The Annals of Statistics
1997, Vol. 25, No. 3, 1001–1013

A BANDWIDTH SELECTOR FOR LOCAL LINEAR DENSITY ESTIMATORS[1]

BY MING-YEN CHENG

National Chung Cheng University

Local linear density estimators achieve automatic boundary corrections and enjoy some typical optimal properties. Proper choice of the smoothing parameters is crucial for their performance. A data-based bandwidth selector is developed in the spirit of plug-in rules. Consistency and asymptotic normality of the selected bandwidth are demonstrated. The bandwidth is very efficient regardless of whether there are non-smooth boundaries in the support of the density or not.

1. Introduction. Local polynomial fitting is becoming widely accepted as an appealing tool for nonparametric curve estimation. Estimators generated by such methods provide many benefits. They are highly intuitive and easy to implement [see Fan and Marron (1994)]. Theoretically, they achieve automatic boundary corrections and enjoy some important optimal properties [see Fan (1993), Fan, et al (1996) and Cheng, Fan and Marron (1997)]. Various density estimators which implement the local polynomial fitting techniques are introduced in Lejeune and Sarda (1992), Jones (1993), Cheng (1994) and Wei and Chu (1994). The goal of this article is to find a simple and effective data-based bandwidth procedure for such estimators. Fan and Gijbels (1992) and Ruppert, Sheather and Wand (1995) discussed bandwidth selection for local linear regression.

Sheather (1992) and Jones, Marron and Sheather (1994), among others, recognized the bandwidth of Sheather and Jones (1991) as a useful and effective tool in both practice and theory. We attempt to mimic its ideas to construct a bandwidth for local linear density estimation. Estimation of integrated squared density derivatives is crucial for plug-in bandwidth procedures. Conventional estimators of these quantities become less efficient when there are nonsmooth boundaries in the support of the density [see Van Es and Hoogstrate (1994)]. As a consequence, traditional plug-in bandwidths are not adequate in that case [see Van Es and Hoogstrate (1993)]. The estimators of integrated density derivative products introduced in Cheng (1997) automatically adjust for boundary effects. So we incorporate them in our bandwidth rule such that it will be proper in both boundary and nonboundary cases.

Weak convergence and asymptotic normality of the resulting bandwidth are established. Interestingly, the rate of convergence depends on the sign of the integrated product of the second and fourth derivatives of the density. The

Received March 1994; revised May 1996.

[1]Supported in part by NSC Grant 85-2121-M-194-001.

AMS 1991 *subject classification*. 62G07.

Key words and phrases. Bandwidth, boundary effects, data binning, local linear density estimation, plug-in bandwidth selector.

rate of convergence is $n^{-5/14}$, the same as that of the Sheather–Jones selector in nonboundary cases if the quantity is negative, and $n^{-2/7}$ otherwise. Note that if the density has essentially no boundary features, the above-mentioned functional is negative. For example, if the second and third derivatives both vanish at the boundaries, then integration by parts yields that the quantity is minus the integrated squared third derivative of the density. However, the bandwidth selector is always consistent for the optimal bandwidth no matter whether there is a nonsmooth boundary or not. Ruppert, Sheather and Wand (1995) developed bandwidth selectors for local linear regression based on the plug-in ideas and established analogous theory on their asymptotic behavior.

Motivation for our bandwidth procedure is entirely analogous to that for the conventional plug-in rules for kernel density estimation [e.g., those discussed in Park and Marron (1990) and Sheather and Jones (1991)]. Yet, contributions of this work include showing that the same approach is applicable for local linear estimation, which has many appealing properties both theoretically and practically, and providing a bandwidth selector that is as effective in the boundary case as in the nonboundary case.

This article is organized as follows. In Section 2, construction and asymptotic performance of local linear density estimators are briefly discussed. Section 3 gives the suggested bandwidth and establishes its consistency and asymptotic normality properties. Section 4 contains some numerical justification of the bandwidth. Proofs are given in Section 5.

2. Local linear estimators. Suppose that $X_1, \dots, X_n$ is an i.i.d. sample observed from a population following an univariate density f. Let $x_i = T + (i - 1/2)b$, $i = 1, \dots, g$, where b is a positive constant which depends on the sample size, and T, g and b are taken such that there is essentially no data point falling outside the interval $[T, T + gb]$. Define the bin counts at the bin centers $\{x_1, \dots, x_g\}$ as

$$c_i = \sum_{j=1}^{n} I_{[x_i - b/2,\, x_i + b/2)}(X_j), \qquad i = 1, \dots, g.$$

Each c_i provides information about $f(x_i)$ in the sense that

$$n^{-1}b^{-1}c_i \to b^{-1}\int_{x_i - b/2}^{x_i + b/2} f(u)\,du \approx f(x_i) \quad \text{a.s.}$$

One can also construct the bin counts by linear binning [see Fan and Marron (1994)] and all the asymptotic results stated in this article will still hold. Then apply the local linear regression ideas to fit

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{g} \big(n^{-1}b^{-1}c_i - \beta_0 - \beta_1(x_i - x)\big)^2 K\left(\frac{x_i - x}{h}\right), \tag{1}$$

where K is a nonnegative weight function and $h > 0$ is the smoothing parameter. Denote the solution of the least squares problem (1) as $\widehat{b}_j(x)$, $j = 0, 1$.

Then the local linear fit estimator $\widehat{f}_h(x)$ of $f(x)$ is defined as $\widehat{b}_0(x)$. Explicitly,

$$\widehat{f}_h(x) \equiv \widehat{b}_0(x) = \frac{S_{n,2}(x)T_{n,0}(x) - S_{n,1}(x)T_{n,1}(x)}{S_{n,2}(x)S_{n,0}(x) - S_{n,1}(x)S_{n,1}(x)},$$

where

$$S_{n,j}(x) = \sum_{i=1}^{g} K\left(\frac{x_i - x}{h}\right)(x_i - x)^j, \qquad j = 0, 1, 2,$$

and

$$T_{n,j}(x) = \sum_{i=1}^{g} K\left(\frac{x_i - x}{h}\right)(x_i - x)^j n^{-1} b^{-1} c_i, \qquad j = 0, 1.$$

It is assumed that f is supported on $[0, \infty)$ whenever boundary effects are discussed. Then typical boundary points are conventionally parametrized as $x = ch$, $c \geq 0$. The following theorem is given in Cheng (1994). Throughout this article, we denote $\mu_l(\psi) = \int u^l \psi(u)\,du$, $l = 0, 1, \dots,$ and $R(\psi) = \int \psi^2$ for any real-valued function ψ on R.

THEOREM 1. *Suppose f and its first two derivatives are bounded, $K^{(l)}$ is bounded and integrable with finite second moments, $l = 0, 1, 2$, and $\mu_0(K) = 1$, $\mu_1(K) = 0$. As $n \to \infty$, $h \to 0$, $nh \to \infty$ and $b/h \to 0$, when x is an interior point,*

$$E\big(\widehat{f}_h(x) - f(x)\big)^2 = \frac{h^4}{4}(f^{(2)}(x))^2 \mu_2(K) + \frac{1}{nh} f(x) R(K) + o\left(h^4 + \frac{1}{nh}\right), \tag{2}$$

and when x is a boundary point, $x = ch, c \geq 0$,

$$E\big(\widehat{f}_h(x) - f(x)\big)^2 = \frac{h^4}{4}(f^{(2)}(0+))^2 \mu_2(K^*_{0,c}) + \frac{f(0+)}{nh} R(K^*_{0,c}) + o\left(h^4 + \frac{1}{nh}\right), \tag{3}$$

where

$$K^*_{0,c}(u) = \frac{S_{2,c} - u S_{1,c}}{S_{2,c}S_{0,c} - S_{1,c}S_{1,c}} K(u) I_{[-c,\infty)}(u), \tag{4}$$

with $S_{j,c} = \int_{-c}^{+\infty} t^j K(t)\,dt$, $j = 0, 1, 2$.

Theorem 1 is analogous to the results for local polynomial regression smoothers given in Ruppert and Wand (1994). Equations (2) and (3) show that $\widehat{f}_h$ behaves like the conventional kernel density estimator in the interior and achieves automatic boundary corrections. Performance of local linear estimators depends on the weight function K and the bandwidth h. Choosing the bandwidth is much more important than deciding the weight function since h controls the amount of the smoothing. We discuss data-based bandwidth procedures in the following sections.

3. A bandwidth selector. For any nonnegative integers γ and ν, define

$$\theta_{\gamma,\nu} = \int f^{(\gamma)}(x) f^{(\nu)}(x)\,dx.$$

For kernel estimation of f based on the kernel K, the asymptotically optimal bandwidth which minimizes the asymptotic mean integrated squared error is

$$h_* = \left\{ \frac{R(K)}{n\mu_2(K)^2\theta_{2,2}} \right\}^{1/5}. \tag{5}$$

Plug-in bandwidth rules use the above expression with the unknown quantity $\theta_{2,2}$ replaced by its estimate. It is shown in Cheng (1994) that the asymptotically optimal bandwidth for local linear estimation is also equal to (5). Hence the plug-in techniques can be implemented in bandwidth selection for local linear density estimation.

Kernel-type estimators of $\theta_{2,2}$ are suggested in Hall and Marron (1987), Jones and Sheather (1991) and others. Those estimators have an efficient rate of convergence in mean squared error provided that the support of f has either no boundaries or no nonsmooth boundaries. Otherwise, the estimators become inefficient and hence the plug-in bandwidths are less efficient. Problems of this kind are investigated by Van Es and Hoogstrate (1993, 1994). We make use of the estimators of $\theta_{\gamma,\nu}$ introduced in Cheng (1997) in our plug-in bandwidth procedure since those estimators have nice boundary adaptive properties. The estimator of $\theta_{2,2}$ is

$$\widehat{\theta}_{2,2}(a) = b\sum_{i=1}^{g} \big(\widehat{f}_a^{(2)}(x_i)\big)^2,$$

where $\widehat{f}_a^{(2)}(x)$ is the estimator of $f^{(2)}(x)$ obtained from a local cubic fitting based on the weight function K and bandwidth $a > 0$. Define $S_n = (S_{n,i+j-2}(x))_{1\le i,j\le 4}$ and $W_2^n(t) = e_3^{\mathrm{T}} S_n^{-1}(1, at, a^2t^2, a^3t^3)^T K(t)$, where

$$S_{n,j}(x) = \sum_{k=1}^{g} K((x_k - x)/a)(x_k - x)^j, \qquad j = 0, 1, \dots, 6,$$

and $e_3^{\mathrm{T}} = (0, 0, 1, 0)$. Then

$$\widehat{\theta}_{2,2}(a) = \frac{4}{n^2 b} \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} W_2^n\left(\frac{x_j - x_i}{a}\right) W_2^n\left(\frac{x_k - x_i}{a}\right) c_j c_k$$

[see Cheng (1996)]. Write

$$K_2^*(t) = e_3^{\mathrm{T}} S^{-1}(1, t, t^2, t^3)^T K(t),$$

where $S = (S_{i+j-2})_{1\le i,j\le 4}$ with $S_j = \int u^j K$, $j = 0, 1, \dots, 6$.

According to Corollary 1 of Cheng (1997), suppose that $R(K_2^*)\mu_4(K_2^*) > 0$, the asymptotically optimal bandwidth for $\widehat{\theta}_{2,2}$ is

$$a_* = \left(\frac{24\chi R(K_2^*)}{n\theta_{2,4}\mu_4(K_2^*)}\right)^{1/7}, \tag{6}$$

where χ equals -1 if $\theta_{2,4} < 0$ and $5/2$ if $\theta_{2,4} > 0$. From (5) and (6),

$$a_* = C(K)D(f)h_*^{5/7}, \tag{7}$$

where

$$C(K) = \left(\frac{24R(K_2^*)\mu_2(K)^2}{R(K)\mu_4(K_2^*)}\right)^{1/7}, \qquad D(f) = \left(\frac{\chi\theta_{2,2}}{\theta_{2,4}}\right)^{1/7}.$$

Analogously to the development of the Sheather–Jones procedure, apply (7) to (5) and find the solution in h of the following equation:

$$h = \left\{\frac{R(K)}{n\mu_2(K)^2\widehat{\theta}_{2,2}(a(h))}\right\}^{1/5}, \tag{8}$$

where

$$a(h) = C(K)D(f)h^{5/7}. \tag{9}$$

Here $D(f)$ involves the unknown quantities $\theta_{2,2}$ and $\theta_{2,4}$. The following describes two possible approaches to estimating $D(f)$ and the corresponding plug-in bandwidths.

If there is information suggesting that the density is close to some scale parametric model with the scale unknown, then $D(f)$ can be estimated by a reference value through the model. Let g_1 be a fixed density function (e.g., the standard normal) that has been normalized so that some measure of scale such as the standard deviation is equal to 1. It is easy to show that $D(g_\lambda) = \lambda^{2/7}D(g_1)$ for any $\lambda > 0$, where $g_\lambda(x) = \lambda^{-1}g_1(x/\lambda)$. Then, from (9), set

$$a_\lambda(h) = C(K)D(g_1)\lambda^{2/7}h^{5/7}.$$

The bandwidth $\widehat{h}_1$ is defined as the solution to the analogous equation of (8) with $a(h)$ replaced by $a_{\widehat{\lambda}}(h)$, where $\widehat{\lambda}$ is a $\sqrt{n}$-consistent estimate of λ. The conventional Sheather–Jones procedure uses parametric reference values in a latter stage. The next bandwidth is its analogy.

First, construct an estimator of $D(f)$ using the estimators of integrated density derivative products given in Cheng (1997). Take the degree of local polynomial fitting as 3 and 5 when estimating $\theta_{2,2}$ and $\theta_{2,4}$, respectively. Let κ_1 and κ_2 be the corresponding asymptotically optimal bandwidths, expressions for which are in Corollary 1 of Cheng (1997), with the quantities depending on f replaced by some scale parametric reference values. The resulting estimators are denoted as $\widehat{\theta}_{2,2}(\kappa_1)$ and $\widehat{\theta}_{2,4}(\kappa_2)$. Let $\widehat{D}(f) = (\widetilde{\chi}\widehat{\theta}_{2,2}(\kappa_1)/\widehat{\theta}_{2,4}(\kappa_2))$, where $\widetilde{\chi}$ equals -1 if $\widehat{\theta}_{2,4}(\kappa_2) < 0$ and $5/2$ if $\widehat{\theta}_{2,4}(\kappa_2) > 0$. Then our second bandwidth $\widehat{h}_2$ is the solution to the analogous equation of (8) with $a(h)$ replaced by $a_{\widehat{D}(f)}(h) = C(K)\widehat{D}(f)h^{5/7}$.

CONDITION 1. (i) There exists a constant $M > 0$ so that, for any x and y in the support of f,

$$|f^{(6)}(x) - f^{(6)}(y)| \le M|x - y|.$$

(ii) The weight function K is supported on $[-1, 1]$ and its first two derivatives exist.

(iii) The weight function K vanishes at -1 and 1 and is symmetric about zero.

Let $\sigma_K^2 = \int u^2 K(u)\,du$. The following theorem establishes that consistency and asymptotic normality properties of $\widehat{h}_2$ hold no matter whether there are boundaries in the support of f or not.

THEOREM 2. *Suppose that Condition* 1 *holds. Then as* $n \to \infty$,

$$\frac{\widehat{h}_2}{h_*} = 1 + O_p(n^{-\alpha}),$$

where α *equals* $5/14$ *if* $\theta_{2,4} < 0$ *and* $2/7$ *if* $\theta_{2,4} > 0$. *Furthermore*,

$$n^{\alpha}\left(\frac{\widehat{h}_2}{h_*} - 1\right) \to_D N(\mu_{PI}, \sigma_{PI}^2), \tag{10}$$

where

$$\mu_{PI} = \tfrac{-7}{150}\sigma_K^{-8/7} R(K)^{2/7} \theta_{2,2}{}^{-9/7} C(K)^2 D(f)^2 \theta_{2,4} \mu_4(K_2^*) I_{(\theta_{2,4}>0)},$$

$$\sigma_{PI}^2 = \tfrac{32}{25} n^{2\alpha-5/7} \sigma_K^{118/35} R(K_2^* * K_2^*) R(K)^{-38/35} \theta_{2,2}^{-5/7} R(f) C(K)^{-9} D(f)^{-9}.$$

REMARK 1. The rate of convergence of the bandwidth $\widehat{h}_2$ depends on the sign of $\theta_{2,4}$ [see (10)]. If there is no boundary or the density is smooth near the boundaries (i.e., second and third derivative of f at zero both vanish), then $\theta_{2,4} = -\theta_{3,3} < 0$ and hence the faster rate $n^{-5/14}$, the same as that of Sheather and Jones (1991), is attained. Otherwise, the Sheather–Jones bandwidth will not be consistent [see Van Es and Hoogstrate (1993)], but $\widehat{h}_2$ still is, only with a slower rate of convergence.

REMARK 2. If f is a member of the parametric family $\{g_\lambda,\ \lambda > 0\}$, then analogous asymptotic properties for the bandwidth $\widehat{h}_1$ can be established similarly.

Since the asymptotically optimal bandwidth (5) is exactly the same as that for the conventional kernel density estimator, one might ask why not simply use the Sheather and Jones (1991) bandwidth procedure. However, the estimators of Jones and Sheather (1991) are influenced by boundary effects and the bandwidth selector will not even be consistent as a consequence. Van Es and Hoogstrate (1993) showed that the bandwidth selector tends to zero at some rate other than $n^{-1/5}$, the rate of the asymptotically optimal bandwidth, in the nonsmooth boundary case.

Variance of local polynomial density derivative estimators can be seriously inflated near the boundary. This can make the bandwidths $\widehat{h}_1$ and $\widehat{h}_2$ very unstable in practice. The following adjustments, which will not change the asymptotic results in Theorem 2, are made to address this problem. Let $\tau \geq 0$ and $\delta = 0$ or $\delta > 1$ be some constants. In the construction of $\widehat{\theta}_{\gamma,\nu}(a)$, estimate the density derivatives using bandwidth δa instead of a at the boundary area $[0, \tau a]$ when $\tau > 0$ and $\delta > 1$, estimate the density derivative only in the interior $(\tau a, \infty)$ when $\delta = 0$ and make no adjustment when $\tau = 0$.

4. A numerical study. The proposed bandwidth selectors were applied to a real data set to examine their practical performance. The data set was obtained from a study on the public water sources in Malawi. The purpose of the study was to determine whether there was any difference in coliform counts between the "improved" and "unimproved" water resources. Hence there are two sets of data, coliform counts from the improved and unimproved water sources, whose sample sizes are 166 and 455, respectively. Adjusted versions of the data sets, with sizes 142 and 305, respectively, were analyzed in the study.

The kernel function was taken as the standard normal density. The parametric scale family for reference values was the family of normal densities with standard deviation as the scale parameter. For each of the data sets, both $\widehat{h}_1$ and $\widehat{h}_2$ were computed under the following three adjusting settings: (i) $\tau = 0$; (ii) $\tau = 1$, $\delta = 0$; and (iii) $\tau = 1$, $\delta = 3$. The Sheather–Jones bandwidth, which is denoted by $\widehat{h}_{SJ}$, was also computed. Table 1 displays the selected bandwidths. The bandwidths selected by the procedure $\widehat{h}_1$ are given in parentheses. Figure 1 depicts local linear density estimates based on the bandwidths $\widehat{h}_{SJ}$ and the three variants of $\widehat{h}_2$. Figure 2 is a zoom-in version of Figure 1.

The unadjusted $\widehat{h}_1$ and $\widehat{h}_2$, i.e. with $\tau = 0$, are too small for smoothing the data sets. This is explained as follows. Local polynomial density derivative estimates could have arbitrarily large absolute values near the boundary. Hence $\widehat{\theta}_{2,2}$ is too large and the plug-in bandwidth is too small. Under the adjustment

TABLE 1

Bandwidths $\widehat{h}_{SJ}$, $\widehat{h}_1$ and $\widehat{h}_2$ for the coliform counts data. Bandwidths for both of the "improved" and "unimproved" data and each of the settings (i) $\tau = 0$, (ii) $\tau = 1$, $\delta = 0$ *and* (iii) $\tau = 1$, $\delta = 3$; *here, $\widehat{h}_1$ is the one parenthesized*

Bandwidth	**Improved data**	**Unimproved data**
$\widehat{h}_{SJ}$	7.7088	10.6608
$\tau = 0$	2.0610 (2.9913)	2.0974 (2.3967)
$\tau = 1,\ \delta = 0$	5.7212 (5.1393)	10.8774 (10.6632)
$\tau = 1,\ \delta = 3$	4.7816 (4.7858)	9.9346 (9.9946)

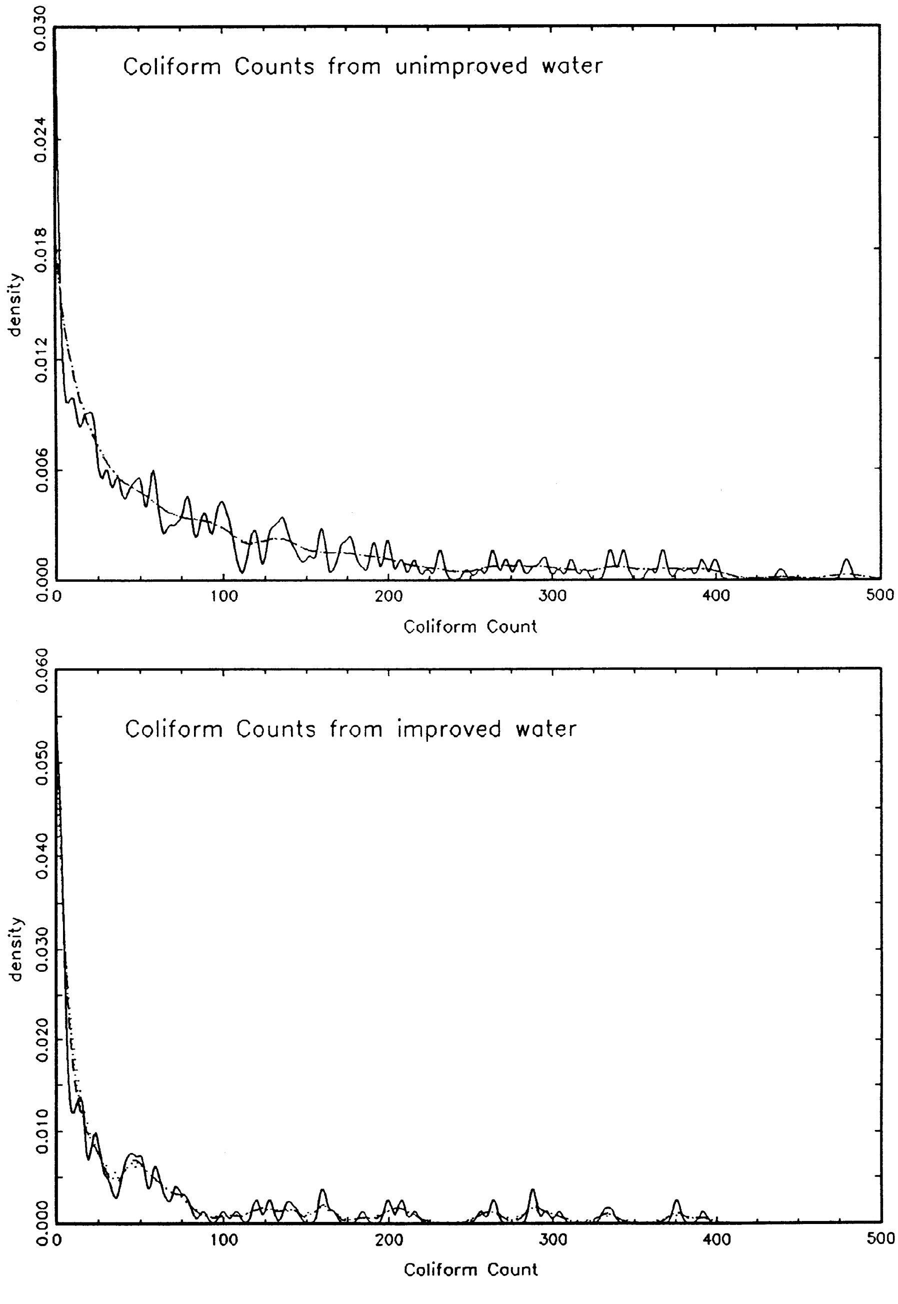

FIG. 1. *Local linear density estimates based on the Malawi data and data-driven bandwidths: kernel function was the standard normal density. Normal densities with the standard deviation as the scale were used to compute the reference values. For each of the "unimproved" (upper panel) and "improved" (lower panel) data, the local linear estimates were constructed using the Sheather–Jones bandwidth (dotted line),* $\widehat{h}_2$ *with* $\tau = 0$ *(solid line),* $\widehat{h}_2$ *with* $\tau = 1, \delta = 0$ *(closely-spaced dotted line) and* $\widehat{h}_2$ *with* $\tau = 1, \delta = 3$ *(short-dashed line).*

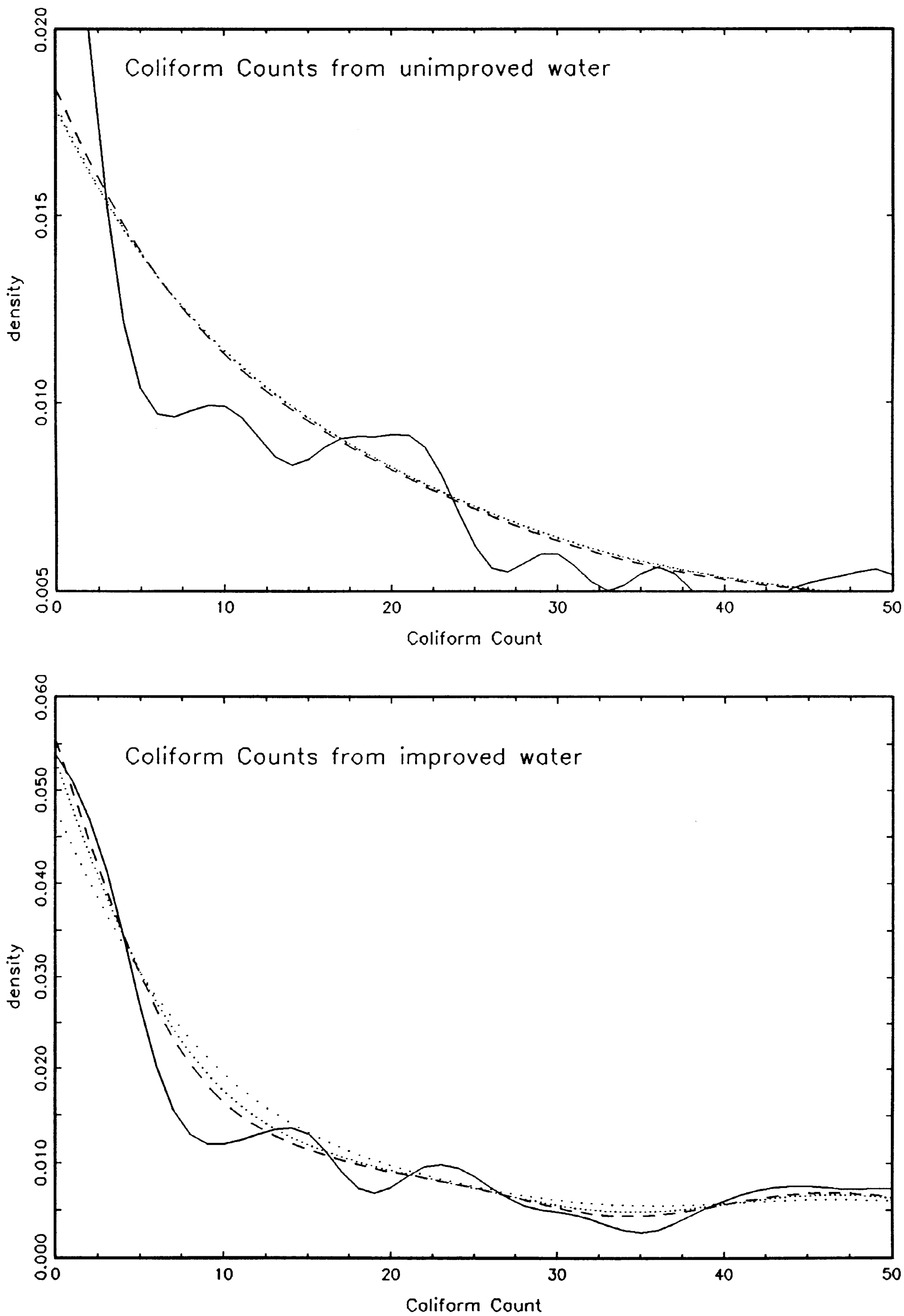

FIG. 2. *The upper (lower) panel is a zoom-in version of the upper (lower) panel of Figure* 1.

$\tau = 1$ and $\delta = 0$, $\widehat{h}_1$ and $\widehat{h}_2$ yield sensible estimates of the density. This shows that disregarding the boundary region in the estimation does stabilize behavior of the bandwidths. However, for the unimproved data, the peak at zero is underestimated; that is, $\widehat{h}_2$ is slightly too large. The reason is that, with the boundary area $[0, a]$ ignored, $\widehat{\theta}_{2,2}$ is adjusted downwards and hence $\widehat{h}_1$ and $\widehat{h}_2$ are adjusted upwards. The above problem does not happen to the bandwidths when $\tau = 1$ and $\delta = 3$. Since enlarging the smoothing parameter near the boundary reduces variation of the density derivative estimators while allowing data points near the boundary involved in the procedure.

Examine the adjusted bandwidths $\widehat{h}_1$ and $\widehat{h}_2$ with $\tau = 1$, $\delta = 3$ and the conventional $\widehat{h}_{SJ}$ more closely. For the unimproved data, $\widehat{h}_{SJ}$, $\widehat{h}_1$ and $\widehat{h}_2$ are comparable to each other. For the improved data, $\widehat{h}_1$ and $\widehat{h}_2$ are smaller than $\widehat{h}_{SJ}$. Notice that $\widehat{h}_1$ and $\widehat{h}_2$ pick up the steep peak near the boundary better than $\widehat{h}_{SJ}$ does. For populations following a density whose derivatives change dramatically like this, one might pursue variable bandwidth procedures. We conclude that our proposed bandwidths are useful as a general tool for exploring data structure, even when the underlying density is not smooth near the boundary.

Finally, we discuss the choice between $\widehat{h}_1$ and $\widehat{h}_2$ (i.e., use reference values directly or do one more step of estimation). Theoretically, $\widehat{h}_2$ is superior to $\widehat{h}_1$ since Theorem 2 guarantees that $\widehat{h}_2$ has an efficient rate of convergence in very general settings. From practical considerations, $\widehat{h}_2$ has the following disadvantages. Computation of $\widehat{\theta}_{2,4}$ is very involved and slow. The estimator $\widehat{\theta}_{2,4}$ could inherit a large variability from the data and hence can make the bandwidth more variable. Furthermore, as observed from the numerical results, $\widehat{h}_2$ is more sensitive to the choice of the adjusting parameters τ and δ than $\widehat{h}_1$. For small to moderate sample sizes, $\widehat{h}_1$ is recommended as a plausible and stable procedure.

5. Proofs.

PROOF OF THEOREM 2. Let $(K^*_{2,c})'(t) = (d/dt)K^*_{2,c}(t)$ and define

$$A_n(a) = \frac{-24b}{n^2a^7}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} K^*_{2,x_i/a}\left(\frac{x_j - x_i}{a}\right)K^*_{2,x_i/a}\left(\frac{x_k - x_i}{a}\right)c_jc_k,$$

$$B_n(a) = \frac{-8b}{n^2a^7}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} \frac{x_j - x_i}{a}(K^*_{2,x_i/a})'\left(\frac{x_j - x_i}{a}\right)K^*_{2,x_i/a}\left(\frac{x_k - x_i}{a}\right)c_jc_k.$$

Proofs of the following lemmas can be found in Cheng (1994).

LEMMA 1. *As $n \to \infty$ and $b/a \to 0$,*

$$\widehat{\theta}_{2,2}(a) = -\frac{a}{6}A_n(a)(1 + o_p(1)),$$

$$\frac{d}{da}\widehat{\theta}_{2,2}(a) = (A_n(a) + B_n(a))(1 + o_p(1)).$$

LEMMA 2. *As* $n \to \infty$, $a \to 0$ *and* $na^5 \to \infty$,

$$E(B_n(a)) = \frac{6}{a}\theta_{2,2} + O\left(\frac{1}{na^6}\right) + O(a),$$

$$\operatorname{Var}(B_n(a)) = \frac{32}{n^2a^{11}} R(f)(R(G * K_2^*))^2(1 + o(1)),$$

where $G(x) = x(K_2^*)'(x)$.

LEMMA 3. *As* $n \to \infty$, $a \to 0$, $na^5 \to \infty$ *and* $b = o(a)$,

$$na^5\left(\widehat{\theta}_{2,2}(a) - \theta_{2,2}\right) \to_D N\left(\mu_*, \sigma_*^2\right),$$

where

$$\mu_* = 4R(K_2^*) + 4a \int u(K_2^*)^2 \int f^{(1)} + \frac{na^7}{6}\theta_{2,4}\mu_4(K_2^*),$$

$$\sigma_*^2 = 32aR(f)R(K_2^* * K_2^*).$$

For any $P \in R$, define functions a_P and L_P as

$$a_P(h) = C(K)Ph^{5/7},$$

$$L_P(h) = h\left[\sigma_K^4 \widehat{\theta}_{2,2}(a_P(h))\right]^{1/5} - n^{-1/5}R(K)^{1/5}.$$

From Lemmas 1, 2 and 3,

$$B_n(a) = \frac{6}{a}\theta_{2,2} + O_p\left(\frac{1}{na^6}\right),$$

$$A_n(a) = -\frac{6}{a}\widehat{\theta}_{2,2}(a)(1 + o_p(1)) = -\frac{6}{a}\theta_{2,2} + O_p\left(\frac{1}{na^6}\right).$$

Then, for some a^* between $a_{\widehat{D}(f)}(h)$ and $a_{D(f)}(h)$,

$$\begin{aligned} &\widehat{\theta}_{2,2}(a_{\widehat{D}(f)}(h)) - \widehat{\theta}_{2,2}(a_{D(f)}(h)) \\ &\quad = \frac{d}{da}\widehat{\theta}_{2,2}(a)\Big|_{a=a^*}\left(a_{\widehat{D}(f)}(h) - a_{D(f)}(h)\right) \\ &\quad = [A_n(a^*) + B_n(a^*)]\left[C(K)\left(\widehat{D}(f) - D(f)\right)h^{5/7}\right](1 + o_p(1)) \\ &\quad = C(K)\left(\widehat{D}(f) - D(f)\right)h^{5/7}O_p\left(\frac{1}{na^{*6}}\right). \end{aligned} \tag{11}$$

From (11), Lemma 3, $\widehat{h}_2 \sim n^{-1/5}$ and $L_{\widehat{D}(f)}(\widehat{h}_2) = 0$,

$$\begin{aligned} L_{D(f)}(\widehat{h}_2) &= L_{D(f)}(\widehat{h}_2) - L_{\widehat{D}(f)}(\widehat{h}_2) \\ &= -\sigma_K^4\widehat{h}_2\left\{\widehat{\theta}_{2,2}(a_{\widehat{D}(f)}(\widehat{h}_2))^{1/5} - \widehat{\theta}_{2,2}(a_{D(f)}(\widehat{h}_2))^{1/5}\right\} \\ &= \left(\widehat{D}(f) - D(f)\right)O_p(n^{-17/35}). \end{aligned} \tag{12}$$

Also,

$$\frac{d}{dh}L_{D(f)}(h) = \left[\sigma_K^4 \widehat{\theta}_{2,2}(a_{D(f)}(h))\right]^{1/5}$$

$$(13) \qquad + \frac{h^{5/7}}{7}\sigma_K^{4/5} C(K)D(f)\left[\widehat{\theta}_{2,2}(a_{D(f)}(h))\right]^{-4/5}\frac{d}{da}\widehat{\theta}_{2,2}(a)\Big|_{a_{D(f)}(h)}$$

$$= \left[\sigma_K^4 \theta_{2,2}\right]^{1/5} + o_p(1).$$

Next, by Lemma 3 and the δ-method,

$$(14) \qquad n^{\alpha_1} L_\lambda(h_*) \to_D N(\mu_1, \sigma_1^2),$$

where α_1 equals 39/70 if $\theta_{2,4} < 0$ and 17/35 if $\theta_{2,4} > 0$,

$$\mu_1 = \begin{cases} O(n^{-1/14}), & \text{if } \theta_{2,4} < 0, \\ \frac{7}{150}\sigma_K^{-8/7} R(K)^{17/35} \theta_{2,2}^{-9/7} C(K)^2 D(f)^2 \theta_{2,4}\mu_4(K_2^*), & \text{if } \theta_{2,4} > 0, \end{cases}$$

and

$$\sigma_1^2 = \tfrac{32}{25} n^{2\alpha_1 - 39/35} \sigma_K^{118/35} R(K_2^* * K_2^*) R(K)^{-31/35} \theta_{2,4}^{-5/7} R(f) C(K)^{-9} D(f)^{-9}.$$

Now,

$$(15) \qquad L_{D(f)}(\widehat{h}_2) = L_{D(f)}(h_*) + \frac{d}{dh}L_{D(f)}(h^{**})(\widehat{h}_2 - h_*),$$

where h^{**} lies in between $\widehat{h}_2$ and h_*. Theorem 2 of Cheng (1997) implies that $\widehat{D}(f) - D(f) = o_p(n^{-1/14})$ if κ_1 and κ_2 are of order $n^{-1/7}$ and $n^{-1/9}$, respectively, which is true by the construction of κ_1 and κ_2. Combining the above fact and (12)–(15), we obtain, as $n \to \infty$,

$$n^\alpha\left(\frac{\widehat{h}_2 - h_*}{h_*}\right) = n^\alpha\left(\frac{L_{D(f)}(\widehat{h}_2) - L_{D(f)}(h_*)}{h_*(d/dh)L_{D(f)}(h^{**})}\right)$$

$$= n^\alpha\left(\frac{n}{R(K)}\right)^{1/5} L_{D(f)}(h_*)(-1 + o_p(1)) \to_D N(\mu_{PI}, \sigma_{PI}^2).$$

Acknowledgments. This research is part of the author's dissertation under the supervision of Professor J. S. Marron at the University of North Carolina.

REFERENCES

CHENG, M. Y. (1994). On boundary effects of smooth curve estimators. Mimeo Series #2319, Institute of Statistics, Univ. North Carolina.

CHENG, M. Y. (1997). Boundary-aware estimators of integrated squared density derivatives. *J. Roy. Statist. Soc. Ser. B* **59** 191–203.

CHENG, M. Y., FAN, J. and MARRON, J. S. (1997). On automatic boundary corrections. *Ann. Statist.* To appear.

FAN, J. (1993). Local linear regression smoothers and their minimax efficiency. *Ann. Statist.* **21** 196–216.

FAN, J. and GIJBELS, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist.* **20** 2008–2036.

FAN, J. and MARRON, J. S. (1994). Fast implementation of nonparametric curve estimators. *J. Comput. Graph. Statist.* **3** 35–56.

FAN, J., GASSER, T., GIJBELS, I., BROCKMANN, M. and ENGELS, J. (1996). Local polynomial fitting: a standard for nonparametric regression. *Ann. Inst. Statist. Math.* To appear.

HALL, P. (1984). Central limit theorem for integrated squared error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14** 1–16.

HALL, P. and MARRON, J. S. (1987). Estimation of integrated squared density derivatives. *Statist. Probab. Lett.* **6** 109–115.

JONES, M. C. (1993). Simple boundary correction for kernel density estimation. *Statist. Comput.* **3** 135–146.

JONES, M. C., MARRON, J. S. and SHEATHER, S. J. (1994). Progress in data-based bandwidth selection for kernel density estimation. Mimeo Series #2088, Institute of Statistics, Univ. North Carolina.

JONES, M. C. and SHEATHER, S. J. (1991). Using non-stochastic terms to advantage in estimating integrated squared density derivatives. *Statist. Probab. Lett.* **11** 511–514.

LEJEUNE, M. and SARDA, P. (1992). Smooth estimators of distribution and density functions. *Comput. Statist. Data Anal.* **14** 457–471.

PARK, B. U. and MARRON, J. S. (1990). Comparison of data-driven bandwidth selectors. *J. Amer. Statist. Assoc.* **85** 66–72.

RUPPERT, D., SHEATHER, S. J. and WAND, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90** 1257–1270.

RUPPERT, D. and WAND, M. P. (1994). Multivariate weighted least squares regression. *Ann. Statist.* **22** 1346–1370.

SHEATHER, S. J. (1992). The performance of six popular bandwidth selection methods on some real data sets. *Comput. Statist.* **7** 225–250.

SHEATHER, S. J. and JONES, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **53** 683–690.

VAN ES, A. J. and HOOGSTRATE, A. J. (1993). Properties of plug-in bandwidth selection procedures in non-smooth cases. Unpublished manuscript.

VAN ES, A. J. and HOOGSTRATE, A. J. (1994). Kernel estimators of integrated squared density derivatives in non-smooth cases. In *Asymptotic Statistics: Proceedings of the Fifth Prague Symposium* (P. Mandl and M. Hušková, eds.) 313–224. Physica, Heidelberg.

WEI, C. Z. and CHU, C. K. (1994). A regression point of view toward density estimation. *J. Nonparametr. Statist.* **4** 191–201.

INSTITUTE OF MATHEMATICAL STATISTICS
NATIONAL CHUNG CHENG UNIVERSITY
MINHSIUNG, CHIAYI 621
TAIWAN
E-MAIL: mycheng@math.nccu.tw