The Annals of Statistics
1991, Vol. 19, No. 3, 1528–1546

SOME STABILIZED BANDWIDTH SELECTORS FOR NONPARAMETRIC REGRESSION[1]

BY SHEAN-TSONG CHIU

Colorado State University and Academia Sinica

The problem of bandwidth selection for nonparametric kernel regression is considered. It is well recognized that the classical bandwidth selectors are subject to large sample variation. Due to the large variation, these selectors might not be very useful in practice. Based on the frequency-domain representation of the residual sum of squares (RSS), the source of the variation is pointed out. The observation leads to consideration of a procedure which stabilizes the RSS by modifying the periodogram of the observations. The stabilized bandwidth selectors are obtained by substituting the stabilized RSS for the RSS in the classical selectors. The strong consistency of the stabilized bandwidth estimate is established. For sufficiently smooth regression functions, it is shown that the stabilized bandwidth is asymptotically normal, and the relative convergence rate of the stabilized bandwidth estimate is $T^{-1/2}$ instead of the rate $T^{-1/10}$ of the classical estimates. In a simulation study, it is confirmed that the stabilized selectors perform much better than the classical selectors. The simulation results are consistent with the theoretic results. The article contains the important message that the commonly used cross-validation can be improved substantially. The procedure and the theoretic results are developed for a rather restrictive case. Further studies are required for more general situations.

1. Introduction. In nonparametric regression estimation, a critical and inevitable step is to choose the smoothing parameter (bandwidth) to control the smoothness of the resulting curve estimate. The smoothing parameter considerably affects the features of the estimated curve. Although in practice one can try several bandwidths and choose a bandwidth subjectively, automatic (data-driven) selection procedures could be useful for many situations; see Silverman (1985) for some examples.

Several automatic bandwidth selectors have been proposed and studied in Craven and Wahba (1979), Silverman (1984), Rice (1984), Li (1985, 1987), Härdle, Hall and Marron (1988), Chiu (1990b) and references given therein. It is well recognized that these bandwidth estimates are subject to large sample variation. As demonstrated in Figure 1, the kernel estimates based on the bandwidths selected by these procedures could have very different appearances. The data sets in Figure 1(a) and (b) are two samples from the simulation study described in Section 5. The solid curves are the kernel estimates

Received February 1989; revised May 1990.

[1]Part of this work was done while the author was visiting the Institute of Statistical Sciences, Academia Sinica, Taiwan.

AMS 1980 *subject classifications*. Primary 62G99; secondary 62M10, 62E20.

Key words and phrases. Nonparametric regression, kernel estimate, bandwidth selection, time series, Fourier transform, periodogram.

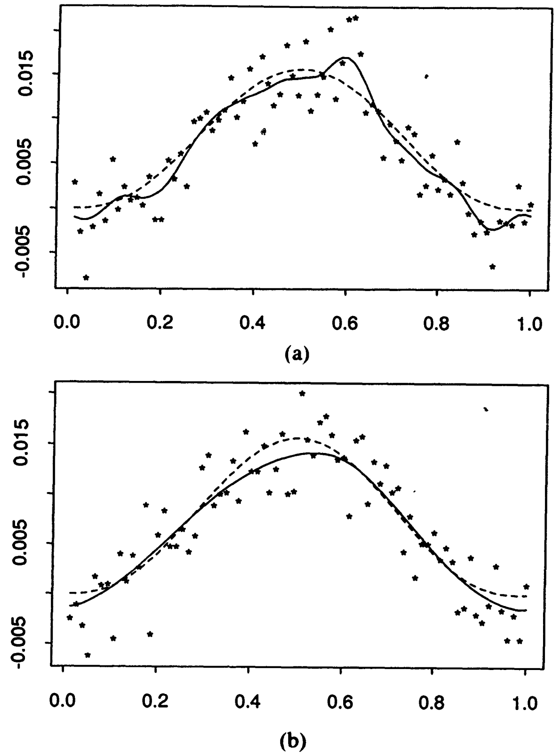

FIG. 1. *Two samples from the simulation study. The dashed curves are the regression function and the solid curves are the kernel estimates based on the bandwidths* (a) 0.157 *and* (b) 0.357 *selected by* GCV.

with the bandwidths chosen by the generalized cross-validation [Craven and Wahba (1979)]. More details concerning Figure 1 can be found in Section 5. Due to the large sample variation, these bandwidth selectors might not be very useful in practice.

In the simulation study of Chiu (1990b), it was observed that Mallows' criterion gives smaller bandwidths more frequently than predicted by the asymptotic theorems. Chiu (1990a) provided an explanation for the cause and suggested a procedure to overcome the difficulty. By applying the procedure, we introduce several bandwidth selectors which give much more stable bandwidth estimates. Section 2 gives a brief background of automatic bandwidth selection. In Section 3, based on the frequency-domain representation of the residual sum of squares (RSS), we point out the source of the variation. The observation leads to consideration of a procedure which stabilizes the RSS by modifying the periodogram of the observations. The stabilized bandwidth selectors are obtained by substituting the stabilized RSS for the RSS in the classical selectors. Some asymptotic results about the stabilized Mallows criterion are given in Section 4. It is shown that the stabilized selector gives a strongly consistent bandwidth estimate. For sufficiently smooth regression functions, we get the quite remarkable result that the relative convergence rate of the stabilized bandwidth estimate is $T^{-1/2}$ instead of the rate $T^{-1/10}$ of the classical estimates. In Section 5 several commonly used criteria are considered. A similar procedure is proposed to modify the well known cross-validation. To compare the bandwidth estimates for finite samples, we carried out a Monte Carlo study in Section 5. The simulation results are consistent with the asymptotic results and confirm that the performance of the stabilized selectors is much better than the performance of the classical selectors. This article contains the important message that the principle of cross-validation can be improved substantially. We remark that the stabilized procedure and the asymptotic results are developed for a rather limited case: positive kernel estimation with a circular design and equally spaced design points. Further research is required for more general situations.

2. Automatic bandwidth selection. We consider the model $Y(t) = m(x_t) + \varepsilon(t)$, $x_t = t/T$, $t = 0, \dots, T-1$, where $\varepsilon(t)$ is a sequence of independent random variables with mean zero and variance σ^2. A commonly used nonparametric method for estimating $m(x)$ is the kernel estimator

$$\hat{m}_\beta(x) = (T\beta)^{-1} \sum_{t=0}^{T-1} w\{(x_t - x)/\beta\} Y(t)$$

[Priestley and Chao (1972)], where β is the bandwidth and $w(x)$ is the kernel function, which is a symmetric probability density function with support on $[-1/2, 1/2]$. Here we consider a "circular design"; that is, $m(x)$ is assumed to be a smooth periodic function and the estimate is obtained by applying the kernel on the extended series $\hat{Y}(t)$, where $\hat{Y}(t + kT) = Y(t)$, $k = 0, \pm 1, \dots$. The circular design is often assumed in theoretic works to avoid difficulties caused by boundary effects. A brief discussion concerning boundary effects is given in Section 6.

The optimal bandwidth considered here is β_{0T}, the minimizer of the risk function

$$R_T(\beta) = E\{\mathrm{SSE}_T(\beta)\} = E \sum_{t=0}^{T-1} \{m(x_t) - \hat{m}_\beta(x_t)\}^2. \tag{2.1}$$

Under some mild conditions, $A_T(\theta) = T^{-1/5} R_T(T^{-1/5}\theta)$ converges to

$$A(\theta) = \theta^{-1}\sigma^2 \int w^2(x)\,dx + 4^{-1}\theta^4 \left\{ \int x^2 w(x)\,dx \right\}^2 \int \{m''(x)\}^2\,dx.$$

$A(\theta)$ has an unique minimum at θ_0, where

$$\theta_0^5 = \sigma^2 \int w^2(x)\,dx \Big/ \left[\left\{ \int x^2 w(x)\,dx \right\}^2 \int \{m''(x)\}^2\,dx \right]. \tag{2.2}$$

Most bandwidth selectors are based on the residual sum of squares

$$\mathrm{RSS}_T(\beta) = \sum_{t=0}^{T-1} \{Y(t) - \hat{m}_\beta(x_t)\}^2.$$

For example, the Mallows criterion [Mallows (1973)],

$$\hat{R}_T(\beta) = \mathrm{RSS}_T(\beta) - T\sigma^2 + 2\sigma^2 w(0)/\beta, \tag{2.3}$$

was considered in Rice (1984). $\hat{R}_T(\beta)$ is an asymptotically unbiased estimate of $R_T(\beta)$. In practice σ^2 in (2.3) is replaced by a $\sqrt{T}$ consistent estimate $\hat{\sigma}^2$. Rice (1984) argued that the error caused by the substitution is negligible. It was shown in Rice (1984) that $\hat{\theta} - \theta_0 = O_p(T^{-1/10})$, where $\hat{\theta} = T^{1/5}\hat{\beta}$ and $\hat{\beta}$ is a local minimizer of $\hat{R}_T(\beta)$. Similar results for other selectors can be found in Rice (1984) and Härdle, Hall and Marron (1988).

3. The motivation and the stabilizing procedure. Following the approach in Chiu (1990b) and Rice (1984), we use the technique of Fourier analysis to study the problem. Most notation and terminology follows Brillinger (1981). In the following discussion, we let $S(t) = m(t/T)$ (signal), and $\hat{S}_\beta(t) = \hat{m}_\beta(t/T)$. The periodogram of the series $Y(t)$, $t = 0, \dots, T-1$, is defined as $I_Y(\lambda) = |d_Y(\lambda)|^2/(2\pi T)$, where

$$d_Y(\lambda) = \sum_{j=0}^{T-1} Y(t)\exp(-i\lambda t), \qquad -\infty < \lambda < \infty,$$

is the (finite) Fourier transform of the series $Y(t)$. The periodograms and Fourier transforms of the series $\varepsilon(t)$, $S(t)$ and $\hat{S}_\beta(t)$ are defined similarly. Applying Parseval's formula yields, for odd T,

$$\mathrm{RSS}_T(\beta) = 4\pi \sum_{j=1}^{N} I_Y(\lambda_j)\{1 - W_\beta(\lambda_j)\}^2 \tag{3.1}$$

[see Chiu (1990b) for details], where $\lambda_j = 2\pi j/T$, $j = 0, \pm 1, \dots,$ are the Fourier frequencies,

$$W_\beta(\lambda) = \sum_t \exp(-i\lambda t) w\{t/(T\beta)\}/(T\beta)$$

is the transfer function of the filter $\{(T\beta)^{-1}w[t/(T\beta)]\}$ and $N = [(T-1)/2]$. Since $w(x)$ is symmetric, $W_\beta(\lambda)$ is a real function. For even T, (3.1) drops the term at frequency π, which has negligible effects. Similarly, for odd T,

$$R_T(\beta) = 4\pi \sum_{j=1}^{N} I_S(\lambda_j)\{1 - W_\beta(\lambda_j)\}^2 + \sigma^2 \sum_{j=1}^{N} W_\beta(\lambda_j)^2 + \sigma^2. \tag{3.2}$$

Under mild conditions, any finite collection of the periodogram ordinates $I_\varepsilon(\lambda_j)$ at λ_j, $j = 1, \dots, N$, is approximately independently and exponentially distributed with mean $\sigma^2/(2\pi)$. See Brillinger (1981) for a more precise statement.

It was shown in Chiu (1990b) that $\hat{\theta} - \theta_0$ is approximately proportional to

$$(3.3) \qquad \sum_{j=1}^{N} \{I_\varepsilon(\lambda_j) - \sigma^2/(2\pi)\}\{1 - W_{\beta_{0T}}(\lambda_j)\}(\partial/\partial\beta) W_{\beta_{0T}}(\lambda_j).$$

A closer look at (3.3) reveals that the periodogram ordinates at $j = O(T^{1/5})$ make the major contribution to the sample variation on the bandwidth estimate. However, under commonly used smoothness conditions of $m(x)$, $m(x)$ has negligible effects on $I_Y(\lambda_j)$ at $j > O(T^{1/6})$. It seems quite unreasonable that the behavior of the bandwidth estimate is essentially determined by the periodogram ordinates $I_Y(\lambda_j)$, which do not contain much information on $m(x)$. This suggests that one could reduce the variation by modifying the periodogram $I_Y(\lambda_j)$ at $j > O(T^{1/6})$.

A procedure for stabilizing the variation is described below. We first obtain an estimate of σ^2. Two estimates, $\hat{\sigma}_1^2 = \pi N^{-1} \sum_{j=1}^{N} I_{\tilde{Y}}(\lambda_j)$ and

$$\hat{\sigma}_2^2 = 2\pi \frac{1}{N - j_0 + 1} \sum_{j=j_0}^{N} \frac{I_{\tilde{Y}}(\lambda_j)}{|1 - \exp(-i\lambda_j)|^2},$$

are considered here, where $\tilde{Y}(t) = Y(t) - Y(t-1)$ is the differenced series of $Y(t)$. The estimate $\hat{\sigma}_1^2$ is a slight modification of the estimate suggested in Rice (1984). Both $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are special cases of

$$(3.4) \qquad \tilde{\sigma}^2 = 2\pi \frac{\sum_{j=1}^{N} I_{\tilde{Y}}(\lambda_j) |1 - \exp(-i\lambda_j)|^2 \phi(\lambda_j)}{\sum_{j=1}^{N} |1 - \exp(-i\lambda_j)|^4 \phi(\lambda_j)},$$

where $\phi(\lambda)$ is a weighting function. More details about the estimation procedure (3.4) can be found in Chiu (1989b). Since the signal may still have some effects on $I_{\tilde{Y}}(\lambda_j)$ at low frequency, we prefer to exclude a few periodogram ordinates when obtaining $\hat{\sigma}_2^2$. By inspecting the plot of $I_{\tilde{Y}}(\lambda_j)/|1 - \exp(-i\lambda_j)|^2$, it is usually not difficult to select a j_0 such that $I_{\tilde{Y}}(\lambda_j)$ at $j \geq j_0$ is not affected much by the signal. With a proper selection of j_0, $\hat{\sigma}_2^2$ is more accurate than $\hat{\sigma}_1^2$. The estimate $\hat{\sigma}_1^2$ has an advantage in that it does not require the selection of j_0. Since a slight change in bandwidth does not affect the kernel estimate much, the estimated curves based on different variance estimates will have similar appearance. In the following discussion, we let $\hat{\sigma}^2$ denote an estimate of σ^2, which could be either $\hat{\sigma}_1^2$ or $\hat{\sigma}_2^2$.

Second, for some constant $c > 1$, find the first integer J such that $I_Y(\lambda_J) < c\hat{\sigma}^2/(2\pi)$. The constant c sets a threshold. As indicated in Theorem 3, the choice of c is not important when $m(x)$ is sufficiently smooth. In our experience, setting $-\log(0.1) < c < -\log(0.05)$ yields good results. For a smooth $m(x)$, the Fourier transform of the series $S(t)$ decays rapidly. Therefore, once $I_Y(\lambda_J)$ is smaller than the threshold, it is reasonable to believe that $I_Y(\lambda)$ at frequency $\lambda \geq \lambda_J$ is not affected much by $m(x)$. The stabilized residual sum of squares (SRSS) is defined by

$$\text{SRSS}_T(\beta) = 4\pi \sum_{j=1}^{N} \tilde{I}_Y(\lambda_j)\{1 - W_\beta(\lambda_j)\}^2, \tag{3.5}$$

where

$$\tilde{I}_Y(\lambda_j) = \begin{cases} I_Y(\lambda_j), & j < J, \\ \hat{\sigma}^2/(2\pi), & j \ge J. \end{cases} \tag{3.6}$$

The stabilized selectors are obtained by substituting the stabilized RSS and RSS in the classical selectors. As an example, we modify (2.3) and get a stabilized criterion

$$\tilde{R}_T(\beta) = \text{SRSS}_T(\beta) - T\hat{\sigma}^2 + 2\hat{\sigma}w(0)/\beta. \tag{3.7}$$

An example of the stabilized criterion is plotted in Figure 3.

4. Assumptions and asymptotic results. We first describe the assumptions which we need in establishing the results.

ASSUMPTION 1. The noise $\varepsilon(t)$ is a sequence of independent random variables with mean zero, variance σ^2 and finite cumulants κ_h of all order.

ASSUMPTION 2. The kernel function $w(x)$ is a symmetric probability density with support on $[-1/2, 1/2]$ and the second-order derivative of $w(x)$ satisfies a Lipschitz condition of order $\alpha > 1/2$.

ASSUMPTION 3. There exist positive constants M_1, M_2, K_1 and K_2 such that $M_1|j|^{-K_1} \ge T^{-1}I_S(\lambda_j)$ for $|j| \le N$ and $T^{-1}I_S(\lambda_j) \ge M_2|j|^{-K_2}$ for $|j| \le T^{1/K_1}$.

Assumptions 1 and 2 are quite standard. The part in Assumptions 3 that requires $I_S(\lambda_j)$ to decay fast is not strict; when $m^{(k)}(x)$ is of bounded variation and $m^{(l)}(0) = m^{(l)}(1)$ for $l < k$, then $T^{-1}I_S(\lambda_j) \le M|j|^{-2(k+1)}$ for some constant M [Zygmund (1959), page 241]. The other part that requires $I_S(\lambda_j)$ to decay in some regular way is a crucial condition. We need this condition to ensure that the stabilized procedure does not cut off (determined by J) too much signal; thus the bias caused by dropping $I_S(\lambda_j)$ at high frequency is negligible. A modification, which requires less strict conditions, is given in Section 6.

In the statements of the theorems below, let $\theta_{0T} = T^{1/5}\beta_{0T}$, $\tilde{\theta} = T^{1/5}\tilde{\beta}$, where $\tilde{\beta}$ is the global minimizer of $\tilde{R}_T(\beta)$ (3.7). Also let $\hat{\beta}$ be a consistent local minimizer of $\hat{R}_T(\beta)$ (2.3). Under commonly used conditions, Theorem 1 establishes the strong consistency of $\tilde{\theta}$.

THEOREM 1. *Under Assumptions 1–3, with $K_1 > 5$ and assuming $\hat{\sigma}^2$ is a strongly consistent estimate of σ^2, $\tilde{\theta}$ converges to θ_0 almost surely.*

With slightly stronger conditions, Theorem 2 shows that the stabilized bandwidth estimate converges to β_{0T} faster than the classical one.

THEOREM 2. *Under the assumptions in Theorem* 1 *with* $K_1 > 5.5$ *and* $(K_1 - 5)/K_2 > 1/10$, $\tilde{\beta} - \beta_{0T} = o_p(\hat{\beta} - \beta_{0T})$.

In fact, if $m(x)$ is as smooth as assumed in Theorem 3, we can get the quite surprising result that the convergence rate of $\tilde{\theta}$ is $T^{-1/2}$ instead of the rate $T^{-1/10}$ of the classical estimates.

THEOREM 3. *Under Assumptions* 1–3 *with* $10 < K_1 \le K_2 < 2K_1 - 10$, *and letting* $\hat{\sigma}^2$ *be either* $\hat{\sigma}_1^2$ *or* $\hat{\sigma}_2^2$ *(with* $j_0/T \to m > 0$*)*; $T^{1/2}(\tilde{\theta} - \theta_{0T})$ *is asymptotically normal with mean zero and variance*

$$\{A''(\theta_0)\}^{-2}\left[\theta_0^{-4}\left\{\int w^2(x)\,dx\right\}^2 V + 4\theta_0^6\sigma^2\left\{\int x^2 w(x)\,dx\right\}^4 \int \{m^{(4)}(x)\}^2\,dx\right],$$

where V is the asymptotic variance of $T^{1/2}(\hat{\sigma}^2 - \sigma^2)$.

The quantity $\int\{m^{(4)}(x)\}^2\,dx$ can be estimated by $2T^{-2}\sum_{j=1}^{J}(2\pi j)^8|d_Y(\lambda_j)|^2$.

To simplify the discussion below, we now assume σ^2 is known. Rewrite (3.7) to

$$\begin{aligned}\tilde{R}_T(\beta) &= 4\pi\sum_{j=1}^{J-1}\{I_Y(\lambda_j) - \sigma^2/(2\pi)\}\{1 - W_\beta(\lambda_j)\}^2 \\ &\quad + \sigma^2\sum_{j=0}^{N} W_\beta^2(\lambda_j) + \sigma^2.\end{aligned} \tag{4.1}$$

Comparing (4.1) with (3.2), we see that $\tilde{R}_T(\beta)$ uses $4\pi\sum_{j=1}^{J}\{I_Y(\lambda_j) - \sigma^2/(2\pi)\}\{1 - W_\beta(\lambda_j)\}^2$ to estimate the bias term in $R_T(\beta)$. Noting that $1 - W_\beta(\lambda_j) \approx 2^{-1}(2\pi\beta j)^2\int w(x)x^2\,dx$ leads to the consideration of a "plug-in" bandwidth estimate, which is obtained by replacing $G = \int\{m''(x)\}^2\,dx$ in (2.2) by the estimate $\hat{G} = 2T^{-2}\sum_{j=1}^{J}(2\pi j)^4|d_Y(\lambda_j)|^2$. Following the arguments in Section 7, similar results for the plug-in estimate can be established. In fact, when σ^2 in (2.2) is also replaced by $\hat{\sigma}^2$, the plug-in estimate has the asymptotic variance described in Theorem 3.

The discussion above shows an important point that estimating the regression function and estimating the bias term in $R_T(\beta)$ are two different and separate problems. Previously, the two problems were often mixed together, which obscures the nature of the bandwidth selection problem. Another important point is that it is not necessary to use the kernel method to estimate $R_T(\beta)$.

5. Other selectors and simulation results. In addition to the selector (2.3), we also consider the generalized cross-validation

$$\text{GCV}_T(\beta) = \text{RSS}_T(\beta)/\{1 - T^{-1}\beta^{-1}w(0)\}^{-2} \tag{5.1}$$

[Craven and Wahba (1979)], the selector T of Rice (1984),

$$\text{T}_T(\beta) = \text{RSS}_T(\beta)/\{1 - 2T^{-1}\beta^{-1}w(0)\} \tag{5.2}$$

and the well known cross-validation $\text{CV}_T(\beta) = \sum_{t=0}^{T-1}\{Y(t) - \bar{S}_\beta(t)\}^2$ [Clark (1975)], where

$$\bar{S}_\beta(t) = \sum_{u \neq t} w\{(t-u)/(\beta T)\}Y(t) \Big/ \sum_{u \neq 0} w\{u/(\beta T)\}$$

is a "leave-one-out" estimate of $S(t)$. The cross-validation can be rewritten as

$$\begin{aligned} \text{CV}_T(\beta) = 4\pi \sum_{j=1}^{N} I_Y(\lambda_j)\left\{1 + U_1^{-1}w(0) - U_1^{-1}U_2W_\beta(\lambda_j)\right\}^2 \\ + O(T^{-1}\beta^{-1}), \end{aligned} \tag{5.3}$$

where $U_2 = \sum w\{u/(\beta T)\}$ and $U_1 = U_2 - w(0)$. Similar to (3.5), the periodogram $I_Y(\lambda_j)$ in (5.3) can be replaced by $\tilde{I}_Y(\lambda_j)$ of (3.6) to obtain the stabilized cross-validation. The stabilized GCV and T selectors are obtained by substituting SRSS for RSS in (5.1) and (5.2), respectively.

We carried out some simulations to compare the performance of the bandwidth estimates. The observations $Y(t)$ were obtained by adding independent Gaussian random variables with mean 0 and variance $\sigma^2 = 0.003^2$ to $m(x_t) = x_t^3(1-x_t)^3$, $x_t = t/T$. The function $m(x)$ was the one considered in Rice (1984). We also used the same kernel function $w(x)$ used in Rice (1984). All random variables were generated by using the function RAND in Fortran 77 on a SUN 3/50 computer.

Five hundred series with $T = 75$ were generated. For each observed series, the minimizers of the criteria were obtained by searching over 451 equally spaced points in the interval [0.04, 0.49]. The noise variance was estimated by $\hat{\sigma}_2^2$ with $j_0 = 5$; j_0 was selected by inspecting the plots of $I_{\tilde{Y}}(\lambda_j)/|1 - \exp(-i\lambda_j)|^2$ of some sample series. The stabilized residual sum of squares was obtained as described in the previous section with $c = 3 \approx -\log(0.05)$.

As mentioned in Section 1, due to the large sample variation of the classical bandwidth estimates, the kernel estimates based on these bandwidths could have very different appearances for different samples. Two sample series are plotted in Figure 1(a) and (b); the dashed curves are the regression function $m(x)$ and the solid curves are the kernel estimates with the bandwidths 0.157 [Figure 1(a)] and 0.357 [Figure 1(b)] selected by GCV. The bandwidths are about, respectively, the 10th and 90th percentiles of the empirical distribution of $\hat{\beta}_{\text{GCV}}$. Since the empirical distribution of $\hat{\beta}_{\text{GCV}}$ has a short right tail, and is biased toward undersmoothing, the bandwidth 0.357 is not too far away from 0.307 (β_{0T}). So the kernel estimate in Figure 1(b) does not give a bad fit. For the same pair of samples, Figure 2(a) and (b) shows the kernel estimates (solid

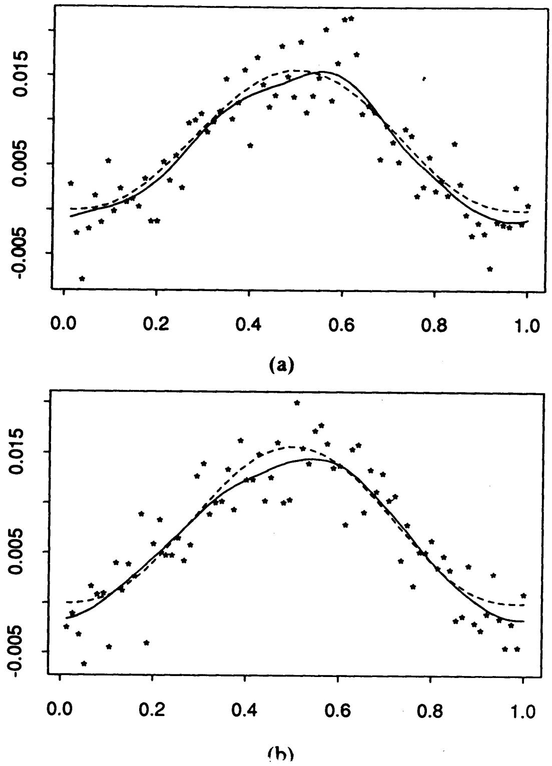

FIG. 2. *The same pair of samples shown in Figure 1. The dashed curves are the regression function and the solid curves are the kernel estimates based on the bandwidths* 0.302 [*Figure* 1(a)] *and* 0.300 [*Figure* 1(b)] *selected by SGCV.*

curves) based on the bandwidths 0.302 and 0.300 selected by the stabilized GCV. The kernel estimates in Figure 2(a) and (b) look more similar than the pair of estimates in Figure 1(a) and (b).

Figure 3 compares $R_T(\beta)$ with $\hat{R}_T(\beta)$ and $\tilde{R}_T(\beta)$ for the data set shown in Figures 1(a) and 2(a). The stars in Figures 3–4 indicate the locations of the optimal bandwidths β_{0T}. While $\hat{R}_T(\beta)$ differs from $R_T(\beta)$ widely, the stabilized criterion $\tilde{R}(\beta)$ is very close to the risk function.

Table 1 summarizes the sample means and standard deviations of the bandwidth estimates and $\mathrm{SSE}_T(\hat{\beta})$. The last row gives β_{0T}, the approximate standard deviation of the stabilized estimates (from Theorem 3), and $R_T(\beta_{0T})$. We should point out that $I_S(\lambda_j) \approx O(Tj^{-7.8})$ and thus $m(x)$ does not satisfy the conditions of Theorem 3. Figure 4 shows the estimated densities of $\hat{\beta}_{\hat{R}}$ and $\hat{\beta}_{\tilde{R}}$. The densities were obtained by using the S function "density" with a Gaussian kernel and bandwidths 0.06 and 0.02, respectively. The normal probability plots of the bandwidth estimates suggest that the empirical distributions of the stabilized bandwidth estimates are close to the normal distributions. The standard deviations of the stabilized estimates are much smaller

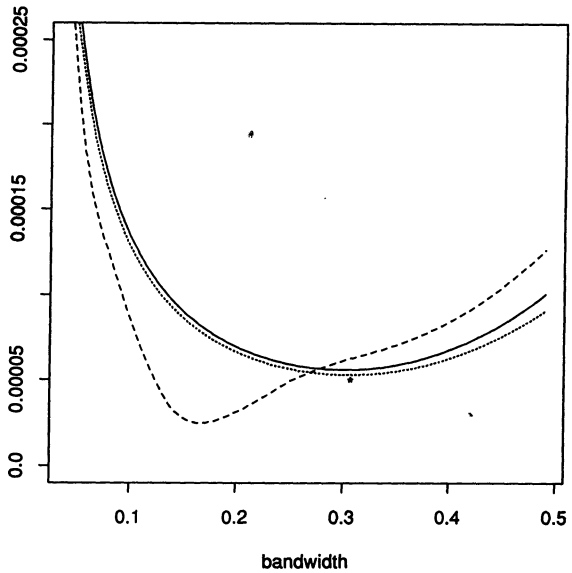

FIG. 3. *Comparison of the risk function* $R_T(\beta)$ *(dotted curve) and the criteria* $\hat{R}_T(\beta)$ *(dashed curve) and* $\tilde{R}_T(\beta)$ *for the data set shown in Figures* 1(a) *and* 2(a). *The star indicates the location of the optimal bandwidth.*

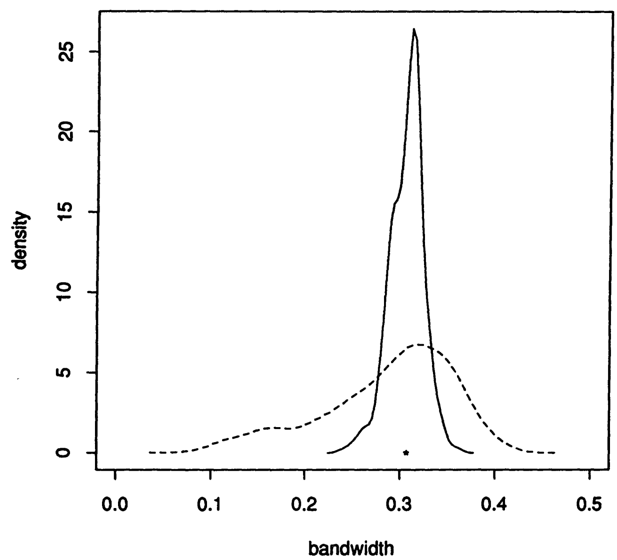

FIG. 4. *Estimated densities of the bandwidth estimates* $\hat{\beta}_{\hat{R}}$ *(dashed curve) and* $\hat{\beta}_{\tilde{R}}$ *(solid curve). The sample size is* 500 *and* $T = 75$. *The star indicates the location of the optimal bandwidth.*

TABLE 1
Summary of the sample means, the standard deviations and SSE. The sample size is 500 *with* $T = 75$. *The last row gives the optimal bandwidth* β_{0T}, *the approximate standard deviation from Theorem* 3 *and* $R_T(\beta_{0T})$

Selector	$E(\hat{\beta})$	$SD(\hat{\beta})$	$E\{SSE_T(\hat{\beta})\}$
$\hat{R}$	0.287	0.0685	6.48e − 5
$\tilde{R}$	0.306	0.0177	5.48e − 5
GCV	0.276	0.0827	7.36e − 5
SGCV	0.306	0.0176	5.48e − 5
T	0.306	0.0598	6.13e − 5
ST	0.321	0.0176	5.50e − 5
CV	0.276	0.0827	7.36e − 5
SCV	0.306	0.0177	5.48e − 5
	0.307	0.0236	5.30e − 5

than those of the classical estimates. It is interesting to note that, although Rice (1984) intentionally designed the selector T to be biased toward oversmoothing, T is the only classical selector that gives an unbiased estimate. However, the stabilized T is substantially biased.

The kernel estimates based on the stabilized bandwidth estimates give better fits. From Table 1, we see that the sample means of SSE are reduced by as much as 25%. The sample means of SSE of the stabilized estimates are remarkably close to $R_T(\beta_{0T})$. Table 2 gives some empirical quantiles of the ratio $\mathrm{SSE}_T(\beta_{\mathrm{GCV}})/\mathrm{SSE}_T(\hat{\beta}_{\mathrm{SGCV}})$. It can be seen that the stabilized estimate often substantially reduces the sum of squares errors. Following Rice (1984) and Härdle, Hall and Marron (1988), Table 3 gives the number of times that the ratio $R_T(\hat{\beta})/R_T(\beta_{0T})$ exceeds certain values. For the stabilized estimates, the ratio is rarely larger than 1.05.

We also obtained similar results from the simulations for different sample sizes and for the regression function $m(x) = 16x^5(1 - x)^5$, which satisfies the conditions of Theorem 3. For the smoother regression function, the approximate standard deviations provided in Theorem 3 agree well with the empirical ones.

TABLE 2
Some empirical quantiles of the ratio $SSE_T(\beta_{GCV})/SSE_T(\hat{\beta}_{SGCV})$. *The sample size is* 500 *with* $T = 75$

p	0.05	0.5	0.75	0.9	0.95
Quantile	0.91	1.06	1.27	2.48	3.77

TABLE 3
Number of exceedances by $R_T(\hat{\beta})/R_T(\beta_{0T})$. The sample size is 500 with $T = 75$

Selector	1.05	1.1	1.2	1.4	1.6	1.8	2	4
T	188	100	46	19	10	7	2	0
$\hat{R}$	206	132	76	48	30	17	11	0
GCV	220	155	105	75	59	49	39	5
CV	220	155	105	75	59	49	39	7
ST	11	0	0	0	0	0	0	0
$\tilde{R}$	7	0	0	0	0	0	0	0
SGCV	7	0	0	0	0	0	0	0
GCV	8	0	0	0	0	0	0	0

6. Remarks. This section gives some remarks. We discuss some interesting problems and suggest some possible approaches. It is clear that for most problems, both theoretic and practical issues deserve further study. It is our hope that the discussion can stimulate more research interest.

REMARK 1. The problem of boundary effects has rarely been studied. Boundary effects also cause trouble on classical selectors. The main difficulty is that when $m(0) \neq m(1)$, $|d_s(\lambda_j)|$ decays slowly (the rate is T/j). The decaying rate can be accelerated by multiplying a smooth function (taper) to $Y(t)$. This is called "tapering" in time-series analysis. A closely related procedure was suggested in Härdle, Hall and Marron (1988), who considered a weighted risk function and a weighted RSS. These are reasonable approaches since there are fewer data available for estimating $m(x)$ at the boundaries. A practical problem is the selection of a proper taper to ensure that the Fourier transform of the tapered signal decays fast enough. A more important problem is that when $|m(0) - m(1)|$ is large, one has to taper the observation heavily. In this case, the bandwidth is essentially determined by the observations at the center. As pointed out by a referee, we can fit a line to the data and obtain the bandwidth from the residuals. We have done some preliminary simulation studies and found that this procedure works well. Since the difference between $m(0)$ and $m(1)$ would be greatly reduced, we would only need to taper the residuals slightly to overcome the boundary effects.

REMARK 2. Following the approach of Chiu (1989b), the proposed procedures can be modified similarly for the case of the corrected noise series, which has a parametric spectral density function.

REMARK 3. When the signal contains periodic components at high frequency, the stabilized procedure will modify the periodogram ordinates at the frequencies of the periodic components. This could be a desirable feature because we might not want the trend estimate to be affected much by the high-frequency components. For extracting such a signal, it might be better to use a combination of a low-pass filter (such as a kernel smoother) and some bandpass filters. See Chiu (1989a) for a discussion of detecting periodic components.

REMARK 4. We now describe a modification of the stabilized procedure. The only change is on the selection of J. We first select a J_1 which is the largest j such that $I_Y(\lambda_j) > \log(T)\hat{\sigma}^2/\pi$. We then choose J to be the first integer larger than J_1 such that $I_Y(\lambda_j) < c\hat{\sigma}^2/(2\pi)$. With a stronger condition on the cumulants of $\varepsilon(t)$, $\sum|\kappa_h|z^h/h! < \infty$ for z in a neighborhood of 0, we have, with probability 1, $\max I_\varepsilon(\lambda_j) \le \log(T)\sigma^2/\pi$ [Theorem 5.3.2 of Brillinger (1981)]. The second part of Assumption 3 can be relaxed to: For any $\delta > 0$, there exists a sequence $j_T > T^{1/K_2-\delta}$ such that $\liminf I_S(\lambda_{j_T})/\log(T) \ge m > \sigma^2/\pi$ (m could be ∞). Under the conditions above, Lemmas 3 and 5 hold for the modified procedure, as do Theorems 1–3. In practice, one can use the modified procedure as a safeguard. When the procedure disagree, the data set contains some subtle features and deserves more study.

REMARK 5. Since this article was first written, we have seen the manuscript of Hall and Marron (1989). For the density case, they showed that the fastest relative convergence rate of bandwidth estimates is $T^{-1/2}$ (T, the sample size), which is the rate given in Theorem 3 for the stabilized bandwidth estimate. By considering a normal density with the standard deviation σ, the reason for the lower bound can be easily understood [Hall and Marron (1989)]. The optimal bandwidth depends on σ, and the best convergence rate is $T^{-1/2}$ for any estimate of σ. For the regression case, the unknown noise variance sets the lower bound even when $m(x)$ is known. We note that the convergence rate of the stabilized bandwidth estimate is faster than the lower bound given in Theorem 3.2 of Hall and Marron (1989). For example, for the class of functions which have bounded second-order derivatives, the minimax lower bound given in Hall and Marron (1989) is $T^{-1/10}$. However, if $m^{(2)}(x)$ is of bounded variation, the relative convergence rate of the stabilized bandwidth estimate is $o(T^{-1/6+\delta})$ for any $\delta > 0$.

REMARK 5. Whether one should try to minimize SSE or MSSE is an important issue. Hall and Marron (1989) discussed this issue in some detail. They argued that since the relative convergence rate of estimates of $\hat{\beta}_0$, the minimizer of ISE, cannot be faster than $T^{-1/10}$, one should try to minimize MISE. Here we give an intuitive explanation for the difficulty in "estimating" $\hat{\beta}_0$. From Chiu and Marron (1990), $\hat{\beta}_0$ is approximately proportional to

$$-8\pi \sum_{j=1}^{N} \{I_\varepsilon(\lambda_j) - \sigma^2/(2\pi)\} W_{\beta_{0T}}(\lambda_j)(\partial/\partial\beta) W_{\beta_{0T}}(\lambda_j)$$

$$+ 4T^{-1}\,\mathrm{Re}\sum_{j=1}^{N} d_S(\lambda_j)\, d_\varepsilon(-\lambda_j)(\partial/\partial\beta) W_{\beta_{0T}}(\lambda_j).$$

The trouble here is that $d_\varepsilon(\lambda_j)$ at low frequency does make a substantial contribution to the second term above (the variance of each term at low frequency is of order $T^{3/10}$). To get a good estimate of $\hat{\beta}_0$, we have to separate the signal and noise at low frequency, which is a difficult job unless $S(t)$ has some special structure. It was observed in Härdle, Hall and Marron (1988), and explained in Chiu and Marron (1990) that the minimizer of SSE and the classical bandwidth estimates are negatively correlated. Therefore, the stabilized bandwidth estimates are closer to $\hat{\beta}_0$ than are the classical estimates.

7. Proofs. This section establishes the results stated in previous sections. We first derive the bounds for J through the following lemmas.

LEMMA 1. *Under Assumption* 1 *and assuming that* $b - a = o(T)$, *for any positive integer* k,

$$(7.1)\quad E\left[\left\{(b-a)^{-1}\sum_{a\le j<b}\left(I_\varepsilon(\lambda_j)-\sigma^2/(2\pi)\right)\right\}^{2k}\right]=O\{(b-a)^{-k}\}.$$

PROOF. First, note that for any $l \ge 1$,

$$(7.2)\qquad \sum_{a\le j_1<b}\cdots\sum_{a\le j_l<b}\operatorname{cum}\{I_\varepsilon(\lambda_{j_1}),\dots,I_\varepsilon(\lambda_{j_l})\}=O\{(b-a)\}.$$

This can be shown by using the technique described on pages 19–21 of Brillinger (1981) and noting that

$$\operatorname{cum}\{d_\varepsilon(\omega_1),\dots,d_\varepsilon(\omega_h)\}=\begin{cases}T\kappa_h, & \omega_1+\cdots+\omega_h\equiv 0\ (\mathrm{mod}\,2\pi),\\ 0, & \text{otherwise,}\end{cases}$$

where ω_j's are Fourier frequencies. Defining $B_h = \sum_{j=a}^{b-1}\{I_\varepsilon(\lambda_j) - \sigma^2/(2\pi)\}$ for $h = 1,\dots,2k$, we see that the left-hand side of (7.1) is equal to

$$(b-a)^{-2k}\sum_{\boldsymbol{\nu}}\operatorname{cum}\{B_h, h\in\nu_1\}\cdots\operatorname{cum}\{B_h, h\in\nu_p\},$$

where the summation above is over all partitions $\boldsymbol{\nu} = \nu_1 \cup \cdots \cup \nu_p$ of the set $\{1,\dots,2k\}$. Since $EB_h = 0$, any partition that gives nonzero contribution must have $p \le k$. The lemma follows this and (7.2). □

LEMMA 2. *Under Assumption* 1 *and assuming that* $b - a = o(T)$ *and* $T^\delta/(b-a) = o(1)$ *for some* $\delta > 0$, *for any constant* $c' > 1$, $\min_{a\le j<b} I_\varepsilon(\lambda_j) \le c'\sigma^2/(2\pi)$ *with probability* 1.

PROOF. It is clear that

$$P\left\{\min_{a\le j<b} I_\varepsilon(\lambda_j) > c'\sigma^2/(2\pi)\right\}$$
$$\le P\left[(b-a)^{-1}\left\{\sum_{j=a}^{b-1}\left(I_\varepsilon(\lambda_j)-\sigma^2/(2\pi)\right)\right\} > (c'-1)\sigma^2/(2\pi)\right].$$

From Lemma 1 and the Chebyshev inequality [Chung (1974), page 48], the right-hand side above is of order $(b-a)^{-k} = o(T^{-k\delta})$. The lemma is established by choosing a large enough k. □

From Lemmas 1 and 2, Lemma 3 gives an upper bound of J.

LEMMA 3. *Under Assumptions* 1 *and* 3 *and assuming* $\hat{\sigma}^2$ *converges to* σ^2 *almost surely, for any* $\delta > 0$, $J \le T^{\delta + 1/K_1}$ *with probability* 1.

PROOF. Let $b = [T^{\delta + 1/K_1}]$ and $a = [T^{\delta_1 + 1/K_1}]$ for some $\delta > \delta_1 > 0$. It is sufficient to show that, with probability 1, only finitely many events $\{\min_{a \le j < b} I_Y(\lambda_j) > c\hat{\sigma}^2/(2\pi)\}$ occur. The event above is contained in the union of the events

$$\min_{a \le j < b} I_Y(\lambda_j) > c'\sigma^2/(2\pi) \tag{7.3}$$

and

$$\hat{\sigma}^2 - \sigma^2 < -(c - c')\sigma^2/c \tag{7.4}$$

for any $c > c' > 1$. From Lemma 2 and Assumption 3, with probability 1, only finitely many events of (7.3) occur. The same statement holds for (7.4) from the strong consistency of $\hat{\sigma}^2$. □

Lemma 4 is a direct consequence of Theorem 4.5.4 of Brillinger (1981).

LEMMA 4. *Under Assumption* 1 *and for any* $\delta > 0$, $\max_{1 \le j \le N} I_\varepsilon(\lambda_j) \le T^\delta$ *with probability* 1.

Similar to Lemma 3, we obtain an upper bound for J.

LEMMA 5. *Under the conditions of Lemma* 3 *and for any* $\delta > 0$, $J \ge T^{1/K_2 - \delta}$ *with probability* 1.

Next we proceed to prove Theorems 1–3. Comparing (3.2) with (3.7) yields, for odd T,

$$\tilde{R}_T(\beta) - R_T(\beta) = D(\beta) = D_1(\beta) + D_2(\beta) + D_3(\beta) + D_4(\beta) + D_5(\beta),$$

where

$$D_1(\beta) = -4\pi \sum_{j=J}^{N} I_S(\lambda_j)\{1 - W_\beta(\lambda_j)\}^2,$$

$$D_2(\beta) = 4\pi \sum_{j=1}^{J-1} \{I_\varepsilon(\lambda_j) - \sigma^2/(2\pi)\}\{1 - W_\beta(\lambda_j)\}^2,$$

$$D_3(\beta) = 4T^{-1}\,\mathrm{Re} \sum_{j=1}^{J-1} d_S(\lambda_j)\, d_\varepsilon(-\lambda_j)\{1 - W_\beta(\lambda_j)\}^2,$$

$$D_4(\beta) = 2(\sigma^2 - \hat{\sigma}^2) \sum_{j=1}^{J-1} \{1 - W_\beta(\lambda_j)\}^2$$

and

$$D_5(\beta) = 2(\hat{\sigma}^2 - \sigma^2) \sum_{j=0}^{N} W_\beta^2(\lambda_j) + \sigma^2.$$

For even T, $D_1(\beta)$ and $D_5(\beta)$ drop the terms at frequency π, which have negligible effects.

In the following discussion, we let M be a generic positive constant. The definition of M will depend on the context in which it is used. Also, unless stated otherwise, all bounds concerning random variables are probability 1 bounds. Letting $\tilde{W}(\beta j) = \int w(x)\exp(-i2\pi\beta jx)\,dx$, Assumption 2 implies that $|\tilde{W}(\beta j)| \le M\beta^{-3}j^{-3}$ and $|(\partial/\partial\beta)\tilde{W}(\beta j)| \le M\beta^{-3}j^{-2}$ for $0 < \beta \le 1$ [see Edwards (1979), pages 32–34 and note that $w(1/2) = w'(1/2) = 0$]. Since $W_\beta(\lambda_j) = \sum_{n=-\infty}^{\infty} \tilde{W}\{\beta(nT + j)\}$,

$$\left|W_\beta(\lambda_j) - \tilde{W}(\beta j)\right| \le M \sum_{n=1}^{\infty} \{\beta(nT + j)\}^{-3} = O(T^{-3}\beta^{-3}).$$

Therefore, $W_\beta(\lambda_j)$ can be replaced by $\tilde{W}(\beta j)$. Similarly, the derivatives $(\partial/\partial\beta)W_\beta(\lambda_j)$ can be replaced by $(\partial/\partial\beta)\tilde{W}(\beta j)$. We will ignore the errors caused by the replacement. Also note that

$$\tilde{W}(\lambda) = 1 - 2^{-1}(2\pi\lambda)^2 \int w(x)x^2\,dx + O(\lambda^4)$$

and so $|1 - \tilde{W}(\beta j)| \le M(\beta^2 j^2)$ for $j \le \beta^{-1}$.

The uniform convergency of $\tilde{R}_T(\beta)$ is given in Lemma 6.

LEMMA 6. *Under the conditions of Theorem* 1, $\tilde{R}_T(\beta)/R_T(\beta)$ *converges to* 1 *almost surely and uniformly on* $T^{-1/5-\tau} \le \beta \le T^{1/5+\tau}$ *for any positive constant* $\tau < \min\{(K_1 - 5)/(4K_2), (K_1 - 5)/(5K_1)\}$.

PROOF. Letting $\delta > 0$ be a constant such that $J < T^{1/K_1+\delta} < T^{1/5-\tau}$, we have

$$|D_1(\beta)| \le M\{\beta^4 T^{1+(-K_1+5)(1/K_2+\delta)} + T\beta^{K_1-1}\} \tag{7.5}$$

and

$$\sum_{j=1}^{J-1} \{1 - W_\beta(\lambda_j)\}^2 \le M\beta^4 T^{5(1/K_1+\delta)}. \tag{7.6}$$

From Lemma 4, and the fact that $\max_\lambda |d_\varepsilon(\lambda)| = o(T^{1/2+\delta})$ for any $\delta > 0$ [Theorem 4.5.4 of Brillinger (1981)], we also have, for some small $\delta_1 > 0$,

$$|D_3(\beta)| \le \begin{cases} M\beta^4 T^{1/2+\delta_1+(-K_1/2+5)(1/K_1+\delta)}, & K_1 \le 10, \\ M\beta^4 T^{1/2+\delta_1}, & K_1 > 10. \end{cases} \tag{7.7}$$

Similarly, since $\max_\lambda I_\varepsilon(\lambda) = o(T^\delta)$ for any $\delta > 0$,

$$|D_2(\beta)| \le M\beta^4 T^{5(1/K_1+\delta)+\delta_1}. \tag{7.8}$$

Also note that $\sum_{j=0}^{N} W_\beta^2(\lambda_j) \le M\beta^{-1}$. From (7.5)–(7.8) and the definition of

$D_4(\beta)$ and $D_5(\beta)$, it can be seen that, by selecting small enough δ and δ_1, $D_k(\beta) = o(T^{1/5})$, $k = 1, \dots, 5$, uniformly on $T^{-1/5-\tau} \le \beta \le T^{-1/5+\tau}$. The lemma follows the fact that $R_T(\beta) \ge MT^{1/5}$. □

PROOF OF THEOREM 1. From the definition of $\tilde{R}_T(\beta)$ [also see (4.1)], we see that

$$\tilde{R}_T(\beta) \ge \begin{cases} MT^{1/5+\tau}, & 0 < \beta < T^{-1/5-\tau}, \\ MT^{1/5+4\tau}, & T^{-1/5+\tau} < \beta < \infty. \end{cases}$$

Since $\tilde{R}_T(T^{-1/5)}) = O(T^{1/5})$, the global minimizer of $\tilde{R}_T(\beta)$ must be inside $T^{-1/5-\tau} \le \beta \le T^{-1/5+\tau}$ for large enough T. The strong consistency of $\tilde{\theta}$ follows this and Lemma 6. □

PROOF OF THEOREM 2. In this and the proof of Theorem 3, we let $\beta = \theta T^{-1/5}$ for some constant $\theta > 0$. From Taylor expansion, we have $D'(\tilde{\beta}) = (\tilde{\beta} - \beta_{0T})R_T''(\bar{\beta})$ for some $\bar{\beta}$ in between β_{0T} and $\tilde{\beta}$. Similarly, $\hat{D}'(\hat{\beta}) = (\hat{\beta} - \beta_{0T})R_T''(\bar{\beta})$ for some $\bar{\beta}$ in between β_{0T} and $\hat{\beta}$, where $\hat{D}(\beta) = \hat{R}_T(\beta) - R_T(\beta)$. Since $\hat{D}'(\beta_{0T})$ is of order $T^{3/10}$ [Rice (1984) and Chiu (1990b)], it is sufficient to prove that $D'(\beta_{0T}) = o(T^{3/10})$. This can be shown by noting that $\{1 - W_\beta(\lambda_j)\}(\partial/\partial\beta)W_\beta(\lambda_j) = O(\beta^3 j^4)$ for $j \le \beta^{-1}$ and is of order $\beta^{-3}j^{-2}$ for $j > \beta^{-1}$ and following the arguments in Lemma 6. □

PROOF OF THEOREM 3. The strong consistency and the asymptotic distribution of $\hat{\sigma}_2^2$ were established in Chiu (1989b). For the estimate $\hat{\sigma}_1^2$, we note that $\sum_{t=0}^{T-1}\tilde{S}^2(t) \le MT^{-1}$ and so

$$\sum_{j=1}^{N} \left|d_{\tilde{Y}}(\lambda_j)\right|^2 = \sum_{j=1}^{N} \left|d_{\tilde{\varepsilon}}(\lambda_j)\right|^2 + O(T),$$

where $\tilde{\varepsilon}(t) = \varepsilon(t) - \varepsilon(t-1)$. Since $\tilde{\varepsilon}(t)$ is independent with $\tilde{\varepsilon}(t')$ for all $|t - t'| > 1$, the strong consistency and the asymptotic distribution of $\hat{\sigma}_1^2$ can be established easily.

Before showing that $D_5'(\beta)$ and $D_3'(\beta)$ are of order $O_p(T^{-1/10})$, we note that the other terms are of smaller order. It is clear that $D_2'(\beta)$ and $D_4'(\beta)$ are of order $o(T^{-1/10})$ when $K_1 > 10$. When $10 < K_1 \le K_2 \le 2K_1 - 10$, $D_1'(\beta)$ is also of order $o(T^{-1/10})$. Since

$$2(\partial/\partial\beta)\sum_{j=1}^{N} W_\beta^2(\lambda_j) = -\beta^{-2}\int w^2(x)\,dx + O(T^{-1}\beta^{-2}),$$

$$D_5'(\beta) \approx (\hat{\sigma}^2 - \sigma^2)\beta^{-2}\int w^2(x)\,dx = O_p(T^{-1/10}).$$

Since

$$T^{-1}\beta^3 \sum_{j=J}^{N} d_S(\lambda_j)\, d_\varepsilon(-\lambda_j) j^4 = o(T^{-1/10}),$$

$$(7.9)\qquad D_3'(\beta) = 4T^{-1}\beta^3\left\{\int x^2 w(x)\,dx\right\}^2 \operatorname{Re}\sum_{j=1}^{N} d_S(\lambda_j)\, d_\varepsilon(-\lambda_j)(2\pi j)^4 + o(T^{-1/10}).$$

The variance of the first term in (7.9) is equal to

$$(7.10)\qquad 8\beta^6 T\sigma^2\left\{\int x^2 w(x)\,dx\right\}^4 \sum_{j=1}^{N} (2\pi j)^8 \left|d_S(\lambda_j)\right|^2 / T^2.$$

Note that $T^{1/5}$ times (7.10) converges to $4\theta^6\sigma^2\{\int x^2 w(x)\,dx\}^4 \int \{m^{(4)}(x)\}^2\,dx$.

Since $d_{\tilde{\varepsilon}}(\lambda_j) = \{1 - \exp(-i\lambda_j)\}\, d_\varepsilon(\lambda_j) + O_p(1)$, and $\operatorname{cum}\{|d_\varepsilon(\lambda_j)|^2, d_\varepsilon(\lambda_k)\} = 0$ for any $1 \le |j|, |k| \le N$, the covariance between the first term of (7.9) and

$$\beta^{-2}T^{-2}\sum_{j=1}^{N} \left|d_{\tilde{\varepsilon}}(\lambda_j)\right|^2 \left|1 - \exp(-i\lambda_j)\right|^2 \phi(\lambda_j)$$

is of order $\beta T^{-1/2} = o(T^{-1/5})$ for any weighting function $\phi(\lambda_j) \le M|1 - \exp(-i\lambda_j)|^{-2}$. Therefore, $D_3'(\beta)$ and $D_5'(\beta)$ are asymptotically uncorrelated.

The asymptotic normality follows the fact that the kth order joint cumulants of $D_5'(\beta)$ and $D_3'(\beta)$ are of order $o(T^{-k/10})$ for $k \ge 3$, which can be shown by some quite straightforward but notationally complicated computation [see the proof of Theorem 5.10.1 of Brillinger (1981)]. □

Acknowledgments. The valuable comments from the referee and Associate Editor are gratefully appreciated. The comments significantly improve the presentation. The author thanks Professor P. Hall and Professor J. S. Marron for providing their preprint. The first version of the article was written at Rice University.

REFERENCES

Brillinger, D. R. (1981). *Time Series Data Analysis and Theory*. Holt, Rinehart, and Winston, New York.

Chiu, S.-T. (1989a). Detecting periodic components in a white Gaussian time series. *J. Roy. Statist. Soc. Ser. B* **51** 249–259.

Chiu, S.-T. (1989b). Bandwidth selection for kernel estimation with correlated noise. *Statist. Probab. Lett.* **8** 347–354.

Chiu, S.-T. (1990a). Why do bandwidth selectors tend to choose smaller bandwidths and a remedy? *Biometrika* **77** 222–226.

Chiu, S.-T. (1990b). On the asymptotic distributions of bandwidth estimates. *Ann. Statist.* **18** 1696–1711.

Chiu, S.-T. and Marron, J. S. (1990). The negative correlations between data-determined bandwidths and the optimal bandwidth. *Statist. Probab. Lett.* **10** 173–180.

Chung, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic, New York.

CLARK, R. M. (1975). A calibration curve for radiocarbon dates. *Antiquity* **49** 251–266.

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

EDWARDS, R. E. (1979). *Fourier Series: A Modern Introduction* **1**, 2nd ed. Springer, New York.

HALL, P. and MARRON, J. S. (1989). Lower bounds for bandwidth selection in density estimation. Unpublished manuscript.

HÄRDLE, W., HALL, P. and MARRON, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *J. Amer. Statist. Assoc.* **83** 86–101.

LI, K. C. (1985). From Stein's unbiased risk estimates to the method of generalized cross-validation. *Ann. Statist.* **13** 1352–1377.

LI, K. C. (1987). Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation. *Ann. Statist.* **15** 958–975.

MALLOWS, C. (1973). Some comments on C_p. *Technometrics* **15** 661–675.

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

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

SILVERMAN, B. W. (1984). A fast and efficient cross-validation method for smoothing parameter choice in spline regression. *J. Amer. Statist. Assoc.* **79** 584–589.

SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. Roy. Statist. Soc. Ser. B* **47** 1–52.

ZYGMUND, A. (1959). *Trigonometric Series* **1**. Cambridge University Press.

DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523