-1, 2)$ are used in producing the estimates in $[0, 1]$. The scale on the left is for the Y_i points, the solid line represents $m(t)$ and the dashed line corresponds to the estimator of m based on our local bandwidth estimates. Superimposed on this graph are curves for the local bandwidths. The scale on the right is for values of h_t^* (solid line) and \hat{h}_t (dotted line). The estimated bandwidths perform as one would hope by increasing and decreasing according to the curvature of m . The spikes in the asymptotically optimal bandwidths correspond to values of $m''(t)$ near zero. That the peaks in \hat{h}_t occur in different places is simply an indication that the finite sample bias has a (estimated) minimum other than where the dominant term vanishes.

5. CONCLUSIONS

In this paper we have proposed a new method for local bandwidth selection. This technique has been shown to be practical and perform well in finite samples. The asymptotic properties of the bandwidth estimator have been derived and it was found to be both consistent and asymptotically normal.

An important question that is now under study is how to adapt our bandwidth selection technique for use when t is near the boundary region. More research is also needed on how to best choose the grid of bandwidths used with the algorithm. Experimentation with different grids suggests that it may improve the procedure to use a larger minimum for the h_i . A different bias estimator may also prove useful in this regard. The possibility exists, for example, of estimating B with a generalized

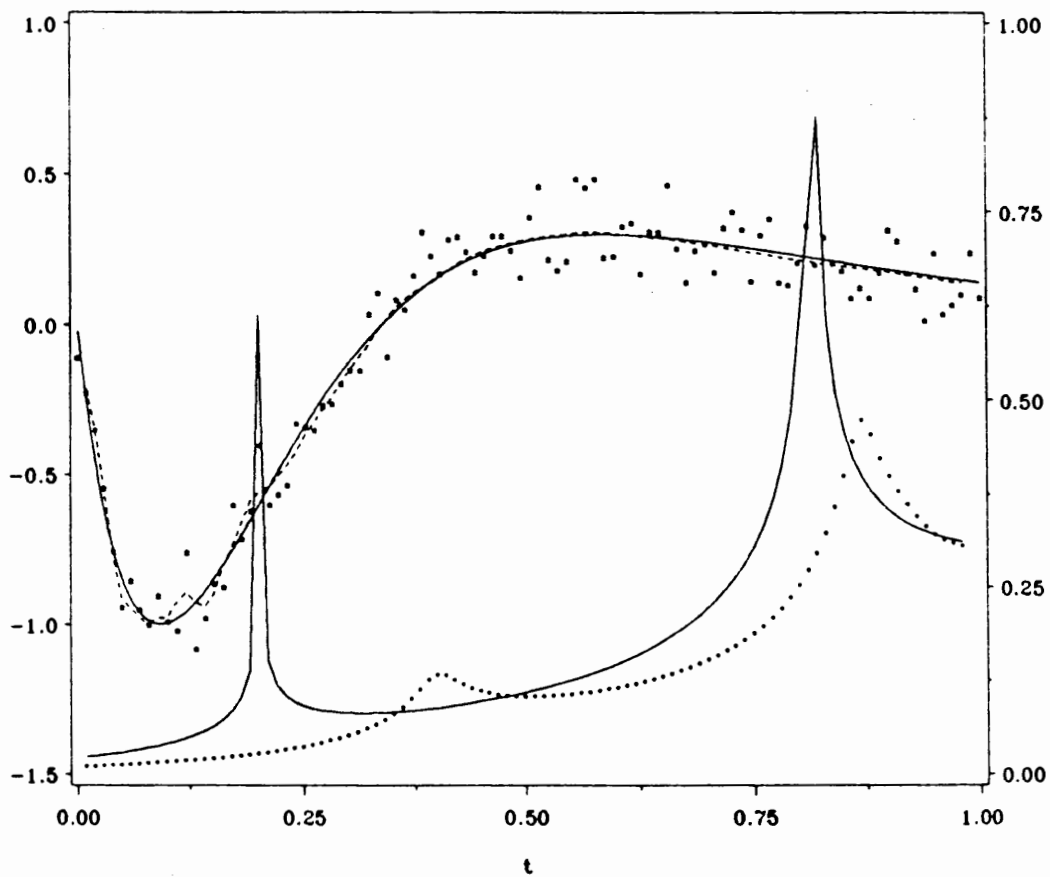


Figure 3.

Illustration of Adaptive Local Bandwidths.

Upper portion displays data(*), $m(t)$ solid and $\hat{m}(t; \hat{h}_t)$ dashed;
 lower two curves are \hat{h}_t^* (solid) and \hat{h}_t (dotted); $n = 100$, $\sigma = .10$

or robust alternative to ordinary least squares. Such an approach has the potential to alleviate effects produced by the grid of bandwidths being either too small (so that the bias estimates are very noisy) or too large (causing extrapolation beyond the range of adequacy of the Taylor approximation). Although these fine tuning issues remain open for study, we believe that the initial version of the algorithm presented here is sufficient to demonstrate its potential superiority over plug-in rules.

To conclude, we note that the basic procedure outlined above can be extended in a number of directions. For example, a similar adaptive approach may be developed for local bandwidth selection in probability density estimation. The variance would, of course, need to be estimated differently in this setting.

REFERENCES

- Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. New York: Marcel-Dekker.
- Gasser, T., Müller, H. G., Köhler, W., Molinari, L. and Prader, A. (1984). "Nonparametric Regression Analysis of Growth Curves," *The Annals of Statistics*, 12, 210-229.
- Gasser, T., Müller, H. G. and Mammitzsch, V. (1985), "Kernels for Nonparametric Curve Estimation," *Journal of the Royal Statistical Society*, B47, 238-252.
- Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986), "Residual Variance and Residual Pattern in Nonlinear Regression," *Biometrika*, 73, 123-127.
- Hall, P. and Schucany, W. R. (1989), "A Local Cross-Validation Algorithm," *Statistics and Probability Letters*, 8, 109-117.
- Härdle, W. and Bowman, A. W. (1988), "Bootstrapping in Nonparametric Regression: Local Adaptive Smoothing and Confidence Bands," *Journal of the American Statistical Association*, 83, 102-110.
- Härdle, W., Hall, P. and Marron, J. S. (1988), "How Far Are Automatically Chosen Regression Smoothers From Their Optimum," *Journal of the American Statistical Association*, 83, 86-101.
- Marron, J. S. (1989), "Automatic Smoothing Parameter Selection: A Survey," *Empirical Economics*, to appear.
- Müller, H. G. (1985), "Empirical Bandwidth Choice for Nonparametric Kernel Regression by Means of Pilot Estimators," *Statistics and Decisions*, 2, 193-206.
- Müller, H. G. (1988), *Nonparametric Regression Analysis of Longitudinal Data*, Heidelberg: Springer-Verlag.
- Müller, H. G. and Stadtmüller, U. (1987), "Variable Bandwidth Kernel Estimators for Regression Curves," *The Annals of Statistics*, 15, 182-201.
- Park, B. U. and Marron, J. S. (1990), "Comparison of Data-Driven Bandwidth Selectors," *Journal of the American Statistical Association*, 85, 66-72.
- Schucany, W. R. (1989) "On Nonparametric Regression with Higher-Order Kernels," *Journal of Statistical Planning and Inference*, 23, 145-151.
- Staniswalis, J. G. (1989), "Local Bandwidth Selection for Kernel Estimates," *Journal of the American Statistical Association*, 84, 284-288.
- Stuetzle, W. and Mittal, Y. (1979), "Some Comments on the Asymptotic Behavior of Robust Smoothers," *Smoothing Techniques for Curve Estimation*, eds. T. Gasser and M. Rosenblatt, Heidelberg: Springer-Verlag, pp. 191-195.
- Vieu, P. (1990), "Nonparametric Regression: Optimal Local Bandwidth Choice," to appear, *Journal of the Royal Statistical Society*, Series B.