The Annals of Statistics
1995, Vol. 23, No. 5, 1474–1495

ADAPTIVE ROOT n ESTIMATES OF INTEGRATED SQUARED DENSITY DERIVATIVES[1]

BY TIEE-JIAN WU

University of Houston

Based on a random sample of size n from an unknown density f on the real line, the nonparametric estimation of $\theta_k = \int\{f^{(k)}(x)\}^2\,dx$, $k = 0, 1, \dots,$ is considered. These functionals are important in a number of contexts. The proposed estimates of θ_k is constructed in the frequency domain by using the sample characteristic function. It is known that the sample characteristic function at high frequency is dominated by sample variation and does not contain much information about f. Hence, the variation of the estimate can be reduced by modifying the sample characteristic function beyond some cutoff frequency. It is proposed to select adaptively the cutoff frequency by a generalization of the (smoothed) cross-validation. The exact convergence rate of the proposed estimate to θ_k is established. It depends solely on the smoothness of f. For sufficiently smooth f, it is shown that the proposed estimate is asymptotically normal, attains the optimal $O_p(n^{-1/2})$ rate and achieves the information bound. Finally, to improve the performance of the proposed estimate at small to moderately large n, two modifications are proposed. One modification is for estimating θ_0; it reduces bias of the estimate. The other modification is for estimating θ_k, $k \geq 1$; it reduces sample variation of the estimate. In simulation studies the superior performance of the proposed procedures is clearly demonstrated.

1. Introduction and motivation. Let $X_1, \dots, X_n$ be a random sample from an unknown density f. Let us write

$$\theta_k = \theta_k(f) = \int_{-\infty}^{\infty}\{f^{(k)}(x)\}^2\,dx, \qquad k = 0, 1, 2, \dots, \tag{1.1}$$

where $f^{(k)}$ is the k-th derivative of f. Note that for smoother f we can express θ_k as

$$\theta_k = (-1)^k \int_{-\infty}^{\infty} f^{(2k)}(x) f(x)\,dx.$$

These nonlinear functionals are of distinct interest from the point of view of actual applications. For example, θ_0 appears in the asymptotic variance of the Wilcoxon rank statistics and in the (Pitman) asymptotic relative efficiency of the Wilcoxon rank test relative to the t-test. The functionals θ_1, θ_2 and θ_3

Received February 1993; revised December 1994.

[1]Research supported by a grant from NSC of Taiwan and a PREF grant from the University of Houston.

AMS 1991 *subject classifications*. Primary 62G05; secondary 62G20.

Key words and phrases. Density derivative, nonparametric information bound, characteristic function, kernel estimate, smoothed cross-validation, bandwidth selection, convergence rate.

appear in the asymptotically optimal bandwidth for histograms, frequency polygons and kernel density estimates. Also, θ_0 has useful applications in projection pursuit because θ_0 appears in the Friedman–Tukey projection index, and $\log \theta_0$ is an upper bound for the Shannon negentropy [we have $E \log f(X_1) \le \log Ef(X_1) = \log \theta_0$ by Jensen's inequality].

The nonparametric estimation of θ_k, based on the observations $X_1, \dots, X_n$, has been investigated by Hall and Marron (1987, 1991a), Bickel and Ritov (1988) and Jones and Sheather (1991). Under suitable conditions, all of these authors obtained asymptotically efficient estimates of θ_k with the optimal convergence rate $O_p(n^{-1/2})$. This means that the mean squared error (MSE) of the estimate is asymptotically equivalent to the following information bound:

$$(1.2) \qquad 4n^{-1}\left\{\int_{-\infty}^{\infty}\{f^{(2k)}(x)\}^2 f(x)\,dx - \theta_k^2\right\} = 4n^{-1}\operatorname{Var} f^{(2k)}(X_1)$$

as $n \to \infty$. Here and below $a_n \sim b_n$ means $a_n b_n^{-1} \to 1$. [See, e.g., Koshevnik and Levit (1976), Ritov and Bickel (1990) and Donoho and Liu (1991) for discussions on the information bound for nonparametric estimates of θ_k.]

The estimates proposed by the above authors are kernel-based and are variations of the estimates $\theta_k(\hat{f}_{h,w})$ or $(-1)^k\int_{-\infty}^{\infty}\hat{f}_{h,w}^{(2k)}(x)\,dF_n(x)$, with $\hat{f}_{h,w}$ being a kernel density estimate of f with kernel w and bandwidth h and F_n the empirical cdf associated with $X_1, \dots, X_n$. Hall and Marron (1987, 1991a) gave "diagonals-out (debiasing)" variations, whereas Jones and Sheather (1991) gave a "diagonals-in" variation. Bickel and Ritov (1988) gave a variation which involved splitting the sample, using an estimated influence function and debiasing. For these estimates, the optimal bound (1.2) can be attained by taking w to be a symmetric kernel of order l, with

$$(1.3) \qquad l \ge 2(m + \alpha - k),$$

and choosing h to be a nonrandom value satisfying

$$(1.4) \qquad hn^{1/(4k+c)} \to \infty \quad \text{and} \quad hn^{1/\{4(m+\alpha-k)\}} \to 0$$

if f has smoothness of order $m + \alpha$ that satisfies

$$(1.5) \qquad m + \alpha > 2k + 4^{-1}c,$$

with $c = 1$ or 2 [here f has smoothness of order $m + \alpha$ means that $f^{(m)}$ exists and is Lipschitz of order $0 < \alpha \le 1$ over $(-\infty, \infty)$]. For example, $c = 2$ in Jones and Sheather (1991), and $c = 1$ in both Hall and Marron (1991a) and Bickel and Ritov (1988) [in fact, they pick $h = n^{-2/(4m+4\alpha+1)}$, which satisfies (1.4)]. However, the optimal rate and bound (1.2) cannot be attained if l is not high enough. In this case, Hall and Marron (1987) and Jones and Sheather (1991) derived asymptotically optimal bandwidths (minimizing asymptotic MSE), which depends on the unknown θ_0 and $\theta_{k+l/2}$.

The above results provide significant insight into the theoretical issue of choosing kernel order and bandwidth. However, their applications require detailed knowledge (at least, on the degree of smoothness) of the unknown f.

Although in practice one can choose them subjectively, there is a great demand for adaptive (data-driven) procedures. Some reasons for using adaptive procedures were given in Silverman (1986).

The purpose of this paper is to propose a kernel-based nonparametric estimate of θ_k which involves using an infinite-order kernel and then selecting the bandwidth adaptively. Furthermore, to improve the performance of the estimate at small to moderately large n, two modifications are also proposed. One modification is for estimating θ_0; it reduces bias of the estimate. The other modification is for estimating θ_k, $k \geq 1$; it reduces sample variation of the estimate. The estimates will be constructed in the frequency domain.

One reason for working with Fourier transforms is that the procedures can be easily implemented in practice by taking fast Fourier transforms [cf. Silverman (1982)]. Another reason for working with Fourier transforms is technical. Since the present problem is estimating θ_k, one need have no qualms over using higher-order kernels in formulating kernel-based estimates. In order to obtain asymptotically best estimates, one needs to take inequality (1.3) into consideration. This leads to the choice of an infinite-order kernel so that (1.3) is always satisfied. Among such kernels, we shall choose the "sync kernel" $K_\infty(x) = (\pi x)^{-1} \sin x$, $-\infty < x < \infty$, because of certain L^2 optimality properties [cf. Davis (1975, 1977), Ibragimov and Khas'minski (1982) and Hall and Marron (1988)]. Note that K_∞ is not in L^1, and it is more convenient to work with its L^2 Fourier transform $\phi_{K_\infty}(\lambda) = I_{[-1,1]}(\lambda)$, $-\infty < \lambda < \infty$, with $I(\cdot)$ denoting the indicator function. Here and below we use $\phi_g(\lambda) = \int_{-\infty}^{\infty} \exp(i\lambda x) g(x)\,dx$ to denote the Fourier transform of any $g \in L^1 \cup L^2$.

By Parseval's formula and the fact that $|\phi_{f^{(k)}}(\lambda)|^2 = \lambda^{2k}|\phi_f(\lambda)|^2$, we can express θ_k as

$$(1.6) \qquad \theta_k = (2\pi)^{-1} \int_{-\infty}^{\infty} |\phi_{f^{(k)}}(\lambda)|^2\,d\lambda = (2\pi)^{-1} \int_{-\infty}^{\infty} \lambda^{2k} |\phi_f(\lambda)|^2\,d\lambda.$$

Indeed, the set of conditions

$$\{f^{(k)} \in L^1 \cap L^2, \text{ and } f^{(j)} \in L^1 \text{ is absolutely continuous for all } j = 0, 1, \dots, k-1\}$$

is sufficient for (1.6) to hold [cf. Hewitt and Stromberg (1969), pages 414–415]. Let $\hat{f}_\Lambda$ denote the sync-kernel estimate of f with bandwidth Λ^{-1}. On replacing $\phi_f(\lambda)$ in (1.6) by the estimate $\phi_{\hat{f}_\Lambda}(\lambda) = \tilde{\phi}(\lambda) I_{[-\Lambda,\Lambda]}(\lambda)$, where

$$(1.7) \qquad \tilde{\phi}(\lambda) = n^{-1} \sum_{j=1}^{n} \exp(i\lambda X_j), \qquad -\infty < \lambda < \infty,$$

is the sample characteristic function, we obtain the family of estimates

$$(1.8) \qquad \tilde{\theta}_k(\Lambda) = (2\pi)^{-1} \int_{-\Lambda}^{\Lambda} \lambda^{2k} |\tilde{\phi}(\lambda)|^2\,d\lambda, \qquad \Lambda > 0.$$

The performance of the estimates depends crucially on how well the cutoff frequency Λ (or, equivalently, the bandwidth Λ^{-1}) can be selected.

The notion of smoothness of f can be expressed in terms of the decay rate of $|\phi_f(\lambda)|$. Throughout the paper we assume the mild conditions given in Condition (A).

CONDITION (A) Relation (1.6) holds and, for some $0 < p < \infty$, $|\lambda|^p|\phi_f(\lambda)| = O(1)$ as $|\lambda| \to \infty$.

According to the earlier discussions, a proper cutoff Λ under conditions analogous to (1.5) should satisfy a criterion analogous to (1.4) (with $h = \Lambda^{-1}$). An important implication here is that the choice of the cutoff Λ should depend on k. It turns out that the adaptive $\hat{\Lambda}_k$ [see (2.1)] we propose does meet the criterion [see (2.10)]. When $p > 2k + 1$, (2.10) is analogous to (1.4) with $c = 2$ (see also Remark 2.2 and Lemma 2.2). Moreover, under Condition (A) the estimate $\tilde{\theta}_k(\hat{\Lambda}_k)$ is consistent for θ_k if $p > k + 2^{-1}$, and is asymptotically optimal if $p > 2k + 1$ (see Theorem 2.1). Note that the condition $p > 2k + 1$ is analogous to (1.5) but it is neither weaker nor stronger than (1.5). For example, the Gamma$(2k + 1 + \delta, \beta)$ density, $0 < \delta < 4^{-1}$, $\beta > 0$, satisfies the condition $p > 2k + 1$ but not (1.5) since the order of smoothness of such density is only $2k + \delta$. Thus, comparing with all the estimates discussed earlier, the proposed adaptive estimate $\tilde{\theta}_k(\hat{\Lambda}_k)$ has equally good theoretical properties.

The source of variation, for estimates using $\tilde{\phi}(\lambda)$, has been clearly pointed out in the important works of Chiu (1991a, 1992). In fact, by noting that

$$E\{|\tilde{\phi}(\lambda)|^2 - n^{-1}\} = n^{-1}(n-1)|\phi_f(\lambda)|^2 \tag{1.9}$$

and that $\tilde{\phi}(\lambda)$ at high frequency is dominated by sample variation and does not contain much information about f, Chiu (1991a) suggested removing the effect from $\tilde{\phi}(\lambda)$ at high frequency and proposed the family of estimates

$$\hat{\theta}_2(\Lambda) = (2\pi)^{-1}\int_{-\Lambda}^{\Lambda} \lambda^4\{|\tilde{\phi}(\lambda)|^2 - n^{-1}\}\,d\lambda, \quad \Lambda > 0, \tag{1.10}$$

to estimate θ_2. The difference between the estimates (1.8) (with $k = 2$) and (1.10) is that the latter does not include the nonstochastic diagonal "$j = l$" terms (bias) in the expansion $|\tilde{\phi}(\lambda)|^2 = n^{-2}\Sigma_j\Sigma_l \exp(i(X_j - X_l)\lambda)$, whereas the former does. A practical advantage of using (1.8) is that it is always positive; whereas (1.10) can be negative for some Λ and to ensure its positivity, one has to choose small Λ which may cause underestimation problems if $|\phi_f(\lambda)|$ does not decay nicely (this will be made clear below). Indeed, the estimates (1.8), which we focus on in this paper, is the sync-kernel, Fourier-domain version of the "diagonals-in" estimates of θ_k proposed by Jones and Sheather (1991), and the readers are referred to their work for further insights on choosing "diagonals-in" over "diagonals-out."

In the context of kernel density estimation, Chiu (1991a) suggested selecting the cutoff frequency, denoted by $\tilde{\Lambda}_c$ here, as the first (smallest positive) λ such that $|\tilde{\phi}(\lambda)|^2 = c/n$ for some constant $c > 1$ [this ensures the positivity of (1.10)]. However, this procedure works only when $|\phi_f(\lambda)|$ never vanishes and decays nicely [see Chiu (1991a), Assumptions 1 and 2, which are much stronger than our Condition (A)]. In order to overcome the difficulty, Chiu (1992) suggested selecting the cutoff frequency as the minimizer of the smoothed cross-validation score [SCV, which is a generalization of the least squares cross-validation and was termed by Hall, Marron and Park (1992)],

$$2\pi \,\mathrm{SCV}_n^{\infty}(\Lambda) = 4\Lambda n^{-1} - \int_{-\Lambda}^{\Lambda} |\tilde{\phi}(\lambda)|^2 \, d\lambda, \qquad \Lambda > 0; \tag{1.11}$$

he also proposed a modification for reducing the chance of selecting a large cutoff frequency. However, this procedure works well only when $\phi_f(\lambda)$ has unbounded support and decays in a regular way [see Theorem 2 of Chiu (1992)]. The $\hat{\Lambda}_k$ we propose here is a modification as well as an adaptation of that procedure. In this paper all the results are derived under Condition (A) only. This condition is much weaker than those in Chiu (1992).

Recent results on nonparametric estimation of θ_k based on observations from a density supported in [0, 1] are given by Fan (1991), who dealt with a white-noise model, and by Goldstein and Messer (1992) [their Example 2 provides discussions on the overlap of their approach with those in Hall and Marron (1987) and Bickel and Ritov (1988)]. Additional results on nonparametric estimation of θ_0 and related functionals are given by Schweder (1975), Hasminskii and Ibragimov (1979) and van Es (1992).

In the beginning of Section 2, we describe a general procedure of selecting the cutoff $\hat{\Lambda}_k$ along with the proposed adaptive estimate. The rest of Section 2 is divided into three subsections. In Section 2.1 we explain the rationale in detail. Lemma 2.1 gives an asymptotic representation of the MSE of the proposed estimate. Remark 2.3 indicates the connection of our approach with the one in Jones and Sheather (1991). Section 2.2 contains the main theoretical result (see Theorem 2.1). In Section 2.3, modification for improving the finite-sample performance of the general procedure is considered. Two different rules are suggested here; one for $k = 0$ and the other for $k \geq 1$. This is mainly due to the concern about the bias–variance trade-off. Bias reduction appears more effective for $k = 0$, while variance reduction is more effective for $k \geq 1$. In Section 3, extensive simulation studies are carried out and the superior performance of the proposed procedures are clearly demonstrated. Section 4 is devoted to proofs.

2. The proposed procedure and the theoretical results.

The proposed cutoff frequency $\hat{\Lambda}_k$ is the minimizer of the smoothed cross-validation score

$$\begin{aligned}\mathrm{SCV}_k(\Lambda) &= 4\Lambda^{2k+1}(n+1)^{-1}(2k+1)^{-1} \\ &\quad - \int_{-\Lambda}^{\Lambda} \lambda^{2k} |\tilde{\phi}(\lambda)|^2 \, d\lambda, \qquad \Lambda > 0,\end{aligned} \tag{2.1}$$

and the proposed estimate of θ_k is [see (1.8)]

$$\tilde{\theta}_k(\hat{\Lambda}_k) = (2\pi)^{-1}\int_{|\lambda|\le\hat{\Lambda}_k} \lambda^{2k}|\tilde{\phi}(\lambda)|^2\, d\lambda. \tag{2.2}$$

Note that, when $k = 0$, $\hat{\Lambda}_k$ and $\mathrm{SCV}_k(\Lambda)$ are asymptotically equivalent to those proposed in Chiu (1992) [see (1.11) above].

2.1. *The rationale of the proposed procedures.* We now explain the rationale of the above scheme. At first, one might attempt to find an estimate which is unbiased, up to a constant shift, for the risk $\mathrm{MSE}_k(\Lambda) = E\{\tilde{\theta}_k(\Lambda) - \theta_k\}^2$ at every n and Λ, and then select the cutoff frequency adaptively by minimizing the estimated risk. This scheme, unfortunately, is not feasible (see Remark 2.1). The alternatives, which we shall consider here, are (i) minimizing an unbiased estimate of a sharp upper bound on a suitable risk and (ii) minimizing an unbiased estimate of a quantity whose minimization is asymptotically (as $n \to \infty$) equivalent to the minimization of $\mathrm{MSE}_k(\Lambda)$. They both lead to the minimization of the same score (2.1), as explained below.

For alternative (i) we set $r_k(\Lambda) = E[\{\tilde{\theta}_k(\Lambda)\}^{1/2} - \theta_k^{1/2}]^2$, the mean squared error of $\{\tilde{\theta}_k(\Lambda)\}^{1/2}$ for estimating $\theta_k^{1/2}$, and consider a sharp upper bound on this risk (this will be made precise later). Let us write $\|\cdot(\lambda)\|_2 = \{\int_{-\infty}^{\infty}|\cdot(\lambda)|^2\, d\lambda\}^{1/2}$. By (1.6), (1.8) and Minkowski's inequality, we have for every $\Lambda > 0$ that

$$\begin{aligned} r_k(\Lambda) &\le (2\pi)^{-1}E\left\|\lambda^k\left\{\tilde{\phi}(\lambda)I_{[-\Lambda,\Lambda]}(\lambda) - \phi_f(\lambda)\right\}\right\|_2^2 \\ &= (2\pi)^{-1}\int_{-\Lambda}^{\Lambda}\lambda^{2k}\left\{E|\tilde{\phi}(\lambda) - \phi_f(\lambda)|^2 - |\phi_f(\lambda)|^2\right\}d\lambda + \theta_k \\ &= (2\pi n)^{-1}\int_{-\Lambda}^{\Lambda}\lambda^{2k}\left\{1 - (n+1)|\phi_f(\lambda)|^2\right\}d\lambda + \theta_k, \end{aligned} \tag{2.3}$$

where the last equality follows from the fact that

$$\begin{aligned} E|\tilde{\phi}(\lambda) - \phi_f(\lambda)|^2 &= \mathrm{cum}\big(\tilde{\phi}(\lambda), \tilde{\phi}(-\lambda)\big) \\ &= \mathrm{Var}\left\{\tilde{\phi}(\lambda)\right\} = n^{-1}\left\{1 - |\phi_f(\lambda)|^2\right\} \end{aligned} \tag{2.4}$$

(see Lemma 4.1). Let us denote the last line of (2.3) by $\mathrm{MISE}_k(\Lambda)$, as a sharp upper bound of $r_k(\Lambda)$. Let Λ_k denote the minimizer of

$$\begin{aligned} Q_k(\Lambda) = {}& 2\Lambda^{2k+1}(n+1)^{-1}(2k+1)^{-1} \\ & -\int_{-\Lambda}^{\Lambda}\lambda^{2k}|\phi_f(\lambda)|^2\, d\lambda, \qquad \Lambda > 0. \end{aligned} \tag{2.5}$$

Then Λ_k is also the minimizer of $\mathrm{MISE}_k(\Lambda)$, $\Lambda > 0$, since $\mathrm{MISE}_k(\Lambda) = (2\pi n)^{-1}(n+1)Q_k(\Lambda) + \theta_k$ and the second term is independent of Λ. By (1.9) and (2.1), we see that $n(n-1)^{-1}\,\mathrm{SCV}_k(\Lambda)$ is an unbiased estimate of $Q_k(\Lambda)$. So $\hat{\Lambda}_k$, the minimizer of $n(n-1)^{-1}\,\mathrm{SCV}_k(\Lambda)$, may be expected to be reasonably close to Λ_k.

For alternative (ii) we analyze the mean squared error $\mathrm{MSE}_k(\Lambda)$. It can be decomposed as $\mathrm{MSE}_k(\Lambda) = \mathrm{Var}\{\tilde{\theta}_k(\Lambda)\} + B_k^2(\Lambda)$, where $B_k(\Lambda) = E\tilde{\theta}_k(\Lambda) - \theta_k$ denotes the bias.

LEMMA 2.1. *Assume Condition* (A) *with* $p > 2k + 1$. *Then, for every* n *and* $\Lambda > 0$,

$$2\pi B_k(\Lambda) = D_k(\Lambda) - h_{k2}(\Lambda) \tag{2.6}$$

and

$$|n \operatorname{Var} \tilde{\theta}_k(\Lambda) - 4 \operatorname{Var} f^{(2k)}(X_1)| \le M_k\{n^{-1} + n^{-1}\Lambda^{4k+1} + h_{k1}(\Lambda)\}, \tag{2.7}$$

where $M_k > 0$ *is a constant not depending on* n *and* Λ,

$$D_k(\Lambda) = 2n^{-1}(2k+1)^{-1}\Lambda^{2k+1} - n^{-1}\int_{-\Lambda}^{\Lambda} \lambda^{2k}|\phi_f(\lambda)|^2\, d\lambda,$$

and $h_{kj}(\Lambda) = \int_{|\lambda| > \Lambda} \lambda^{2k}|\phi_f(\lambda)|^j\, d\lambda$, *for* $j = 1, 2$.

Let Λ_b denote the least upper bound of the support of $|\phi_f|$, and consider the following modification of $\mathrm{MSE}_k(\Lambda)$:

$$\begin{aligned} \mathrm{MSE}_k^*(\Lambda) &= 4n^{-1} \operatorname{Var} f^{(2k)}(X_1) \\ &\quad + (2\pi)^{-2}\{D_k(\Lambda) + h_{k2}(\Lambda)\}^2, \qquad \Lambda > 0. \end{aligned} \tag{2.8}$$

Note that $\mathrm{MSE}_k^*(\Lambda)$ is greater than or equal to the information bound (1.2) for all $\Lambda > 0$. Evidently, Lemma 2.1 implies that, as $n \to \infty$,

$$\mathrm{MSE}_k(\Lambda) \sim \{\mathrm{MSE}_k^*(\Lambda) - \pi^{-2}D_k(\Lambda)h_{k2}(\Lambda)\} \tag{2.9}$$

uniformly in $\Lambda \in J_{nk} = \{\Lambda(n) \le \Lambda \le o(n^{1/(4k+1)})\}$, where $\{\Lambda(n)\}$ is any sequence satisfying $\liminf_{n\to\infty} \Lambda(n) = \Lambda_b$. Since both $D_k(\Lambda)$ and $h_{k2}(\Lambda)$ are nonnegative for all $\Lambda \ge 0$, $\mathrm{MSE}_k^*(\Lambda)$ can be viewed as an asymptotic upper bound on $\mathrm{MSE}_k(\Lambda)$ over J_{nk}. The asymptotic difference $\pi^{-2}D_k(\Lambda)h_{k2}(\Lambda)$ has order $o(n^{-1})$ uniformly on any interval J_{nk}^* of the form

$$J_{nk}^* = \{\Lambda^*(n) \le \Lambda \le o(n^{1/(4k+2)})\}, \tag{2.10}$$

with $\{\Lambda^*(n)\}$ being a sequence satisfying $h_{k2}(\Lambda^*(n)) = o(n^{-1/2})$. Note that $\{J_{nk}^*\} \subset \{J_{nk}\}$. It follows that, as $n \to \infty$,

$$\mathrm{MSE}_k(\Lambda) \sim \mathrm{MSE}_k^*(\Lambda) \sim 4n^{-1} \operatorname{Var} f^{(2k)}(X_1) \tag{2.11}$$

uniformly in $\Lambda \in J_{nk}^*$. This and (1.2) together indicate that the minimization of $\mathrm{MSE}_k(\Lambda)$, $\Lambda > 0$, can be asymptotically achieved through the minimization of $\mathrm{MSE}_k^*(\Lambda)$, $\Lambda > 0$, if the minimizer of the latter is contained in some J_{nk}^*. By (2.8) and noting that $D_k(\Lambda) + h_{k2}(\Lambda) = (2\pi)\mathrm{MISE}_k(\Lambda)$, we see that Λ_k [cf. (2.5)] is the minimizer of $\mathrm{MSE}_k^*(\Lambda)$, $\Lambda > 0$. Now $Q_k'(\Lambda) = 0$ implies $|\phi_f(\Lambda)|^2 = (n+1)^{-1}$ and so, under Condition (A) with $p > k + 2^{-1}$, there exists a positive constant M_0 such that the inequality

$$\Lambda \le (M_0 n)^{1/(2p)}, \qquad n \ge 1, \tag{2.12}$$

holds for every critical value Λ. In particular, it holds for Λ_k. Combining this with the inequality $Q_k(\Lambda_k) \le Q_k(b_n)$, where $b_n = (M_0 n)^{1/(2p)}$, yields

$$(2.13)\quad h_{k2}(\Lambda_k) - h_{k2}(b_n) \le (n+1)^{-1}(2k+1)^{-1}\left(b_n^{2k+1} - \Lambda_k^{2k+1}\right).$$

From (2.12), (2.13) and the fact that $h_{k2}(b_n) = O(b_n^{-2p+2k+1})$, we conclude that

$$(2.14)\qquad h_{k2}(\Lambda_k) = O(n^{-1+[(2k+1)/(2p)]}) \quad \text{and} \quad \Lambda_k = O(n^{1/(2p)})$$

if $p > k + 2^{-1}$, and Λ_k is contained in some J_{nk}^* [see (2.10)] if $p > 2k+1$. The preceding discussions, together with those immediately below (2.5), indicate that asymptotically $\hat{\Lambda}_k$, Λ_k and Λ_{opt} [the minimizer of $\text{MSE}_k(\Lambda)$, $\Lambda > 0$] may be expected to be close to one another (i.e., they are of the same order in probability).

REMARK 2.1. Equation (2.6) indicates that $B_k^2(\Lambda)$, and hence $\text{MSE}_k(\Lambda)$, cannot be unbiasedly (up to a constant shift) estimated. This can be seen by noting that $h_{k2}(\Lambda) = 2\pi\theta_k - \int_{|\lambda| \le \Lambda} \lambda^{2k} |\phi_f(\lambda)|^2 \, d\lambda$, and θ_k is not estimable.

REMARK 2.2. From (1.2), (2.10) and (2.11), we see that $\tilde{\theta}_k(\Lambda)$ is asymptotically efficient if Λ is contained in some J_{nk}^* for each n.

REMARK 2.3. Set $\text{MSE}_k^{**}(\Lambda) = 4n^{-1}\,\text{Var}\, f^{(2k)}(X_1) + B_k^2(\Lambda)$ [equals the right-hand side of (2.9)]. Then, as $n \to \infty$,

$$(2.15)\qquad \text{MSE}_k^{**}(\Lambda) \sim \text{MSE}_k^{*}(\Lambda)$$

uniformly in $\Lambda \in J_{nk}^*$. Let Λ_0 denote the minimizer of $B_k^2(\Lambda)$ as well as $\text{MSE}_k^{**}(\Lambda)$, $\Lambda > 0$. Then, upon noting that the bias $B_k(\Lambda)$ is strictly increasing on $(0, \infty)$, we see that Λ_0 is the unique solution of the equation $B_k(\Lambda) = 0$ [i.e., $D_k(\Lambda) = h_{k2}(\Lambda)$], $\Lambda > 0$. Note that Λ_0^{-1} is the sync-kernel, Fourier-domain version of the solution (denoted by α_*) in Jones and Sheather (1991) and involves functionals which are not estimable nonparametrically. However, we can easily show that (2.14) remains true if Λ_k is replaced by Λ_0 throughout, and Λ_0 is contained in some J_{nk}^* if $p > 2k + 1$. This, together with (2.15) and the arguments immediately below (2.14), indicates that asymptotically $\hat{\Lambda}_k$, Λ_0, Λ_k and Λ_{opt} may be expected to be close to one another (i.e., they are of the same order in probability).

2.2. *The main theoretical results.* Remark 2.2 suggests that a proper adaptive cutoff frequency must be asymptotically (as $n \to \infty$) contained in the region of the form (2.10). The next lemma indicates that the proposed $\hat{\Lambda}_k$ indeed meets the criterion.

LEMMA 2.2. *Assume Condition* (A) *with* $p > k + 2^{-1}$. *Then*

$$(2.16)\qquad \hat{\Lambda}_k = O_p(n^{1/(2p)})$$

and

$$(2.17)\qquad h_{k2}(\hat{\Lambda}_k) = O_p(n^{-1+\{(2k+1)/(2p)\}}),$$

where the function h_{k2} *is defined in Lemma* 2.1.

By comparing this lemma with (2.14), we see that $\hat{\Lambda}_k$ and Λ_k have the same order, and they both are contained in region (2.10) if $p > 2k + 1$.

The main results concerning the asymptotic (as $n \to \infty$) properties of the estimate (2.2) are contained in the next theorem, and equation (2.16) is the key to proving them.

THEOREM 2.1. *Assume Condition* (A) *with* $p > k + 2^{-1}$. *Then the following two assertions hold*:

$$\text{(i)} \qquad \tilde{\theta}_k(\hat{\Lambda}_k) - \theta_k = \begin{cases} O_p(n^{-1+[(2k+1)/(2p)]}), & \text{if } p < 2k+1, \\ O_p(n^{-1/2}\log n), & \text{if } p = 2k+1. \end{cases}$$

(ii) *If* $p > 2k + 1$, *then*

$$\tilde{\theta}_k(\hat{\Lambda}_k) \to N(\theta_k, 4n^{-1}\operatorname{Var} f^{(2k)}(X_1)) \quad \textit{in law}.$$

As expected, the order p which governs the decay rate of the characteristic function for f is crucial. For large p, our estimators are $\sqrt{n}$-consistent, and they achieve the information bound. For smaller p, they are still consistent but with slower convergence rates.

2.3. *Two modifications of the proposed procedures.* This subsection is devoted to two modifications which improve the performance of the proposed estimate (2.2) in practice (while n is small to moderately large). One modification is for estimating θ_0; it reduces the bias of the estimate. The other modification is for estimating θ_k, $k \geq 1$; it reduces the chance of mistakenly selecting a too large cutoff frequency, and thus it reduces the sample variation of the estimate.

By Remark 2.3 and by the works of Jones and Sheather (1991) and Sheather and Jones (1991), we know that the minimizer of $B_k^2(\Lambda)$, which is the unique solution of $B_k(\Lambda) = 0$, plays an important role. Since $B_k(\Lambda)$ is not estimable (see Remark 2.1), we need to find a reasonable estimate $\hat{B}_k(\Lambda)$ and to solve $\hat{B}_k(\Lambda) = 0$. Using (2.6), the equation in Remark 2.1 and following the scale-model approach for f in Park and Marron (1990), we shall use

$$\text{(2.18)} \quad \hat{B}_k(\Lambda) = \pi^{-1}n^{-1}(2k+1)^{-1}\Lambda^{2k+1} + \tilde{\theta}_k(\Lambda) - \hat{\sigma}^{-(2k+1)}\theta_k(g)$$

to estimate $B_k(\Lambda)$, where $\tilde{\theta}_k(\Lambda)$ is given by (1.8), $\theta_k(g)$ is given by (1.1) with f being replaced by g throughout, g is a reference density with unit variance and $\hat{\sigma}$, the estimate of σ_f (the s.d. of f), equals min{sample s.d., (sample interquartile range)/1.349} [cf. Silverman (1986), page 47]. It is easy to see that $\hat{B}_k(\Lambda)$ is strictly increasing in $\Lambda > 0$, and the equation $\hat{B}_k(\Lambda) = 0$, $\Lambda > 0$, has a unique solution, say $\hat{\Lambda}_{k,g}$.

Both modifications involve using $\hat{\Lambda}_{k,g}$ at some stage, as a reference frequency, to adjust the magnitude of the selected cutoff frequency. Since g is subjectively chosen, the modified procedures are no longer completely data-driven.

The first modified estimate is $\tilde{\theta}_k(\hat{\Lambda}_{k,M})$, where $\hat{\Lambda}_{k,M} = \max\{\hat{\Lambda}_k, \hat{\Lambda}_{k,g}\}$, with g being the standardized beta$(k+2, k+2)$ density. This modification is recommended in practice only for estimating θ_0 (the reason for not recommending it for $k \geq 1$ will be made clear in Section 3). Evidently, $\tilde{\theta}_k(\hat{\Lambda}_{k,M}) \geq \tilde{\theta}_k(\hat{\Lambda}_k)$. Hence, when $k = 0$, the probable underestimation by using $\tilde{\theta}_k(\hat{\Lambda}_k)$, as implied by the simulation results of Chiu (1992), can be improved by using $\tilde{\theta}_k(\hat{\Lambda}_{k,M})$. By (1.9), (2.18) and the fact $\hat{\sigma} \approx \sigma_f$ asymptotically, we see that $\hat{B}_k(\Lambda) = 0$ implies, in a rough sense, that asymptotically,

$$(2.19)\quad \begin{aligned} n^{-1}(2k+1)^{-1}\Lambda^{2k+1} &\approx 2^{-1}h_{k2}(\Lambda) - \pi\left[\theta_k(f) - \sigma_f^{-(2k+1)}\theta_k(g)\right] \\ &\leq 2^{-1}h_{k2}(\Lambda), \end{aligned}$$

where h_{k2} is defined in Lemma 2.1 and the last inequality is by Terrell (1990), who proves that, for our present choice of g, the expression within the brackets in (2.19) is nonnegative for all f. It follows that the solution to (2.19) is of order $O(n^{1/(2p)})$ under Condition (A) with $p > k + 2^{-1}$. The preceding heuristic arguments justify that the conclusions of Lemma 2.2 and Theorem 2.1 hold for $\hat{\Lambda}_{k,M}$ and $\tilde{\theta}_k(\hat{\Lambda}_{k,M})$, respectively. A formal proof can be given, but is omitted.

The second modification is recommended only for estimating θ_k, $k \geq 1$. It is mainly an adaptation and a refinement of the approach in Chiu (1992). It also relates to the approach in Sheather and Jones (1991) through using $\hat{\Lambda}_{k,g}$. The basic idea here is to use some suitably chosen $\hat{\Lambda}_{\text{mod}}$, say, as the cutoff frequency unless $\tilde{\phi}(\lambda)$ at higher frequency contains significant information about f. The issue of how to choose $\hat{\Lambda}_{\text{mod}}$ will be addressed later. We now describe how to modify the SCV score (2.1) beyond $\hat{\Lambda}_{\text{mod}}$. It can be shown that, for any fixed n,

$$(2.20)\quad \operatorname{Var}(\mathrm{SCV}_k(\lambda) - \mathrm{SCV}_k(\mu)) \sim Q_{k1}(\mu, \lambda, \phi_f) + Q_{k2}(\mu, \lambda, \phi_f)$$

as $\mu \to \infty$ and $\lambda - \mu \to \infty$ [see (4.5), (4.12), (4.14) and (4.15)], where

$$Q_{k1}(\mu, \lambda, \phi_f) = 4n^{-2}\int_\mu^\lambda\int_\mu^\lambda \lambda_1^{2k}\lambda_2^{2k}\left\{|\phi_f(\lambda_1+\lambda_2)|^2 + |\phi_f(\lambda_2-\lambda_1)|^2\right\}d\lambda_1\,d\lambda_2,$$

$$Q_{k2}(\mu, \lambda, \phi_f) = 4(n-3)n^{-2}$$

$$\times\int_J\int_J \lambda_1^{2k}\lambda_2^{2k}\{\phi_f(-\lambda_1)\phi_f(-\lambda_2)\phi_f(\lambda_1+\lambda_2)\}\,d\lambda_1\,d\lambda_2$$

with $J = [-\lambda, -\mu] \cup [\mu, \lambda]$. Therefore, the left-hand side of (2.20) is estimated by $\hat{V}_k(\mu, \lambda) = Q_{k1}(\mu, \lambda, \tilde{\phi}) + Q_{k2}(\mu, \lambda, \tilde{\phi})$. Let us write

$$(2.21)\quad \mathrm{SCV}_k^*(\mu, \Lambda) = \begin{cases} \mathrm{SCV}_k(\Lambda), & \text{if } 0 < \Lambda \leq \mu, \\ \mathrm{SCV}_k(\Lambda) + z(\alpha_k)\left\{\hat{V}_k(\mu, \Lambda)\right\}^{1/2}, & \text{if } \Lambda > \mu, \end{cases}$$

where $z(\alpha_k)$ is the upper α_k point of the standard normal distribution, and $\alpha_1 = 0.09$, $\alpha_2 = 0.06$, $\alpha_3 = 0.035$ and $\alpha_k \leq 0.035$ for $k \geq 4$. The exact choice of α_k depends on the user's objective. The proposed modified estimate is $\tilde{\theta}_k(\hat{\Lambda}_k^*)$ where $\hat{\Lambda}_k^*$ is the minimizer of the modified score $\mathrm{SCV}_k^*(\hat{\Lambda}_{\mathrm{mod}}, \Lambda)$, $\Lambda > 0$.

It is left to describe how to select $\hat{\Lambda}_{\mathrm{mod}}$. For the rest, the g in $\hat{\Lambda}_{k,g}$ is set to be the standard normal density. Let $\hat{\Lambda}_{\mathrm{loc}}$ denote the first local minimizer of (2.1). It plays a pivotal role [cf. Hall and Marron (1991b)]. The basic idea is letting $\hat{\Lambda}_{\mathrm{mod}}$ equal $\hat{\Lambda}_{\mathrm{loc}}$ unless $\hat{\Lambda}_{\mathrm{loc}}$ is too large; and, in that case, letting $\hat{\Lambda}_{\mathrm{mod}}$ equal $\hat{\Lambda}_{k,g}$. Specifically, let $\hat{\Lambda}_{k,g}^*$ denote the minimizer of $\mathrm{SCV}_k^*(\hat{\Lambda}_{k,g}, \Lambda)$, $\Lambda > 0$ and define

$$(2.22) \qquad \hat{\Lambda}_{\mathrm{mod}} = \begin{cases} \hat{\Lambda}_{k,g}, & \text{if } \hat{\Lambda}_{k,g} < \hat{\Lambda}_{\mathrm{loc}} \text{ and} \\ & \qquad \hat{\Lambda}_{k,g} = \hat{\Lambda}_{k,g}^* \text{ (if } E_1 \text{ and } E_2\text{, say)}, \\ \hat{\Lambda}_{\mathrm{loc}}, & \text{otherwise.} \end{cases}$$

Here E_1 implies $\hat{\Lambda}_{\mathrm{mod}} \le \hat{\Lambda}_{\mathrm{loc}}$, and E_2 indicates that there is no need to look beyond $\hat{\Lambda}_{k,g}$ since $\tilde{\phi}(\lambda)$ does not contain much information about f beyond it. This explains (2.22).

The estimate $\tilde{\theta}_k(\hat{\Lambda}_k^*)$ has smaller sample variation than the estimate $\tilde{\theta}_k(\hat{\Lambda}_k)$ due to

$$(2.23) \qquad \hat{\Lambda}_{\mathrm{mod}} \le \hat{\Lambda}_k^* \le \hat{\Lambda}_k \quad \text{and} \quad \tilde{\theta}_k\big(\hat{\Lambda}_{\mathrm{mod}}\big) \le \tilde{\theta}_k\big(\hat{\Lambda}_k^*\big) \le \tilde{\theta}_k\big(\hat{\Lambda}_k\big).$$

Note that $\tilde{\theta}_k(\hat{\Lambda}_k^*) = \tilde{\theta}_k(\hat{\Lambda}_{k,g}) < \tilde{\theta}_k(\hat{\Lambda}_{\mathrm{loc}})$ when $\hat{\Lambda}_{\mathrm{mod}} = \hat{\Lambda}_{k,g}$. This can happen in practice, especially at small n, but cannot happen as $n \to \infty$. Since, for any $\Lambda > 0$, the difference between $\mathrm{SCV}_k^*(\hat{\Lambda}_{k,g},\Lambda)$ and $\mathrm{SCV}_k(\Lambda)$ is asymptotically negligible, the event $E_1 \cap E_2$ is asymptotically null. Consequently, $\hat{\Lambda}_{\mathrm{mod}}$ and $\hat{\Lambda}_{\mathrm{loc}}$ are asymptotically indistinguishable. This and (2.23) together imply that $\hat{\Lambda}_k^*$ and $\tilde{\theta}_k(\hat{\Lambda}_k^*)$ are asymptotically bounded below by $\hat{\Lambda}_{\mathrm{loc}}$ and $\tilde{\theta}_k(\hat{\Lambda}_{\mathrm{loc}})$, respectively. Next, by the fact that $\mathrm{SCV}_k'(\Lambda) = 2\Lambda^{2k}[(2/(n+1)) - |\tilde{\phi}(\Lambda)|^2]$, we see that $\hat{\Lambda}_{\mathrm{loc}}$ is equal to the first Λ such that $|\tilde{\phi}(\Lambda)|^2 = 2/(n+1)$. It follows that $\hat{\Lambda}_{\mathrm{loc}} = \tilde{\Lambda}_{2n/(n+1)}$ [recall that $\tilde{\Lambda}_c$, the cutoff frequency used by Chiu (1991a), is the first λ such that $|\tilde{\phi}(\lambda)|^2 = c/n$, $c > 1$]. Therefore, if Condition (A) is fulfilled and, in addition, $|\phi_f(\lambda)| \sim C|\lambda|^{-p}$ (algebraic decay) or $|\phi_f(\lambda)| \sim C\exp(-q|\lambda|^r)$ (exponential decay) as $\lambda \to \infty$, where $C > 0$, $q > 0$ and $1 \le r \le 2$ are constants, then, following the proofs in Chiu (1991a), we can show that the conclusions of Lemma 2.2 hold for $\hat{\Lambda}_{\mathrm{mod}}$ and, consequently, hold for $\hat{\Lambda}_k^*$; and thus the conclusions of Theorem 2.1 hold for $\tilde{\theta}_k(\hat{\Lambda}_k^*)$.

Finally, we note that, when $k = 0$, the term $Q_{k1}(\mu, \lambda, \phi_f)$ in (2.20) is asymptotically equivalent to, but more precise than, the one in Chiu (1992). The extra term $Q_{k2}(\mu, \lambda, \phi_f)$ that we have introduced has much larger order than the former term, and so it cannot be ignored, especially at small n.

3. Simulation results. We have carried out extensive simulation studies. The complete report of the simulation is available from the author upon request. Throughout this section we use $\tilde{\theta}_k$, $\tilde{\theta}_{k,M}$ and $\tilde{\theta}_k^*$ to denote the proposed estimates $\hat{\theta}_k(\hat{\Lambda}_k)$, $\tilde{\theta}_k(\hat{\Lambda}_{k,M})$ and $\tilde{\theta}_k(\hat{\Lambda}_k^*)$ of Section 2, respectively.

For estimating θ_k, $k = 0$, we compare our estimates $\tilde{\theta}_k$ and $\tilde{\theta}_{k,M}$ with $\hat{\theta}_C$ and $\hat{\theta}_S$, where $\hat{\theta}_C$ denotes Chiu's (1991a) estimate $\hat{\theta}_k(\tilde{\Lambda}_c)$ with $c = 3$ [here $\hat{\theta}_k(\Lambda)$ is defined by (1.10) with λ^4 being replaced by λ^{2k}], and $\hat{\theta}_S$ denotes the estimate of Sheather and Jones (1991) by using the normal density as both the kernel and the reference density. For estimating θ_k, $1 \le k \le 3$, we compare our modified estimate $\tilde{\theta}_k^*$ with $\hat{\theta}_C$, $\hat{\theta}_S$ and $\hat{\theta}_H$, where $\hat{\theta}_H$ denotes the estimate of θ_2 of Hall, Sheather, Jones and Marron (1991). We use $\hat{\sigma}$ [see (2.18)] to estimate the scale parameter σ wherever needed in all these estimates. For $1 \le k \le 3$, the simulation results on the estimate $\tilde{\theta}_k$ are not reported here because sometimes they still show quite large sample variations even when $n = 1600$. We do not know how large an n is needed before the asymptotics take effect for $\tilde{\theta}_k$. However, the simulation results do suggest that n should substantially increase with k and, moreover, the estimates $\tilde{\theta}_0$ as well as $\tilde{\theta}_k^*$, $1 \le k \le 3$, being the substitute for $\tilde{\theta}_k$, perform excellently and so the asymptotics take effect at relatively small n (see Tables 1 and 2). A referee pointed out that in an unpublished work B. Aldershof has obtained some results relating to the important question of "when do the asymptotics take effect?" in general settings.

We generated 200 realizations of data sets of size $n = 100$, $n = 400$ and $n = 1600$ from each of eight normal mixture densities: (#1) Gaussian, $N(0,1)$; (#2) strongly skewed (resembles lognormal), $8^{-1}\sum_{j=0}^{7} N(3\{(\frac{2}{3})^j - 1\}, (\frac{2}{3})^{2j})$; (#3) kurtotic, $(\frac{2}{3})N(0,1) + (\frac{1}{3})N(0, \frac{1}{100})$; (#4) outlier, $0.1N(0,1) + 0.9N(0, \frac{1}{100})$; (#5) separated bimodal, $0.5N(0,1) + 0.5N(8,1)$; (#6) skewed bimodal, $0.75N(0,1) + 0.25N(2, \frac{1}{9})$; (#7) trimodal, $0.25N(-2, \frac{1}{16}) + 0.5N(0,1) + 0.25N(2, \frac{1}{16})$; (#8) claw (five-modal), $0.5N(0,1) + 0.1\sum_{j=0}^{4} N(j/2 - 1, \frac{1}{100})$. Densities (#1)–(#4) are unimodal, whereas (#5)–(#8) are multimodal [and $|\phi_f(\lambda)|^2$ has sidelobes]. These eight densities have been carefully chosen because they typify different types of challenges [see Marron and Wand (1992) for more details on these densities]. We remark that density (#6) was also considered in Scott and Terrell (1987). The random samples were generated by the function RAND in FORTRAN 77 on a Sun-Sparc workstation. For each sample we applied the fast Fourier transform to evaluate $\tilde{\phi}(\lambda)$. The implementation is the same as in Chiu (1991a, b).

For comparison of the above estimates (say, $\hat{\theta}_n$), we choose to compare $\text{MSE} = m^{-1}\sum_{j=1}^{m}(\hat{\theta}_n/\theta_k - 1)^2$ and $\text{MAE} = m^{-1}\sum_{j=1}^{m}|\hat{\theta}_n/\theta_k - 1|$, respectively, where $m = 200$.

Tables 1 and 2 summarize the simulation results. To save space, the case $n = 1600$ is not included here. Table 1 clearly indicates that $\tilde{\theta}_{0,M}$ and $\tilde{\theta}_0$ are conclusively the best two estimates of θ_0 for all the cases except when $n = 100$ and density is (#4). For that case all four estimates perform equally well. Table 2 indicates that, for the multimodal densities (#5)–(#8), $\tilde{\theta}_k^*$ is overwhelmingly the best estimate of θ_k for all the possible cases of k and n except when $n = 100$, $1 \le k \le 2$, and the density is (#6) or (#8). For these cases $\tilde{\theta}_k^*$ is comparable with other estimates (its MSE or MAE may be a little larger, but its bias is smaller than those of other estimates). For the unimodal

TABLE 1

Simulation result on estimating θ_0: sample mean, mean squared error and mean absolute error of the ratio $\hat{\theta}_n/\theta_0$ are given for $n = 100$ and 400 from each of eight different underlying densities (200 replications in each case); the value under each density is the true θ_0

		$n = 100$			$n = 400$		
Density	$\hat{\boldsymbol{\theta}}_n$	**MEAN**	**MSE**	**MAE**	**MEAN**	**MSE**	**MAE**
#1	$\hat{\theta}_C$	0.880	0.021	0.127	0.893	0.013	0.107
0.2821	$\tilde{\theta}_0$	0.923	0.015	0.107	0.904	0.011	0.098
	$\tilde{\theta}_{0,M}$	0.947	0.012	0.090	0.936	0.006	0.068
	$\hat{\theta}_S$	0.886	0.019	0.122	0.894	0.013	0.106
#2	$\hat{\theta}_C$	0.815	0.054	0.201	0.913	0.014	0.100
0.5569	$\tilde{\theta}_0$	1.000	0.042	0.144	0.974	0.008	0.071
	$\tilde{\theta}_{0,M}$	1.000	0.042	0.144	0.974	0.008	0.071
	$\hat{\theta}_S$	0.635	0.140	0.365	0.737	0.073	0.263
#3	$\hat{\theta}_C$	0.905	0.047	0.175	0.935	0.011	0.088
0.6152	$\tilde{\theta}_0$	1.053	0.042	0.158	0.971	0.008	0.073
	$\tilde{\theta}_{0,M}$	1.065	0.048	0.167	0.975	0.009	0.075
	$\hat{\theta}_S$	0.809	0.068	0.225	0.841	0.032	0.161
#4	$\hat{\theta}_C$	0.963	0.011	0.087	0.982	0.003	0.041
2.3592	$\tilde{\theta}_0$	1.009	0.012	0.087	0.996	0.003	0.040
	$\tilde{\theta}_{0,M}$	1.019	0.013	0.088	1.006	0.003	0.040
	$\hat{\theta}_S$	0.965	0.012	0.087	0.981	0.003	0.042
#5	$\hat{\theta}_C$	0.249	0.584	0.751	0.254	0.576	0.746
0.1410	$\tilde{\theta}_0$	0.824	0.043	0.187	0.795	0.044	0.205
	$\tilde{\theta}_{0,M}$	0.824	0.043	0.187	0.795	0.044	0.205
	$\hat{\theta}_S$	0.456	0.297	0.545	0.603	0.158	0.397
#6	$\hat{\theta}_C$	0.772	0.055	0.229	0.779	0.050	0.222
0.2350	$\tilde{\theta}_0$	0.912	0.018	0.112	0.879	0.016	0.121
	$\tilde{\theta}_{0,M}$	0.926	0.013	0.098	0.893	0.012	0.107
	$\hat{\theta}_S$	0.813	0.038	0.187	0.835	0.028	0.165
#7	$\hat{\theta}_C$	0.541	0.213	0.459	0.602	0.162	0.398
0.2410	$\tilde{\theta}_0$	0.961	0.015	0.104	0.897	0.013	0.105
	$\tilde{\theta}_{0,M}$	0.961	0.015	0.104	0.897	0.013	0.105
	$\hat{\theta}_S$	0.624	0.142	0.376	0.701	0.090	0.299
#8	$\hat{\theta}_C$	0.771	0.056	0.229	0.779	0.050	0.221
0.3702	$\tilde{\theta}_0$	0.989	0.030	0.144	0.953	0.005	0.060
	$\tilde{\theta}_{0,M}$	0.994	0.028	0.139	0.953	0.005	0.060
	$\hat{\theta}_S$	0.785	0.050	0.215	0.801	0.041	0.199

TABLE 2

Simulation results on estimating θ_k, $1 \le k \le 3$: sample mean, mean squared error and mean absolute error of the ratio $\hat{\theta}_n/\theta_k$ are given for $n = 100$ and 400 from each of eight different underlying densities (200 replications in each case); the top of each MSE column is the true θ_k; the zero entries in the mean columns are due to rounding

Density	$\hat{\theta}_n$	$k = 1$ MEAN	$k = 1$ MSE	$k = 1$ MAE	$k = 2$ MEAN	$k = 2$ MSE	$k = 2$ MAE	$k = 3$ MEAN	$k = 3$ MSE	$k = 3$ MAE
#1			0.1410			0.2116			0.5289	
$n = 100$	$\hat{\theta}_C$	1.017	0.128	0.249	1.191	2.323	0.646	1.214	4.884	0.968
	$\tilde{\theta}_k^*$	1.047	0.092	0.222	1.089	0.359	0.402	1.089	0.912	0.620
	$\hat{\theta}_S$	1.017	0.089	0.227	1.373	0.968	0.587	1.984	4.837	1.213
	$\hat{\theta}_H$				1.505	1.496	0.701			
$n = 400$	$\hat{\theta}_C$	1.025	0.024	0.118	1.088	0.289	0.288	1.137	3.919	0.561
	$\tilde{\theta}_k^*$	1.043	0.025	0.119	1.060	0.090	0.216	1.005	0.210	0.317
	$\hat{\theta}_S$	1.006	0.019	0.109	1.165	0.163	0.271	1.354	0.850	0.536
	$\hat{\theta}_H$				1.211	0.212	0.310			
#2			15.4191			4595.1899			3066960.0	
$n = 100$	$\hat{\theta}_C$	0.279	0.588	0.740	0.050	0.912	0.950	0.007	0.987	0.993
	$\tilde{\theta}_k^*$	0.471	0.607	0.678	0.149	0.946	0.932	0.028	0.976	0.987
	$\hat{\theta}_S$	0.034	0.933	0.966	0.001	0.999	0.999	0.000	1.000	1.000
	$\hat{\theta}_H$				0.001	0.999	1.000			
$n = 400$	$\hat{\theta}_C$	0.562	0.243	0.453	0.186	0.683	0.814	0.045	0.915	0.955
	$\tilde{\theta}_k^*$	0.772	0.162	0.338	0.342	0.527	0.689	0.109	0.816	0.893
	$\hat{\theta}_S$	0.057	0.890	0.943	0.001	0.998	0.999	0.000	1.000	1.000
	$\hat{\theta}_H$				0.001	0.998	0.999			
#3			15.9093			2351.3999			587701.0	
$n = 100$	$\hat{\theta}_C$	0.627	0.370	0.530	0.359	0.645	0.749	0.177	0.877	0.873
	$\tilde{\theta}_k^*$	0.888	0.430	0.511	0.604	0.839	0.744	0.339	1.421	0.933
	$\hat{\theta}_S$	0.188	0.688	0.813	0.037	0.930	0.963	0.006	0.988	0.994
	$\hat{\theta}_H$				0.042	0.924	0.958			
$n = 400$	$\hat{\theta}_C$	0.885	0.088	0.238	0.714	0.244	0.422	0.562	0.428	0.599
	$\tilde{\theta}_k^*$	1.034	0.108	0.246	0.987	0.430	0.431	0.917	0.958	0.683
	$\hat{\theta}_S$	0.217	0.621	0.783	0.034	0.933	0.966	0.004	0.992	0.996
	$\hat{\theta}_H$				0.027	0.947	0.973			
#4			114.321			17137.5			4284320.0	
$n = 100$	$\hat{\theta}_C$	0.954	0.116	0.259	0.954	0.930	0.522	1.068	8.081	0.922
	$\tilde{\theta}_k^*$	0.992	0.114	0.252	0.829	0.266	0.415	0.674	0.562	0.596
	$\hat{\theta}_S$	0.910	0.099	0.256	1.010	0.460	0.488	1.158	1.622	0.773
	$\hat{\theta}_H$				1.147	0.752	0.533			
$n = 400$	$\hat{\theta}_C$	0.981	0.019	0.110	0.952	0.099	0.228	0.919	0.524	0.405
	$\tilde{\theta}_k^*$	1.015	0.024	0.116	1.013	0.156	0.252	0.821	0.247	0.384
	$\hat{\theta}_S$	0.930	0.024	0.128	0.916	0.101	0.255	0.887	0.273	0.406
	$\hat{\theta}_H$				1.004	0.085	0.224			

TABLE 2
(*Continued*)

Density	$\hat{\theta}_n$	$k=1$ MEAN	$k=1$ MSE	$k=1$ MAE	$k=2$ MEAN	$k=2$ MSE	$k=2$ MAE	$k=3$ MEAN	$k=3$ MSE	$k=3$ MAE
#5			0.0705			0.1058			0.2644	
$n=100$	$\hat{\theta}_C$	0.069	0.911	0.943	0.039	0.966	0.980	0.006	0.992	0.994
	$\tilde{\theta}_k^*$	0.961	0.105	0.257	0.741	0.531	0.504	0.388	0.903	0.796
	$\hat{\theta}_S$	0.135	0.749	0.865	0.026	0.948	0.974	0.003	0.994	0.997
	$\hat{\theta}_H$				0.022	0.956	0.978			
$n=400$	$\hat{\theta}_C$	0.072	0.900	0.930	0.037	0.951	0.964	0.010	0.986	0.990
	$\tilde{\theta}_k^*$	1.010	0.021	0.100	0.931	0.078	0.187	0.796	0.425	0.430
	$\hat{\theta}_S$	0.251	0.561	0.749	0.054	0.894	0.946	0.007	0.987	0.993
	$\hat{\theta}_H$				0.052	0.898	0.948			
#6			0.2625			3.2251			72.6081	
$n=100$	$\hat{\theta}_C$	0.208	0.629	0.792	0.012	0.977	0.988	0.001	0.999	0.999
	$\tilde{\theta}_k^*$	0.379	0.612	0.750	0.082	0.980	0.973	0.006	0.992	0.994
	$\hat{\theta}_S$	0.304	0.488	0.696	0.045	0.912	0.955	0.006	0.989	0.994
	$\hat{\theta}_H$				0.036	0.930	0.964			
$n=400$	$\hat{\theta}_C$	0.220	0.618	0.781	0.021	0.966	0.979	0.007	0.990	0.993
	$\tilde{\theta}_k^*$	0.898	0.149	0.310	0.522	0.495	0.629	0.233	0.823	0.864
	$\hat{\theta}_S$	0.387	0.379	0.613	0.072	0.861	0.928	0.010	0.980	0.990
	$\hat{\theta}_H$				0.055	0.894	0.945			
#7			1.0870			26.9927			1083.49	
$n=100$	$\hat{\theta}_C$	0.026	0.951	0.975	0.001	0.997	0.999	0.000	1.000	1.000
	$\tilde{\theta}_k^*$	0.830	0.558	0.513	0.347	0.806	0.792	0.095	0.917	0.950
	$\hat{\theta}_S$	0.049	0.905	0.951	0.003	0.994	0.997	0.000	1.000	1.000
	$\hat{\theta}_H$				0.003	0.995	0.997			
$n=400$	$\hat{\theta}_C$	0.072	0.870	0.928	0.012	0.977	0.988	0.003	0.995	0.997
	$\tilde{\theta}_k^*$	0.994	0.062	0.193	0.748	0.177	0.355	0.405	0.463	0.634
	$\hat{\theta}_S$	0.089	0.831	0.911	0.007	0.986	0.993	0.000	0.999	1.000
	$\hat{\theta}_H$				0.003	0.994	0.997			
#8			6.9292			1149.8			255921.0	
$n=100$	$\hat{\theta}_C$	0.028	0.945	0.972	0.000	0.999	0.999	0.000	1.000	1.000
	$\tilde{\theta}_k^*$	0.351	1.080	1.018	0.245	1.018	0.994	0.113	0.941	0.946
	$\hat{\theta}_S$	0.031	0.939	0.969	0.001	0.999	1.000	0.000	1.000	1.000
	$\hat{\theta}_H$				0.001	0.999	1.000			
$n=400$	$\hat{\theta}_C$	0.028	0.944	0.972	0.000	0.999	1.000	0.000	1.000	1.000
	$\tilde{\theta}_k^*$	1.088	0.123	0.265	0.915	0.132	0.277	0.606	0.221	0.421
	$\hat{\theta}_S$	0.034	0.933	0.966	0.001	0.999	0.999	0.000	1.000	1.000
	$\hat{\theta}_H$				0.001	0.999	0.999			

densities (#1) and (#2), $\tilde{\theta}_k^*$ is apparently the best estimate of θ_k for all the possible cases of k and n except when $n = 400$, $k = 1$ and the density is (#1). For that case all three estimates perform well, but $\hat{\theta}_S$ is a little better than the others. For the unimodal density (#3), $\tilde{\theta}_k^*$ and $\hat{\theta}_C$ are evidently the best two estimates of θ_k for all the possible cases of k and n. The estimate $\hat{\theta}_C$ in general has smaller MSE or MAE, but always has much larger bias than that of $\tilde{\theta}_k^*$. Finally, for the unimodal density (#4), $\tilde{\theta}_k^*$ is the best estimate of θ_3 and also of θ_2 when $n = 100$. As for the remaining cases of k and n, all the estimates perform well and are comparable with one another. When $n = 100$ and $k = 1$, $\tilde{\theta}_k^*$ has the smallest MAE, whereas $\hat{\theta}_S$ has the smallest MSE. When $n = 400$, $\hat{\theta}_H$ is the best estimate of θ_2, and $\hat{\theta}_C$ performs very well for $1 \le k \le 2$.

In summary, the simulation results reveal that over a wide range of smooth density shapes and at practical sample sizes, the overall performances of the proposed estimates $\tilde{\theta}_0$, $\tilde{\theta}_{0,M}$ and $\tilde{\theta}_k^*$, $1 \le k \le 3$, are excellent and are much better than the overall performances of other estimates included in the study. Furthermore, Tables 1 and 2 also reveal that for smooth densities the convergence rates (as $n \to \infty$) of the proposed estimates $\tilde{\theta}_0$, $\tilde{\theta}_{0,M}$ and $\tilde{\theta}_k^*$, $1 \le k \le 3$, to their respective target values are very fast. This agrees well with the earlier theoretic results.

4. Proofs. Set

$$\tilde{\phi}_d(\lambda) = \tilde{\phi}(\lambda) - \phi_f(\lambda), \tag{4.1}$$

$$T_1(A) = \int_A \lambda^{2k} |\tilde{\phi}_d(\lambda)|^2 \, d\lambda, \tag{4.2}$$

$$T_2(A) = \int_A \lambda^{2k} \phi_f(-\lambda) \tilde{\phi}_d(\lambda) \, d\lambda, \tag{4.3}$$

$$T_3(A) = \int_A \lambda^{2k} \left\{ |\tilde{\phi}(\lambda)|^2 - |\phi_f(\lambda)|^2 \right\} d\lambda, \tag{4.4}$$

where A is any measurable set. Note that $E\tilde{\phi}_d(\lambda) = 0$ and

$$T_3(A) = T_1(A) + 2\,\mathrm{Re}(T_2(A)). \tag{4.5}$$

Observe that $\mathrm{Re}(T_2(A)) = T_2(A)$ if $A = -A$. The notation $T_j^0(A) = T_j(A) - ET_j(A)$ shall be used.

In the sequel we shall write ϕ for ϕ_f and suppress the subscript k in $\hat{\Lambda}_k$, $\tilde{\theta}_k$ and h_{kj}, $j = 1, 2$, whenever it causes no confusion.

LEMMA 4.1. *The following hold for any* λ_1 *and* λ_2:

$$\mathrm{cum}\left(\tilde{\phi}_d(\lambda_1), \tilde{\phi}_d(\lambda_2)\right) = n^{-1}\{\phi(\lambda_1 + \lambda_2) - \phi(\lambda_1)\phi(\lambda_2)\}; \tag{4.6}$$

$$\begin{aligned} &\mathrm{cum}\left(|\tilde{\phi}_d(\lambda_1)|^2, \tilde{\phi}_d(\lambda_2)\right) \\ &\quad = n^{-2}\left\{2|\phi(\lambda_1)|^2\phi(\lambda_2) - \phi(-\lambda_1)\phi(\lambda_1 + \lambda_2) \right. \\ &\qquad \left. - \phi(\lambda_1)\phi(\lambda_2 - \lambda_1)\right\}; \end{aligned} \tag{4.7}$$

and

$$\operatorname{cum}\big(\tilde{\phi}_d(\lambda_1), \tilde{\phi}_d(-\lambda_1), \tilde{\phi}_d(\lambda_2), \tilde{\phi}_d(-\lambda_2)\big)$$

$$(4.8) \qquad = n^{-3}\Big\{2\sum \phi(a)\phi(b)\phi(-a-b) - 6|\phi(\lambda_1)|^2|\phi(\lambda_2)|^2 - |\phi(\lambda_1+\lambda_2)|^2 - |\phi(\lambda_1-\lambda_2)|^2\Big\},$$

where the summation is over $\{(a,b)\colon a = \lambda_1$ *or* $-\lambda_1$, $b = \lambda_2$ *or* $-\lambda_2\}$. *Furthermore, for any* $s \ge 2$ *it holds uniformly in* $\lambda_1, \dots, \lambda_s$ *that*

$$(4.9) \qquad \operatorname{cum}\big(\tilde{\phi}_d(\lambda_1), \dots, \tilde{\phi}_d(\lambda_s)\big) = O(n^{-s+1}), \qquad n \to \infty.$$

The proof follows from straightforward computations [see Brillinger (1981), pages 19–21, for general methods of computing cumulants; see also Chiu (1991a), Lemma 1].

For the rest of the paper $M > 0$ denotes a finite generic constant that does not depend on n (but that may depend on k and/or p); $A^c = (-\infty, \infty) - A$ for any set A; and $I[\cdot]$ denotes the indicator function on the set $[\cdot]$.

PROOF OF LEMMA 2.1. We note that $h_2 \le h_1$ and, under the conditions of the present lemma, $h_1(0) = \int_{-\infty}^{\infty}\lambda^{2k}|\phi(\lambda)|\,d\lambda$ is finite and $f^{(2k)}$ exists and is bounded over the whole real line [hence $\operatorname{Var} f^{(2k)}(X_1)$ is finite]. Now, (2.6) can be readily derived from (1.6), (1.8) and (1.9). Next, we have by (1.8) and (4.5) that, for all $\Lambda > 0$,

$$(4.10) \qquad \begin{aligned} 4\pi^2 \operatorname{Var}\tilde{\theta}(\Lambda) &= \operatorname{Var} T_3^0([-\Lambda, \Lambda]) \\ &= \operatorname{Var}\{T_1^0([-\Lambda,\Lambda]) + 2T_2([-\Lambda,\Lambda])\}. \end{aligned}$$

By arguments analogous to those in Chiu [(1991a), Lemma 4], we obtain

$$(4.11) \qquad \operatorname{Var} T_2((-\infty,\infty)) = 4\pi^2 n^{-1}\operatorname{Var} f^{(2k)}(X_1).$$

By (4.6) we have, for all $\Lambda > 0$,

$$(4.12) \qquad \begin{aligned} n\operatorname{Var} T_2([-\Lambda,\Lambda]^c) &= \int_{|\lambda_1|>\Lambda}\int_{|\lambda_2|>\Lambda}(\lambda_1\lambda_2)^{2k} \\ &\qquad \times \phi(-\lambda_1)\phi(-\lambda_2)\phi(\lambda_1+\lambda_2)\,d\lambda_1\,d\lambda_2 - h_2^2(\Lambda) \\ &\le 2h_1^2(\Lambda). \end{aligned}$$

Equations (4.11) and (4.12) and an application of the Cauchy–Schwarz inequality give, for all $\Lambda > 0$,

$$(4.13) \qquad \big|\operatorname{Var} T_2([-\Lambda,\Lambda]) - 4\pi^2 n^{-1}\operatorname{Var} f^{(2k)}(X_1)\big| \le Mn^{-1}h_1(\Lambda).$$

[Here and below M does not depend on Λ either.] Using (4.7) we get, for all $\Lambda > 0$,

$$(4.14) \qquad \begin{aligned} &\operatorname{Cov}\big(T_1^0([-\Lambda,\Lambda]), T_2([-\Lambda,\Lambda])\big) \\ &\quad = \int_{-\Lambda}^{\Lambda}\int_{-\Lambda}^{\Lambda}(\lambda_1\lambda_2)^{2k}\phi(-\lambda_2)\operatorname{cum}\big(|\tilde{\phi}_d(\lambda_1)|^2, \tilde{\phi}_d(\lambda_2)\big)\,d\lambda_1 d\lambda_2 \\ &\quad \le Mn^{-2}. \end{aligned}$$

Let us denote the left-hand sides of (4.6) and (4.8) by $c(\lambda_1, \lambda_2)$ and $d(\lambda_1, \lambda_2)$, respectively. Then, for all $\Lambda > 0$,

$$\begin{aligned}
&\operatorname{Var} T_1^0([-\Lambda, \Lambda]) \\
&\quad = \int_{-\Lambda}^{\Lambda}\int_{-\Lambda}^{\Lambda} (\lambda_1\lambda_2)^{2k} \operatorname{cum}\left(|\tilde{\phi}_d(\lambda_1)|^2, |\tilde{\phi}_d(\lambda_2)|^2\right) d\lambda_1\, d\lambda_2 \\
&\quad = \int_{-\Lambda}^{\Lambda}\int_{-\Lambda}^{\Lambda} (\lambda_1\lambda_2)^{2k}\{2c(\lambda_1, \lambda_2)c(-\lambda_1, -\lambda_2) + d(\lambda_1, \lambda_2)\}\, d\lambda_1\, d\lambda_2 \qquad (4.15)\\
&\quad \le 2n^{-2}\left\{R(\Lambda) + 3(h_1(0) - h_1(\Lambda))^2\right\} \\
&\qquad + n^{-3}\left\{2R(\Lambda) + 14(h_1(0) - h_1(\Lambda))^2\right\},
\end{aligned}$$

where $R(\Lambda) = \int_{-\Lambda}^{\Lambda}\int_{-\Lambda}^{\Lambda}(\lambda_1\lambda_2)^{2k}|\phi(\lambda_1 + \Lambda_2)|^2\, d\lambda_1\, d\lambda_2$ is the dominating term here. Evidently,

$$R(\Lambda) \le \Lambda^{2k}\int_{-\Lambda}^{\Lambda} \lambda_2^{2k}\int_{-\infty}^{\infty} |\phi(\lambda_1 + \lambda_2)^2\, d\lambda_1\, d\lambda_2 \le M\Lambda^{4k+1}, \qquad \Lambda > 0. \tag{4.16}$$

Inequality (2.7) can be concluded from (4.10) and (4.13)–(4.16) immediately. □

PROOF OF LEMMA 2.2. From (1.8), (2.1) and the fact that $\mathrm{SCV}_k(\hat{\Lambda}) \le \mathrm{SCV}_k(\Lambda)$, for all $\Lambda > 0$, we get

$$\hat{\Lambda}^{2k+1} - \Lambda^{2k+1} \le 2^{-1}d_n\pi\left(\tilde{\theta}(\hat{\Lambda}) - \tilde{\theta}(\Lambda)\right), \qquad \Lambda > 0 \tag{4.17}$$

where $d_n = (n + 1)(2k + 1)$. Pick and fix any ε from the interval $(2^{-1/2}, 1)$ and set

$$u_n = \left\{2\varepsilon(2k + 1)\left[(2p - 2k - 1)(2\varepsilon^2 - 1)\right]^{-1} n\right\}^{1/(2p)}. \tag{4.18}$$

By denoting the interval $I_m = [u_n, u_n\varepsilon^{-(m+1)/(2k+1)}]$, $m = 0, 1, \dots,$ we have, by (4.17),

$$\begin{aligned}
&P\left[\hat{\Lambda}^{2k+1} > (1 - \varepsilon)^{-1}u_n^{2k+1}\right] \\
&\quad = P\left[\hat{\Lambda}^{2k+1} - u_n^{2k+1} > \varepsilon\hat{\Lambda}^{2k+1}\right] \\
&\quad \le \sum_{m=0}^{\infty} P\Big[2^{-1}d_n\pi\left(\tilde{\theta}(\hat{\Lambda}) - \tilde{\theta}(u_n)\right) > \varepsilon\hat{\Lambda}^{2k+1}, \\
&\qquad\qquad u_n\varepsilon^{-m/(2k+1)} < \hat{\Lambda} \le u_n\varepsilon^{-(m+1)/(2k+1)}\Big] \qquad (4.19)\\
&\quad \le \sum_{m=0}^{\infty} P\Big[2^{-1}d_nT_3^0(I_m) > \varepsilon^{-m+1}u_n^{2k+1} \\
&\qquad\qquad -2^{-1}d_n\int_{I_m} \lambda^{2k}E|\tilde{\phi}(\lambda)|^2\, d\lambda\Big].
\end{aligned}$$

Note that (1.9), together with Condition (A) with $p > k + 2^{-1}$, implies for all $m \geq 0$,

$$
\begin{aligned}
&2^{-1}d_n\int_{I_m}\lambda^{2k}E|\tilde{\phi}(\lambda)|^2\,d\lambda \\
&\quad = 2^{-1}d_n\int_{I_m}\lambda^{2k}\big(n^{-1}+n^{-1}(n-1)|\phi(\lambda)|^2\big)\,d\lambda \\
&\quad < 2^{-1}(1+n^{-1})\varepsilon^{-m-1}u_n^{2k+1} \\
&\qquad + 2^{-1}d_n n^{-1}(n-1)(2p-2k-1)^{-1}u_n^{2k+1-2p} \\
&\quad < 2^{-1}(1+n^{-1})\big\{\varepsilon^{-m-1}+(2\varepsilon)^{-1}(2\varepsilon^2-1)\big\}u_n^{2k+1} \\
&\quad \leq 2^{-1}(1+n^{-1})\varepsilon^{-m+1}\big\{\varepsilon^{-2}+(2\varepsilon^2)^{-1}(2\varepsilon^2-1)\big\}u_n^{2k+1} \\
&\quad = \varepsilon^{-m+1}u_n^{2k+1}\Big\{(1+n^{-1})2^{-1}\big(1+(2\varepsilon^2)^{-1}\big)\Big\} \\
&\quad \leq \varepsilon^{-m+1}u_n^{2k+1}2^{-1}(1+\delta)
\end{aligned}
\tag{4.20}
$$

for all $n \geq 2\delta(1-\delta)^{-1}$, where $\delta = 2^{-1}[1+(2\varepsilon^2)^{-1}] < 1$. From arguments similar to those in (4.12) and applying the Cauchy–Schwarz inequality we have, for all $1 \leq a < b < \infty$,

$$
\begin{aligned}
nET_2^2(J) &\leq \int_J \lambda_2^{2k}|\phi(\lambda_2)|\int_J \lambda_1^{2k}|\phi(\lambda_1)|\,|\phi(\lambda_1+\lambda_2)|\,d\lambda_1\,d\lambda_2 \\
&\leq \int_J \lambda_2^{2k}|\phi(\lambda_2)|h_2^{1/2}(a)\left\{\int_J \lambda_1^{2k}|\phi(\lambda_1+\lambda_2)|^2\,d\lambda_1\right\}^{1/2}d\lambda_2 \\
&\leq h_2^{1/2}(a)Mb^k\int_J \lambda_2^{2k}|\phi(\lambda_2)|\,d\lambda_2 \leq h_2(a)Mb^{2k+(1/2)},
\end{aligned}
\tag{4.21}
$$

where $J = \{\lambda\colon a \leq |\lambda| \leq b\}$. It follows from arguments analogous to those in (4.10), (4.15) and (4.16), and from (4.21) that

$$
\operatorname{Var} T_3^0([a,b]) \leq M\{n^{-2}b^{4k+1} + n^{-1}h_2(a)b^{2k+(1/2)}\}. \tag{4.22}
$$

Here we note that $h_2(u_n) = O(u_n^{-2p+2k+1})$ when $p > k + 2^{-1}$. Applying Chebyshev's inequality to (4.19) and using (4.20) and (4.22) yield, as $n \to \infty$,

$$
\begin{aligned}
&P\big[\hat{\Lambda}^{2k+1} > (1-\varepsilon)^{-1}u_n^{2k+1}\big] \\
&\qquad = O\left\{\varepsilon^{-4}(1-\delta)^{-2}u_n^{-1/2}\sum_{m=0}^{\infty}\varepsilon^{(m+1)/(2k+1)}\right\} \\
&\qquad = O(n^{-1/(4p)})
\end{aligned}
\tag{4.23}
$$

and so (2.16) holds. Next, we have $h_2(\hat{\Lambda})I[\hat{\Lambda} \geq \Lambda_k] \leq h_2(\Lambda_k)$. Set $J_1 = I[\hat{\Lambda} < \Lambda_k]$ and $R_2 = h_2(\hat{\Lambda}) - h_2(\Lambda_k)$. Then (2.4), (2.14), (4.5), (4.17) and an application of the Cauchy–Schwarz inequality give

$$
\begin{aligned}
&2^{-1}R_2J_1\\
&\quad= \Big\{-T_3\big(\big[\hat{\Lambda},\Lambda_k\big]\big) + \pi\big(\tilde{\theta}(\Lambda_k) - \tilde{\theta}(\hat{\Lambda})\big)\Big\}J_1\\
&\quad\le \Big\{T_1([0,\Lambda_k]) + 2\Big|T_2\big(\big[\hat{\Lambda},\Lambda_k\big]\big)\Big| + 2d_n^{-1}\big(\Lambda_k^{2k+1} - \hat{\Lambda}^{2k+1}\big)\Big\}J_1\\
&\quad= O_p(r_n) + 2\Big|T_2\big(\big[\hat{\Lambda},\Lambda_k\big]\big)\Big|J_1\\
&\quad\le O_p(r_n) + \{2T_1([0,\Lambda_k])R_2\}^{1/2}J_1\\
&\quad= O_p(r_n) + O_p\big(r_n^{1/2}\big)R_2^{1/2}J_1,
\end{aligned}
\tag{4.24}
$$

where $r_n = n^{-1+[(2k+1)/(2p)]}$. Set $J_2 = J_1 I[R_2 > r_n]$. Multiplying (4.24) by $R_2^{-1/2}J_2$ yields

$$2^{-1}R_2^{1/2}J_2 \le \big\{O_p(r_n)R_2^{-1/2} + O_p\big(r_n^{1/2}\big)\big\}J_2 = O_p\big(r_n^{1/2}\big),$$

and so, using (2.14), we may conclude (2.17) quickly. □

Proof of Theorem 2.1. By (4.17) and (4.18) we get, for all $p > k + 2^{-1}$, $\{\tilde{\theta}(u_n) - \tilde{\theta}(\hat{\Lambda})\}I[\hat{\Lambda} \le u_n] = O_p(r_n)$, where r_n was defined in (4.24). Furthermore, upon denoting the interval $B = (u_n, (1-\varepsilon)^{-1/(2k+1)}u_n]$, we get from (2.4), (4.5) and (4.18) that

$$
\begin{aligned}
\pi\big(\tilde{\theta}(\hat{\Lambda}) - \tilde{\theta}(u_n)\big)I[\hat{\Lambda} \in B] &\le |T_3(B)| + 2^{-1}h_2(u_n)\\
&\le T_1(B) + 2|T_2(B)| + 2^{-1}h_2(u_n) = O_p(r_n),
\end{aligned}
$$

for all $p > k + 2^{-1}$. This and (4.23) together ensures that $\tilde{\theta}(\hat{\Lambda}) - \tilde{\theta}(u_n) = O_p(r_n)$ for all $p > k + 2^{-1}$. Hence the theorem will be proved if we show that assertions (i) and (ii) remain true if $\hat{\Lambda}$ (i.e., $\hat{\Lambda}_k$) is replaced by u_n throughout. By (4.5), for $p > k + 2^{-1}$,

$$
\begin{aligned}
\pi\big(\tilde{\theta}(u_n) - \theta_k\big) &= T_3([0,u_n]) - 2^{-1}h_2(u_n)\\
&= T_1([0,u_n]) + T_2([-u_n,u_n]) - 2^{-1}h_2(u_n)
\end{aligned}
\tag{4.25}
$$

and so, using (2.4) and (4.18), we see that assertion (i) remains true with $\hat{\Lambda}$ being replaced by u_n. For the rest of the proof we assume $p > 2k+1$. On the rightmost side of (4.25), the first and third terms are of order $o_p(n^{-1/2})$. Furthermore, (4.12) implies that $T_2([-u_n, u_n]^c) = o_p(n^{-1/2})$, and so $n^{1/2}T_2((-\infty,\infty))$ and $n^{1/2}\pi(\tilde{\theta}(u_n) - \theta_k)$ have the same asymptotic distribution. To establish the desired asymptotic normality, by virtue of (4.11), it suffices to show that the r-th order cumulant of $n^{1/2}T_2((-\infty,\infty))$ converges to zero for all $r \ge 3$. However, this may be concluded by noting that such an r-th-order cumulant, by (4.9), is equal to

$$
\begin{aligned}
&n^{r/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{j=1}^{r}\lambda_j^{2k}\phi(-\lambda_j)\mathrm{cum}\big(\tilde{\phi}_d(\lambda_1),\dots,\tilde{\phi}_d(\lambda_r)\big)\,d\lambda_1\cdots d\lambda_r\\
&\qquad= O\big(n^{(-r/2)+1}\big).
\end{aligned}
$$

This completes the proof. □

Acknowledgments. The author is very grateful to Professor S. T. Chiu for many helpful discussions and for assistance in simulation. His thoughtful comments and suggestions contributed significantly to the development of this work. Also, the author is grateful to an Associate Editor and two referees for many valuable suggestions which improved the presentation of the paper. In particular, the critical suggestions by an Associate Editor and a referee of including simulation studies led the author to reconstruct the modified procedure of Section 2.3 that significantly improved its first version. Furthermore, the author is indebted to an Associate Editor for valuable comments on the first revision which improved the quality of this paper considerably.

REFERENCES

BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393.

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

CHIU, S. T. (1991a). Bandwidth selection for kernel density estimation. *Ann. Statist.* **19** 1883–1905.

CHIU, S. T. (1991b). The effect of discretization error on bandwidth selection for kernel density estimation. *Biometrika* **78** 436–441.

CHIU, S. T. (1992). An automatic bandwidth selector for kernel density estimation. *Biometrika* **79** 771–782.

DAVIS, K. B. (1975). Mean square error properties of density estimates. *Ann. Statist.* **3** 1025–1030.

DAVIS, K. B. (1977). Mean integrated square error properties of density estimates. *Ann. Statist.* **5** 530–535.

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. II. *Ann. Statist.* **19** 633–667.

FAN, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* **19** 1273–1294.

GOLDSTEIN, L. and MESSER, K. (1992). Optimal plug-in estimators for nonparametric functional estimation. *Ann. Statist.* **20** 1306–1328.

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

HALL, P. and MARRON, J. S. (1988). Choice of kernel order in density estimation. *Ann. Statist.* **16** 161–173.

HALL, P. and MARRON, J. S. (1991a). Lower bounds for bandwidth selection in density estimation. *Probab. Theory Related Fields* **90** 149–173.

HALL, P. and MARRON, J. S. (1991b). Local minimum in cross-validation functions. *J. Roy. Statist. Soc. Ser. B* **53** 245–252.

HALL, P., MARRON, J. S. and PARK, B. U. (1992). Smoothed cross-validation. *Probab. Theory Related Fields* **92** 1–20.

HALL, P., SHEATHER, S. J., JONES, M. C. and MARRON, J. S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* **78** 521–530.

HASMINSKII, R. Z. and IBRAGIMOV, I. A. (1979). On the nonparametric estimation of functionals. In *Proceedings of the Second Prague Symposium on Asymptotic Statistics* (P. Mandl and M. Hušková, eds.) 41–51. North-Holland, Amsterdam.

HEWITT, E. and STROMBERG, K. (1969). *Real and Abstract Analysis*. Springer, New York

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1982). Estimation of distribution density belonging to a class of entire functions. *Theory Probab. Appl.* **27** 551–562.

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

KOSHEVNIK, YU. A. and LEVIT, B. YA. (1976). On a non-parametric analogue of the information matrix. *Theory Probab. Appl.* **21** 738–753.

MARRON, J. S. and WAND, M. P. (1992). Exact mean integrated squared error. *Ann. Statist.* **20** 712–736.

PARK, B. U. and MARRON, J. S. (1990). Comparison of data-driven bandwidth selectors. *J. Amer. Statist. Assoc.* **85** 66–78.

RITOV, Y. and BICKEL, P. J. (1990). Achieving information bounds in non and semiparametric models. *Ann. Statist.* **18** 925–938.

SCHWEDER, T. (1975). Window estimation of the asymptotic variance of rank estimators of location. *Scand. J. Statist.* **2** 113–126.

SCOTT, D. W. and TERRELL, G. R. (1987). Biased and unbiased cross-validation in density estimation. *J. Amer. Statist. Assoc.* **82** 1131–1146.

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.

SILVERMAN, B. W. (1982). Kernel density estimation using the fast Fourier transform. *J. Roy. Statist. Soc. Ser. C* **31** 93–99.

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

TERRELL, G. R. (1990). The maximal smoothing principle in density estimation. *J. Amer. Statist. Assoc.* **85** 470–477.

VAN ES, B. (1992). Estimating functionals related to a density by a class of statistics based on spacings. *Scand. J. Statist.* **19** 61–72.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON
HOUSTON, TEXAS 77204-3476