The Annals of Statistics
2002, Vol. 30, No. 2, 431–454

EFFECT OF DEPENDENCE ON STOCHASTIC MEASURES OF ACCURACY OF DENSITY ESTIMATORS

BY GERDA CLAESKENS AND PETER HALL

Australian National University and Texas A&M University, and Australian National University

In kernel density estimation, those data values that make a nondegenerate contribution to the estimator (computed at a given point) tend to be spaced well apart. This property has the effect of suppressing many of the conventional consequences of long-range dependence, for example, slower rates of convergence, which might otherwise be revealed by a traditional loss- or risk-based assessment of performance. From that viewpoint, dependence has to be very long-range indeed before a density estimator experiences any first-order effects. However, an analysis in terms of the convergence rate for a particular realization, rather than the rate averaged over all realizations, reveals a very different picture. We show that from that viewpoint, and in the context of functions of Gaussian processes, effects on rates of convergence can become apparent as soon as the boundary between short- and long-range dependence is crossed. For example, the distance between ISE- and MISE-optimal bandwidths is generally of larger order for long-range dependent data. We shed new light on cross-validation, too. In particular we show that the variance of the cross-validation bandwidth is generally larger for long-range dependent data, and that the first-order properties of this bandwidth do not depend on how many data are left out when constructing the cross-validation criterion. Moreover, for long-range dependent data the cross-validation bandwidth is usually perfectly negatively correlated, in the limit, with the optimal stochastic bandwidth.

1. Introduction. In many problems of point estimation from time-series data, estimators fail to be root-n consistent if and only if the sum of the correlations at successive lags diverges. However, this condition, which is usually taken to determine the barrier between short- and long-range dependence [see, e.g., Barndorff-Nielsen and Cox (1989), page 13], appears to have little connection to problems of nonparametric density estimation. Indeed, the data can be either short- or relatively long-range dependent, yet density estimators and their optimal bandwidths can converge at the same rate as they would under the assumption of independence. Results of this type are available from, for example, work of Roussas (1969, 1990), Rosenblatt (1970), Prakasa Rao (1978), Nguyen (1979), Ahmad (1982), Castellana and Leadbetter (1986), Roussas and Ioannides (1987), Castellana (1989), Tran (1989, 1990), Hall and Hart (1990), Hall, Lahiri and

Received July 2000; revised August 2001.
AMS 2000 subject classifications. Primary 62G07; secondary 62M10.
Key words and phrases. Bandwidth, cross-validation, Gaussian process, integrated squared error, kernel methods, long-range dependence, nonparametric density estimator, risk-based analysis.

Truong (1995), Hall (1997), Hart (1996), Kim and Cox (1997) and Masry and Fan (1997). See also the review by Härdle, Lütkepohl and Chen (1997). Accounts of properties of density estimators under long-range dependence have been given by, for example, Cheng and Robinson (1991), Robinson (1991) and Csörgő and Mielniczuk (1995). This work characterizes the strength of dependence that is sufficient for traditional, first-order properties to break down.

The majority of this research has focused on conventional risk- or loss-based measures of performance, however. There, the bandwidth is chosen to minimize the *expected* distance between the estimator and the true density. In the present paper we argue that the differences between short- and long-range dependence are concealed by such an approach. The differences occur at the level of stochastic fluctuations in the distance. We derive results which show that these stochastic terms can be important as soon as the classical divide between short- and long-range dependence is crossed. Prior to that point, properties of stochastic fluctuations of the distance are virtually identical for independent and dependent data; after that point, the order of the stochastic terms can increase sharply with strength of dependence.

The stochastic terms have zero mean, and so their properties have gone unnoticed in traditional loss- and risk-based analyses. As a result, the literature has tended to overlook the more direct effects of long-range dependence on properties of density estimators and empirically chosen bandwidths. Our results will be derived in the case of time series that may be represented as functions of Gaussian processes; this model plays a central role in a variety of practical applications. It allows the strength of dependence to be represented simply and unambiguously in terms of lagged correlations. For a more general process the strength would be relatively difficult to describe, and the results would be more difficult to present, discuss and derive. In particular, an account of strength of dependence based on mixing conditions gives only an upper bound to the strength for a given separation of time points at which the process is observed. The detailed results that we shall give require considerably more precision than a single bound can supply; we need bounds on both sides, and they must be asymptotically equivalent in the context of relatively long separation in time.

To appreciate our results in more detail, consider measuring the accuracy of a kernel estimator $\hat{f}$ of a density f, in terms of either mean integrated squared error (MISE) or its stochastic counterpart, integrated squared error (ISE). Let h^{opt} and $\hat{h}^{\mathrm{opt}}$ be the respective bandwidths that minimize these criteria, and let n denote sample size. It is known that, under the assumption of independence, these two bandwidths are generally close, and in fact the ratio $\hat{h}^{\mathrm{opt}}/h^{\mathrm{opt}}$ converges to 1 at rate $O_p(n^{-1/10})$; see Hall and Marron (1987). We shall show that this rate persists for dependent data, up to the point where the condition of short-range dependence fails (i.e., where the sample mean fails to be root-n consistent for the population mean). From there on, the ratio $\hat{h}^{\mathrm{opt}}/h^{\mathrm{opt}}$ can converge to 1 more slowly than it would for independent data.

Similarly, the size of the difference between values of ISE at the respective bandwidths h^{opt} and $\hat{h}^{\mathrm{opt}}$ can increase sharply as the classical boundary between short- and long-range dependence is crossed. Prior to that point, the size is the same as it would be for independent data, but after the boundary is crossed, size can be of strictly larger order. "Size" here refers to fluctuations that have zero mean and which would not be noticed in a purely loss- or risk-based assessment of performance.

These examples show that by focusing on stochastic aspects of convergence criteria we obtain a particularly simple way of characterizing the effects of long-range dependence on density estimators. The situation is very different if we treat only MISE. Indeed, the ratio of MISE for time-series data to its counterpart for independent data converges to 1, as sample size increases, if and only if the rate of convergence of the sample mean equals $o_p(n^{-1/5})$. If this condition is violated, then the time series is very long-range dependent indeed. Therefore, one has to move well beyond the traditional barrier between short- and long-range dependence before a significant effect is noticed on a loss- or risk-based assessment of performance.

There is an intuitive explanation for these results, which helps explain the apparent contradictions between them. Suppose the density estimator $\hat{f}$ is constructed using a bandwidth h, and is computed from a segment $\mathcal{X} = \{X_1, \dots, X_n\}$ of the time series. The contribution of X_i to $\hat{f}(x)$ will be significant only if X_i lies within a small interval of width $O(h)$ centered at x. The probability of this occurring is small, and so the expected values of the distances between indices of successive X_i's that make significant contributions to $\hat{f}(x)$ will tend to be large, unless the time series is exceptionally long-range dependent. (In the latter case, conditional on $X_i = x$, the probability that X_j is close to X_i for a run of values j close to i can be relatively high.) As a result, the effects of long-range dependence tend to be dampened down by the very nature of density estimation—estimation of f at x inherently involves only a sparse subsequence of the time series.

However, while the expected value of the distance between successive non-negligible components in the series may be large, the variability of fluctuations in that distance can be substantial, even in the case of relatively moderate long-range dependence. As a result, a view of the effects of long-range dependence which ignores these stochastic fluctuations can underestimate the impact of relatively low-level departures of the time series from independence. Interestingly, the same arguments do not apply to nonparametric regression estimators, at least in cases where dependence among errors is indexed in the same order as the explanatory variable. (This is the usual time-series model in regression.) There, the smoothing operation does not have nearly the same tendency to mask the effects of long-range dependence, and the results described in the present paper do not have close analogues.

In the context of density estimation our results have immediate implications for time-series applications of the popular cross-validation technique for bandwidth choice; see Hart and Vieu (1990) for discussion of that method. Cross-validation is sometimes interpreted as a method for selecting the bandwidth that minimises ISE, rather than one that minimises MISE; see, for example, Hall and Johnstone (1992). Our results lend considerable weight to this viewpoint. We show that the cross-validation bandwidth $\hat{h}^{\mathrm{cv}}$ exhibits the properties noted above for $\hat{h}^{\mathrm{opt}}$. In particular, the ratio $\hat{h}^{\mathrm{cv}}/h^{\mathrm{opt}}$ converges to 1 at the same rate as for dependent data if the data are short-range dependent; and beyond that point, it generally converges to 1 at the same, slow rate as $\hat{h}^{\mathrm{opt}}/h^{\mathrm{opt}}$.

Our work also reveals that the variability of $\hat{h}^{\mathrm{cv}}$ can increase markedly with increasing range of dependence of the time series. We show that this property does not depend on the leave-out number that is used to construct $\hat{h}^{\mathrm{cv}}$. The effects of that quantity go into higher-order terms. Hence, there does not seem to be a good reason for taking the leave-out number to be other than 1. Numerical simulations, such as those of Hart and Vieu (1990), lend weight to this suggestion.

Furthermore, in the context of long-range dependence and except for degenerate cases, the asymptotic correlation between $\hat{h}^{\mathrm{cv}}$ and $\hat{h}^{\mathrm{opt}}$ equals -1. That is, $\hat{h}^{\mathrm{cv}}$ makes a stochastic correction in exactly the wrong direction, relative to the correction supplied by $\hat{h}^{\mathrm{opt}}$. In the more traditional context of independent data the correction is negative, but not -1. See Härdle, Hall and Marron [(1988), with discussion by Johnstone and Scott] and Hall and Johnstone (1992).

We shall focus on L_2 properties, and in particular on rates of convergence of ISE and MISE. It is readily shown that the density estimators we consider are consistent in the supremum metric, as well being consistent in L_2.

2. Main results.

2.1. *Expansion of integrated squared error for dependent data.* We begin by defining the density estimator. Let $\mathcal{X} = \{X_1, \ldots, X_n\}$ denote an n-segment of an infinite stationary time series $\mathcal{X}_\infty = \{X_1, X_2, \ldots\}$ with marginal density f. A nonparametric estimator of f is given by

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where h is a bandwidth and K a kernel function.

For $i \geq 1$ let $f_i(x_1, x_2)$ denote the joint density of (X_1, X_{i+1}), and define $g_i(x_1, x_2) = f_i(x_1, x_2) - f(x_1) f(x_2)$. Write $g_i^{(r,s)}(x_1, x_2)$ for the $(r+s)$th derivative of $g_i(x_1, x_2)$, r times with respect to x_1 and s times with respect to x_2. We assume that

(2.1) K is a compactly supported, bounded, symmetric probability density and

(2.2) f has six bounded, uniformly integrable derivatives, each g_i is four times differentiable and each of the first four derivatives of g_i is bounded uniformly in i.

Suppose too that, for some $\varepsilon > 0$ and a positive sequence $\gamma = \gamma(n)$,

$$\begin{aligned}&\int \sup_{|z|\le\varepsilon} |f^{(6)}(x+z)|\,dx < \infty,\\ &\sum_{i=1}^{n}\int \sup_{\substack{j_1+j_2=4\\ |z_1|,|z_2|\le\varepsilon}} |g_i^{(j_1,j_2)}(x+z_1,x+z_2) - g_i^{(j_1,j_2)}(x,x)|\,dx = O(\gamma).\end{aligned} \tag{2.3}$$

Define $\kappa = \int K^2$, $\mu_j = \int x^j K(x)\,dx$, $I_j = \int (f^{(j)})^2$,

$$\tau_{n1} = 2\sum_{i=1}^{n-1}(1-n^{-1}i)\int g_i(x,x)\,dx,$$

$$\tau_{n2} = \mu_2\sum_{i=1}^{n-1}(1-n^{-1}i)\int \{g_i^{(2,0)}(x,x) + g_i^{(0,2)}(x,x)\}\,dx,$$

$$M(h) = \int E(\hat{f}_h - f)^2, \qquad N(h) = (nh)^{-1}\kappa - n^{-1}I_0 + h^4\tfrac{1}{4}\mu_2^2 I_2 - h^6\tfrac{1}{24}\mu_2\mu_4 I_3.$$

Thus, $M(h)$ represents mean integrated squared error of the estimator $\hat{f}_h$. The following result describes errors that arise in approximating $M(h)$ by $N(h)$.

THEOREM 2.1. *For dependent data satisfying conditions* (2.1), (2.2) *and* (2.3),

$$M(h) = N(h) + n^{-1}(\tau_{n1} + \tau_{n2}h^2) + O(n^{-1}h^2 + n^{-1}h^4\gamma + h^8), \tag{2.4}$$

where, for any functions $h_1(n), h_2(n)$ satisfying $0 < h_1(n) < h_2(n) < \infty$, $h_j(n) \to 0$ and $nh_j(n) \to \infty$, the remainder terms are of the stated orders uniformly in $h \in [h_1(n), h_2(n)]$.

2.2. *Function of a Gaussian process.* Here we note that the results in Section 2.1 are valid if $\mathcal{X}_\infty$ is a sufficiently regular function of a Gaussian process. Suppose $X_i = a(Y_i)$, where a is a function and the process $\mathcal{Y}_\infty = \{Y_1, Y_2, \ldots\}$ is stationary and Gaussian with zero mean, unit variance and correlation τ_i between Y_1 and Y_{i+1}. To simplify our theoretical arguments we shall assume that a is infinitely differentiable and monotone, and that the function $b = a^{-1}$ (i.e., the inverse of the function a) satisfies, for all x and each $\varepsilon > 0$,

$$
\begin{aligned}
&|b'(x)|^{-1} \le C(\varepsilon)\exp\{\varepsilon b(x)^2\}, \\
&|b^{(j)}(x)| \le C_j(\varepsilon)\exp\{\varepsilon b(x)^2\} \qquad \text{for } j \ge 0,
\end{aligned}
\tag{2.5}
$$

where $C(\varepsilon)$ and $C_j(\varepsilon)$ are constants. A more elaborate argument will show that the number of derivatives needed of a is in fact bounded and that ε may be kept fixed but sufficiently small. Examples of functions a satisfying these conditions include smooth distribution functions of distributions with heavier tails than the Normal, for example, where the tails decrease polynomially rather than exponentially fast.

We shall suppose too that $E(X^2) = 1$; this assumption will simplify formulae such as (2.6) below. Since a is monotone, then, with N a standard Normal random variable, $E\{a'(N)\} \ne 0$, which implies that the Hermite rank of a [see Taqqu (1975)] equals 1.

Let $\rho_i = \rho(i)$ denote the correlation between X_1 and X_{i+1}. If $\tau_i \to 0$ as $i \to \infty$, then ρ_i also converges to 0 with increasing i. If in addition (2.5) holds, then either directly by Taylor expansion or using the definition of Hermite rank we have

$$
\rho_i \sim \tau_i\{Ea'(N)\}^2 \qquad \text{as } i \to \infty.
\tag{2.6}
$$

To rule out cases where ρ_i is changing very erratically we shall often suppose in addition to (2.5) that

$$
\begin{aligned}
&\rho_i \text{ is of one sign for all sufficiently large } i, \\
&\text{and for some } \varepsilon > 0,\ |\rho_i| = O(i^{-\varepsilon}) \text{ as } i \to \infty.
\end{aligned}
\tag{2.7}
$$

Under (2.5), and provided $\rho_i \to 0$, the function g_i defined in Section 2.1 admits a Taylor expansion in powers of ρ_i: for each $r \ge 1$, and uniformly in x_1 and x_2,

$$
g_i(x_1, x_2) = \sum_{j=1}^{r} \rho_i^j \psi_j(x_1, x_2) + O(|\rho_i|^{r+1})
\tag{2.8}
$$

as $i \to \infty$, where the functions ψ_j depend only on a. Further, (2.8) continues to hold if we apply the operator $(\partial/\partial x_1)^r(\partial/\partial x_2)^s$ to both g_i and ψ_j. This suggests that (2.3) holds if we take

$$
\gamma = \sum_{i=1}^{n} |\rho_i|,
\tag{2.9}
$$

and also that $\tau_{n1} = O(\gamma)$ and $\tau_{n2} = O(\gamma)$. In fact all these results are implied by (2.5).

In classical problems of inference for time series the case where γ is bounded, that is,

$$
\sum_{i=1}^{\infty} |\rho_i| < \infty,
\tag{2.10}
$$

is usually referred to as short-range dependence. Under mild additional assumptions on the sequence of correlations, for example, finite variance and (2.7), boundedness of γ is necessary and sufficient for the mean of the data $\mathcal{X}$ to be root-n consistent for the population mean.

Much of our interest will center on the case of long-range dependence, since otherwise, as we shall show, the convergence rates of bandwidth selection rules are identical to their counterparts under independence. In such contexts, if (2.5) and (2.7) hold, then the ratios $|\tau_{n1}|/\gamma$ and $|\tau_{n2}|/\gamma$ [with γ defined as in (2.9)] are bounded away from zero and infinity. For reasons such as these, γ defined by (2.9) provides a particularly effective, and simple, description of different orders of magnitude that affect properties of integrated squared error.

2.3. *The MISE-optimal bandwidth.* Let M_0 denote the version of M in the case where the components of $\mathcal{X}$ are independent and identically distributed with marginal density f. Note that, in this setting, each g_i vanishes. Therefore the same is true of τ_{n1} and τ_{n2}, and so (2.4) simplifies to

$$M_0(h) = N(h) + O(n^{-1}h^2 + h^8). \tag{2.11}$$

(The remainder terms here are of the stated orders uniformly in $h \in [h_1(n), h_2(n)]$, where $h_1(n)$ and $h_2(n)$ are as in Theorem 2.1.) It follows from (2.4) and (2.11) that the ratio of MISE for time-series data to its counterpart for independent data converges to 1 if and only if $\gamma = o(n^{1/5})$, or equivalently, if and only if the sample mean, $\bar{X} = n^{-1}\sum_i X_i$, converges to the population mean at rate $o_p(n^{-2/5})$. Long-range dependence is usually thought of as starting when the rate $o_p(n^{-1/2})$ is violated.

As Hall, Lahiri and Truong (1995) noted, (2.11) implies that the bandwidth h_0^{opt} that is optimal for independent data, in the sense of minimizing $M_0(h)$, satisfies

$$h_0^{\mathrm{opt}} = (J_1/n)^{1/5} + J_2(J_1/n)^{3/5} + O(n^{-4/5}), \tag{2.12}$$

where $J_1 = \kappa/(\mu_2^2 I_2)$, $J_2 = \mu_4 I_3/(20\mu_2 I_2)$ and $I_j = \int (f^{(j)})^2$. To obtain the analogue of this formula for a dependent process, assume that

$$\tau_{n2} = O(\gamma) \quad \text{and} \quad \gamma = o(n^{7/10}).$$

Then in view of (2.4) the bandwidth h^{opt} that minimizes $M(h)$ satisfies

$$h^{\mathrm{opt}} = h_0^{\mathrm{opt}} - J_3 n^{-4/5}\tau_{n2} + O(\gamma n^{-6/5} + \gamma^2 n^{-7/5}), \tag{2.13}$$

where $J_3 = \frac{1}{5}(\kappa^2 \mu_2^6 I_2^3)^{-1/5}$.

In the case of short-range dependence, (2.12) and (2.13) are identical. Furthermore, it follows from (2.13) that $h^{\mathrm{opt}} \sim h_0^{\mathrm{opt}}$ provided $\gamma = o(n^{3/5})$. A more detailed analysis, based on a longer version of the expansion (2.4), will show that $h^{\mathrm{opt}} \sim h_0^{\mathrm{opt}}$ if and only if $\gamma = o(n^{3/5})$. Now, the latter condition fails if and only if the L^2 rate of convergence of the sample mean to the population mean is no better than $O_p(n^{-1/5})$. By this time the series $\mathcal{X}_\infty$ is particularly long-range dependent. In summary, only for very long-range dependence do MISE properties of a time series differ from those in the case of independence.

2.4. *The ISE-optimal bandwidth.* Integrated squared error is defined by

$$\widehat{M}(h) = \int (\hat{f}_h - f)^2,$$

where we use the "hat" notation to indicate that $\widehat{M}$ is a random quantity. Let $\hat{h}^{\mathrm{opt}} = \operatorname{argmin}_h \widehat{M}(h)$ denote the bandwidth that is optimal in the sense of minimizing the distance of $\hat{f}_h$ from f for the particular dataset $\mathcal{X}$. When the data are independent it is known that $\hat{h}^{\mathrm{opt}}$ is asymptotic to h^{opt} (the MISE-optimal bandwidth), although the rate of convergence is slow. In fact, $n^{1/10}\{(\hat{h}^{\mathrm{opt}}/h^{\mathrm{opt}}) - 1\}$ has a limiting Normal distribution with zero mean and nonzero variance; see Hall and Marron (1987). This result continues to hold if the data are short-range dependent, that is, if (2.10) holds. However, if that condition is violated, then the rate of convergence of $\hat{h}^{\mathrm{opt}}/h^{\mathrm{opt}}$ to 1 can be very slow indeed, as our next result will show.

Recall from Sections 2.1 and 2.3 that $h^{\mathrm{opt}} \sim c_0 n^{-1/5}$ and $M(h) \sim c_1(nh)^{-1} + c_2h^4$, where $c_0 = J_1^{1/5}$, $c_1 = \kappa$ and $c_2 = 4^{-1}\mu_2^2 I_2$; here it is assumed that $\gamma = o(n^{1/5})$. It may similarly be shown that $M''(h^{\mathrm{opt}}) \sim c_3 n^{-2/5}$, where $c_3 = 2c_1c_0^{-3} + 12c_2c_0^2$. Define

$$s_1^2 = (2/c_0)^3 \left(\int f^2\right) \int \left[K(u+v)\{K(v) - K_1(v)\}\, dv\right]^2 du,$$

where $K_1(x) = -xK'(x)$, and if (2.5), (2.7) and (2.10) hold, put

$$s_2^2 = \operatorname{var}\{f''(X_1)\} + 2\sum_{i=1}^{\infty} \operatorname{cov}\{f''(X_1), f''(X_{i+1})\}.$$

Under (2.5) and (2.7), $\operatorname{cov}\{f''(X_i), f''(X_{i+1})\} = O(|\rho_i|)$ as $i \to \infty$. Hence, if in addition (2.10) is valid, the infinite series on the right-hand side in the definition of s_2^2 converges absolutely. Put

$$\sigma_1^2 = c_3^{-2}\{s_1^2 + (2\mu_2 c_0)^2 s_2^2\},$$

$$\sigma_2^2 = 8c_3^{-2}\mu_2^2 \left(\frac{E[f''\{a(N)\}a'(N)]}{E\{a'(N)\}}\right)^2.$$

We shall suppose that

(2.14) $\mathcal{X}_\infty$ is a function of a Gaussian process, (2.5) and (2.7) hold, and the kernel K is a compactly supported, symmetric probability density with two Hölder-continuous derivatives.

THEOREM 2.2. *Assume* (2.14). *If the time series $\mathcal{X}_\infty$ is short-range dependent in the sense of* (2.10), *then*

(2.15) $$n^{1/10}\{(\hat{h}^{\mathrm{opt}}/h^{\mathrm{opt}}) - 1\} \to N(0, \sigma_1^2) \qquad \textit{in distribution}.$$

On the other hand, if (2.10) *is violated and* $\gamma = O(n^{(1/5)-\varepsilon})$ *for some* $\varepsilon > 0$, *then*

$$n^{1/10}\gamma^{-1/2}\{(\hat{h}^{\text{opt}}/h^{\text{opt}}) - 1\} \to N(0, \sigma_2^2) \qquad \textit{in distribution}. \tag{2.16}$$

If the function a is antisymmetric, then f is symmetric, and so $f''\{a(N)\}a'(N)$ is antisymmetric. In this special case, $\sigma_2^2 = 0$ and the limit distribution at (2.16) is degenerate. As a rule, however, $\sigma_2^2 \neq 0$ and (2.16) describes a nontrivial limit theorem. The case $\sigma_2^2 = 0$ is of completely different character from when $\sigma_2^2 \neq 0$, and a nondegenerate limit theorem there is highly complex, depending intimately on details of the behavior of ρ_i as $i \to \infty$. In general the limit distribution has two parts, one deriving from the limit distribution of $\sum_i f''(X_i)$, which can be non-Normal when $\sigma_2^2 = 0$ [compare Taqqu (1975)], and the other coming from a term of U-statistic character in an expansion of $\hat{h}^{\text{opt}}$. This term is negligible when $\sigma_2^2 \neq 0$, but can be significant when $\sigma_2^2 = 0$. In the case of short-range dependence, where (2.10) holds, the U-statistic term gives rise to the contribution $c_3^{-2}s_1^2$ to the variance σ_1^2.

Results (2.15) and (2.16) demonstrate that the well-known order of the difference between the MISE- and ISE-optimal bandwidths in the independence case is preserved under dependence if the dependence is of short range. When the latter condition fails, however, the MISE-optimal bandwidth can be approximated very poorly by its ISE counterpart. Indeed, if $\gamma \sim cn^{1/5}$ for a constant $c > 0$, then it is no longer true that $\hat{h}^{\text{opt}}/h^{\text{opt}}$ converges to 1 in probability. In this case (2.16) fails, as too do (2.18) and (2.21)–(2.23) below. However, our results continue to hold in cases where the condition $\gamma \sim cn^{1/5}$ "only just" fails, for example, where $\gamma = n^{1/5}L(n)$, L is slowly varying at infinity, and $L(n) \to 0$ as $n \to \infty$, although they require strengthening of other regularity conditions.

Both (2.15) and (2.16) have analogues for convergence of ISE itself. It is known that, in the case of independent data, $\widehat{M}(h^{\text{opt}}) - \widehat{M}(\hat{h}^{\text{opt}})$ converges to 0 at rate n^{-1}. Our next result shows that this remains true for dependent data, provided the dependence is of short range in the sense of (2.10). For more strongly dependent data, the difference between values of ISE at bandwidths h^{opt} and $\hat{h}^{\text{opt}}$ is generally of larger order.

Define $\sigma_3^2 = \frac{1}{2}c_0^2 c_3 \sigma_1^2$ and $\sigma_4^2 = \frac{1}{2}c_0^2 c_3 \sigma_2^2$. Let χ_1^2 denote a random variable having the chi-squared distribution on one degree of freedom.

THEOREM 2.3. *Assume* (2.14). *If the time series* $\mathcal{X}_\infty$ *is short-range dependent in the sense of* (2.10), *then*

$$n\{\widehat{M}(h^{\text{opt}}) - \widehat{M}(\hat{h}^{\text{opt}})\} \to \sigma_3^2 \chi_1^2 \qquad \textit{in distribution}. \tag{2.17}$$

If (2.10) *fails and* $\gamma = O(n^{(1/5)-\varepsilon})$ *for some* $\varepsilon > 0$, *then*

$$n\gamma^{-1}\{\widehat{M}(h^{\text{opt}}) - \widehat{M}(\hat{h}^{\text{opt}})\} \to \sigma_4^2 \chi_1^2 \qquad \textit{in distribution}. \tag{2.18}$$

Comparing (2.17) and (2.18) with the analogues of those results in Theorems 2.2 and 2.3 we see that the influence of long-range dependence on the optimal empirical bandwidth can be very different from that when optimality is measured in purely deterministic, MISE terms.

Of course, $\hat{h}^{\mathrm{opt}}$ is not a practical choice as a bandwidth; although it is optimal from an empirical, rather than risk-based, viewpoint, it still depends heavily on the unknown density. However, the fact that $\hat{h}^{\mathrm{opt}}$ can be so highly variable, for long-range dependent data, indicates that in practical problems it will be very difficult to develop an accurate, data-based approximation. In the next section we produce evidence to support this claim, showing that the cross-validation bandwidth is at least as variable as $\hat{h}^{\mathrm{opt}}$.

2.5. *The cross-validation bandwidth.* Let $\ell \geq 1$ be an integer; it will determine how many data values we omit at the leave-one-out step of cross-validation. Put

$$\hat{f}_{h,-i}(x) = \frac{1}{n_\ell h} \sum_{j:|i-j|>\ell} K\left(\frac{x - X_j}{h}\right),$$

where

$$n_\ell = n\#\{(i, j) : |i - j| > \ell \text{ and } 1 \leq i, j \leq n\}.$$

The cross-validation criterion is

$$\mathrm{CV}(h) = \int \hat{f}_h^2 - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{h,-i}(X_i)$$

and represents an approximation to $\widehat{M}(h) + \int f^2$. The cross-validation bandwidth $\hat{h}^{\mathrm{cv}}$ is the one that minimizes $\mathrm{CV}(\cdot)$.

The properties of the optimal empirical bandwidth $\hat{h}^{\mathrm{opt}}$ evinced in Theorems 2.2 and 2.3 have direct analogues for its cross-validation counterpart, as we now show. Critically, we shall show that the leave-one-out number ℓ has negligible influence on the convergence rate. Even in the case of long-range dependence, where (2.10) is violated, its effect is expressed only through relatively high-order terms.

Define

$$s_3^2 = (2/c_0)^3 \left(\int f^2\right) \int \{K(u) + uK'(u)\}^2 \, du,$$

$$\sigma_5^2 = c_3^{-2}\{s_3^2 + (2\mu_2 c_0)^2 s_2^2\}, \qquad \sigma_6^2 = \tfrac{1}{2} c_0^2 c_3 \sigma_5^2.$$

For simplicity we shall suppose that $\hat{h}^{\mathrm{cv}}$ is taken to equal the minimum of $\mathrm{CV}(h)$ in an interval $[C_1 n^{-1/5}, C_2 n^{-1/5}]$, where (C_1, C_2) contains the optimal constant $c_0 = J_1^{1/5}$ in the formula $h^{\mathrm{opt}} \sim c_0 n^{-1/5}$. This avoids difficulties with spurious small minima, which can persist when using cross-validation for dependent data.

THEOREM 2.4. *Assume* (2.14). *Suppose too that* $\ell = O(n^{1-\varepsilon})$ *for some* $\varepsilon > 0$. *If the time series* $\mathcal{X}_\infty$ *is short-range dependent in the sense of* (2.10), *then*

$$n^{1/10}\{(\hat{h}^{\mathrm{cv}}/\hat{h}^{\mathrm{opt}}) - 1\} \to N(0, \sigma_5^2) \qquad \text{in distribution}, \tag{2.19}$$

$$n\{\widehat{M}(\hat{h}^{\mathrm{cv}}) - \widehat{M}(\hat{h}^{\mathrm{opt}})\} \to \sigma_6^2 \chi_1^2 \qquad \text{in distribution}. \tag{2.20}$$

On the other hand, if (2.10) *is violated and* $\gamma = O(n^{(1/5)-\varepsilon})$ *for some* $\varepsilon > 0$, *then*

$$(h^{\mathrm{opt}}\gamma)^{-1/2}\{(\hat{h}^{\mathrm{cv}}/h^{\mathrm{opt}}) - 1\} \to N(0, \sigma_2^2) \qquad \text{in distribution}, \tag{2.21}$$

$$n\gamma^{-1}\{\widehat{M}(\hat{h}^{\mathrm{cv}}) - \widehat{M}(\hat{h}^{\mathrm{opt}})\} \to 4\sigma_4^2 \chi_1^2 \qquad \text{in distribution}, \tag{2.22}$$

$$n^{3/10}\gamma^{-1/2}(\hat{h}^{\mathrm{opt}} + \hat{h}^{\mathrm{cv}} - 2h^{\mathrm{opt}}) \to 0 \qquad \text{in probability}. \tag{2.23}$$

Result (2.23) shows that, except in degenerate cases where $\sigma_2^2 = 0$, the asymptotic correlation between $\hat{h}^{\mathrm{opt}}$ and $\hat{h}^{\mathrm{cv}}$ equals -1. This result in particular, but also (2.21) and (2.22), argues that cross-validation may not be an ideal bandwidth choice method for long-range dependent data. A plug-in rule would suffer less from variability, and on those grounds would be more attractive. However, as the strength of dependence approaches a point where where the condition $\gamma = O(n^{(1/5)-\varepsilon})$ fails for all $\varepsilon > 0$, the asymptotic equivalence of the MISE-optimal bandwidth h^{opt} and the ISE-optimal bandwidth $\hat{h}^{\mathrm{opt}}$ also tends to disappear. See the discussion two paragraphs below Theorem 2.2. Therefore the advantages of plug-in rules, which focus on estimating h^{opt}, become less clear.

The fact that the value of ℓ plays no first-order role in Theorem 2.4 indicates that the extent to which there is an "optimal" ℓ will be determined by relatively high-order properties. There seems little to be gained from pursuing this issue from a practical viewpoint, however, since the variability of $\hat{h}^{\mathrm{cv}}$, and hence the difficulty of identifying high-order effects of choice of ℓ, increase with increasing range of dependence.

It is of interest to address the difference between $\widehat{M}(\hat{h}^{\mathrm{cv}})$ and $\widehat{M}(h^{\mathrm{opt}})$, rather than between $\widehat{M}(\hat{h}^{\mathrm{cv}})$ and $\widehat{M}(\hat{h}^{\mathrm{opt}})$ as at (2.20) and (2.22). In the case of short-range dependence the results are effectively the same as for independence and may be deduced from the more extensive properties reported by Hall and Marron (1987) for that setting. In the context of long-range dependence they follow from Theorems 2.3 and 2.4, as follows:

$$\begin{aligned}
n\gamma^{-1}\{\widehat{M}(\hat{h}^{\mathrm{cv}}) - \widehat{M}(h^{\mathrm{opt}})\} &\sim \tfrac{1}{2}n\gamma^{-1}\{(\hat{h}^{\mathrm{cv}} - \hat{h}^{\mathrm{opt}})^2 - (h^{\mathrm{opt}} - \hat{h}^{\mathrm{opt}})^2\}\widehat{M}''(\hat{h}^{\mathrm{opt}}) \\
&\sim \tfrac{3}{2}n\gamma^{-1}(\hat{h}^{\mathrm{opt}} - h^{\mathrm{opt}})^2\widehat{M}''(\hat{h}^{\mathrm{opt}}) \\
&\sim 3n\gamma^{-1}\{\widehat{M}(h^{\mathrm{opt}}) - \widehat{M}(\hat{h}^{\mathrm{opt}})\} \to 3\sigma_4^2\chi_1^2
\end{aligned}$$

in distribution. Here the first relation follows by Taylor expansion; the second from (2.23); the third by Taylor expansion and the final limiting result, from (2.18).

2.6. *Pointwise convergence of $\hat{f}$.* Here we show that the condition $h\gamma \to 0$, which was imposed in earlier sections when describing the effects of long-range dependence on stochastic measures of the general performance of $\hat{f}$, also implies that $\hat{f}$ has, locally, the same first-order asymptotic properties that it would enjoy in the case of independent data. Of course, the expected value of $\hat{f}$ depends only on the marginal distribution of $\mathcal{X}$, not on the strength of dependence, so we shall confine attention to local properties of $\hat{f} - E(\hat{f})$.

THEOREM 2.5. *Suppose $\mathcal{X}_\infty$ is a function of a Gaussian process*, (2.1), (2.5) *and* (2.7) *hold*, $h = h(n) \to 0$ *and* $nh \to \infty$, *and* $h\gamma \to 0$. *Then for each* x,

$$(nh)^{1/2}\{\hat{f}(x) - E\hat{f}(x)\} \to N\{0, \sigma(x)^2\} \qquad \textit{in distribution},$$

where $\sigma(x)^2 = f(x)\int K^2$.

The argument that we shall use to derive Theorem 2.5 may be employed to establish a wide variety of related results, for example, about pointwise convergence of the jth derivative of $\hat{f} - E(\hat{f})$. Again, the result asserts that the limit distribution is identical to that under independence. The only additional regularity condition is the assumption that K has j bounded derivatives.

Perhaps of more interest are results which establish the validity of so-called data sharpening methods applied to time-series data. We shall briefly describe them here. Details are given by Hall and Minnotte (2000) in the case of independent data. Let $\widehat{F}(x) = \int_{y \le x} \hat{f}(y)\,dy$ denote the distribution function corresponding to $\hat{f}$, let I represent the identity function and define functions

$$\hat{\gamma}_4 = I + h^2 \frac{\mu_2}{2} \frac{\hat{f}'}{\hat{f}},$$

$$\hat{\gamma}_6 = \hat{\gamma}_4 + h^4 \left\{ \left(\frac{\mu_4}{24} - \frac{\mu_2^2}{2} \right) \frac{\hat{f}^{(3)}}{\hat{f}} + \frac{\mu_2^2}{2} \frac{\hat{f}'' \hat{f}'}{\hat{f}^2} - \frac{\mu_2^2}{8} \frac{(\hat{f}')^3}{\hat{f}^3} \right\},$$

and so on; $\hat{\gamma}_r$ may be defined for any even integer r. Assume the conditions of Theorem 2.5, and in addition that K has sufficiently many bounded derivatives and $(\log n)/(nh^r)$ is bounded as $n \to \infty$. Then the estimator

$$\hat{f}_r(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - \hat{\gamma}_r(X_i)}{h} \right)$$

satisfies

$$E\{\hat{f}_r(x)\} = f(x) + O\{h^r + (nh)^{-1}\},$$

$$\operatorname{var}\{\hat{f}_r(x)\} \sim (nh)^{-1} t_r^2 f(x),$$

$$(nh)^{1/2}\{\hat{f}_r(x) - E\hat{f}_r(x)\} \to N\{0, t_r^2 f(x)\} \qquad \text{in distribution},$$

where $t_r > 0$ depends only on K.

3. Technical arguments.

3.1. *Proof of Theorem* 2.1. The proof is based on that of Theorem 2.1 of Hall, Lahiri and Truong (1995), referred to below as HLT. Specifically, note that the argument given on the lower half of page 2245 of HLT shows that

$$n \int \operatorname{var} \hat{f}(x)\, dx = \int \operatorname{var} K_h(x - X_i) + \tau_{n1} + 2 \sum_{i=1}^{n-1} (1 - n^{-1} i) \int r_{i1}(x)\, dx,$$

where $K_h(u) = h^{-1} K(u/h)$,

$$r_{i\ell}(x) = \iint K(y_1) K(y_2) \biggl\{ g_i(x - hy_1, x - hy_2) - \sum_{k=0}^{\ell} (1/k!) \sum_{j_1, j_1:\ j_1 + j_2 = k} g_i^{(j_1, j_2)}(x, x)(-hy_1)^{j_1} (-hy_2)^{j_2} \biggr\}\, dy_1\, dy_2.$$

To obtain our Theorem 2.1, first carry the expansion of $n \int \operatorname{var} \hat{f}$ to one further term:

$$n \int \operatorname{var} \hat{f}(x)\, dx = \int \operatorname{var} K_h(x - X_i) + \tau_{n1} + \tau_{n2} h^2 + 2 \sum_{i=1}^{n-1} (1 - n^{-1} i) \int r_{i3}(x)\, dx.$$

Then apply the Taylor expansion argument on page 2245 of HLT to r_{i3} rather than r_{i1}, obtaining, in place of the fifth line on page 2246 of HLT,

$$\int \operatorname{var} \hat{f}(x)\, dx = (nh)^{-1} \kappa - n^{-1} I_0 + n^{-1} (\tau_{n1} + \tau_{n2} h^2) + O(n^{-1} h^2 + n^{-1} h^4 \gamma + h^8).$$

A similar but simpler argument leads to the following analogue of (2.11) of HLT:

$$\int (E\hat{f} - f)^2 = h^4 \tfrac{1}{4} \mu^2 I_2 - h^6 \tfrac{1}{24} \mu_2 \mu_4 I_3 + O(h^8).$$

Theorem 2.1 follows on adding the last two displayed formulae.

3.2. *Density expansions for functions of Gaussian processes.* Put $k_0 = 0$. Given an r-vector $v = (1, k_1 + 1, \dots, k_{r-1} + 1)$, where the integers k_i satisfy $0 < k_1 < \cdots < k_{r-1}$, let f_v denote the joint density of $(X_1, X_{k_1+1}, \dots, X_{k_{r-1}+1})$ and let f_{v0}, a product of r marginal densities, be the function that f_v would equal if the components of $\mathcal{X}_\infty$ were independent with the marginal distribution of X_1. Now, f_v may be regarded as as a functional of the sequence of correlations of the Gaussian process of which $\mathcal{X}_\infty$ is a function. Those correlations may in turn be expressed through the correlations of $\mathcal{X}_\infty$ itself, and so we may treat f_v as a functional of the values $\rho(k_i - k_j)$ for $i > j$.

Index these values as $\rho^{(1)}, \dots, \rho^{(p)}$, where $p = r(r-1)/2$. Given an integer $\nu \geq 1$, expand f_v as a Taylor series in $\rho^{(1)}$ about $\rho^{(1)} = 0$, terminating with a term that involves the νth power of $\rho^{(1)}$ and expressing this term using the exact form for the remainder in a Taylor series. Now Taylor-expand each of these $\nu+1$ terms as a series in $\rho^{(2)}$ about $\rho^{(2)}=0$, terminating with the term involving the νth power of $\rho^{(2)}$, and again using the exact form of this term. Next, expand each of the $(\nu+1)^2$ terms as a Taylor series in $\rho^{(3)}$ about $\rho^{(3)}=0$; and so on. This process will produce a series that consists of $(\nu+1)^p$ terms. We now describe the forms of those terms.

If $i_1,\ j_1,\dots,i_\ell,\ j_\ell$ are components of v with the property that each $i_s<j_s$, let $v_+(i_1,\dots,j_\ell)$ be the vector obtained from v by deleting all components except those that appear at least once in the sequence of i_s's and j_s's, and define $v_-(i_1,\dots,j_\ell)$ to be the vector made up of the deleted components of v. Put $x=(x_1,x_{k_1+1},\dots,x_{k_{r-1}+1})$. Then for each v and each integer $\nu\ge 1$, the expansion to $(\nu+1)^p$ terms (and neglecting the remainder), described in the previous paragraph, may be written as

$$\begin{aligned}f_v(x)=f_{v0}(x)+\sum_{\ell\ge 1}\sum_{i_1<j_1}\sum\cdots\sum_{i_\ell<j_\ell}\rho(j_1-i_1)\cdots\rho(j_\ell-i_\ell)\\ \times\psi_{v_+(i_1,\dots,j_\ell)}(x)f_{v_-(i_1,\dots,j_\ell),0}(x),\end{aligned}\tag{3.1}$$

where with a single exception [see (d) below], (a) for $\alpha_u=f_{u,0}$ or ψ_u, the function $\beta=\alpha_{v_\pm(i_1,\dots,j_\ell)}(x)$ depends nondegenerately only on those components of x that are listed in $v_\pm(i_1,\dots,j_\ell)$, (b) for $u=v_-(i_1,\dots,j_\ell)$ (a t-vector, say) the function $f_{u,0}$ denotes the joint density of t independent random variables each having the distribution of X_i and (c)

$$\sup_x|\beta(x)|\le C(r,\nu),\qquad \int|\beta(x)|\,dx\le C(r,\nu),\tag{3.2}$$

where $C(r,\nu)$ depends on r but not otherwise on v. The exceptional term in the expansion at (3.1) is one that has the form $\chi(x)\prod_{1\le i\le p}(\rho^{(i)})^\nu$, where (d) the function $\beta=\chi$ satisfies (3.2). Result (3.1) is linked to (2.8) in that, in the special case $r=2$ where $k_1=i$, $g_i(x_1,x_2)=f_v(x)-f_{v0}(x)$.

3.3. *Proof of Theorem* 2.2. Sections 3.3.1–3.3.5 will derive result (2.16), and Section 3.3.6 will outline the simpler proof of (2.15).

3.3.1. *Decomposition of* $\widehat{M}'(h)$. Put

$$\widehat{M}_1(h)\equiv-\tfrac{1}{2}h\widehat{M}'(h)=\int(\hat f_h-f)(\hat f_h-\hat g_h),\tag{3.3}$$

where $\hat g_h(x)=(nh)^{-1}\sum_i K_1\{(x-X_i)/h\}$ and $K_1(x)=-xK'(x)$. Define

$$K_j(x)=\int K(u)L(u+x)\,du$$

for $(j, L) = (2, K)$ and $(j, L) = (3, K_1)$, and let $K_4 = K_2 - K_3$,

$$k_4(x) = E\left\{K_4\left(\frac{x - X_i}{h}\right)\right\}, \qquad \kappa_4 = E\{k_4(X_i)\},$$

$$T_1 = (n^2h)^{-1} \sum\sum_{i \neq j} \left\{K_4\left(\frac{X_i - X_j}{h}\right) - k_4(X_i) - k_4(X_j) + \kappa_4\right\},$$

$$k_5(x) = \int \{2(1 - n^{-1})K_4(u) + K_1(u) - K(u)\} f(x - hu)\, du,$$

$$\kappa_5 = E\{k_5(X_i)\},$$

$$T_2 = n^{-1} \sum_{i=1}^{n} \{k_5(X_i) - \kappa_5\},$$

$$\kappa_0 = \iint \{(1 - n^{-1})K_4(u) + K_1(u) - K(u)\} f(v) f(v - hu)\, du\, dv.$$

Both T_1 and T_2 are functions of h. In this notation,

$$\widehat{M}_1(h) = T_1(h) + T_2(h) + (nh)^{-1}K_4(0) + \kappa_0. \tag{3.4}$$

3.3.2. *Bound for* $T_1(h)$. Put $t_1(X_i, X_j) = K_4\{(X_i - X_j)/h\} - k_4(X_i) - k_4(X_j) + \kappa_4$, and note that

$$\begin{aligned}(n^2h)^2 E(T_1^2) = O\Bigg[&\sum\sum_{i \neq j} E\{t_1(X_i, X_j)^2\} \\ &+ \sum_{i, j_1, j_2}\sum\sum_{\text{all distinct}} \left|E\{t_1(X_i, X_{j_1}) t_1(X_i, X_{j_2})\}\right| \\ &+ \sum_{i_1, j_1, i_2, j_2}\sum\sum\sum_{\text{all distinct}} \left|E\{t_1(X_{i_1}, X_{j_1}) t_1(X_{i_2}, X_{j_2})\}\right|\Bigg].\end{aligned} \tag{3.5}$$

To bound the double series on the right-hand side of (3.5), assume without loss of generality that $0 < i < j$ and put $v = (0, j - i)$. Then f_v is the joint density of (X_i, X_j), and

$$\begin{aligned}E\{t_1(X_i, X_j)^2\} &= h \iint t_1(x_1, x_1 - hx_2)^2 f_v(x_1, x_1 - hx_2)\, dx_1\, dx_2 \\ &= O(h),\end{aligned} \tag{3.6}$$

uniformly in $i \neq j$. [The second identity at (3.6) follows after squaring the right-hand side of the quantity

$$t_1(x_1, x_1 - hx_2) = K_4(x_2) - k_4(x_1) - k_4(x_1 - hx_2) + \kappa_4, \tag{3.7}$$

integrating term by term, and noting that $\sup_x |k_4(x)| = O(h)$.] It follows that the double series on the right-hand side of (3.5) equals $O(n^2 h)$.

To bound the triple series on the right-hand side of (3.5), assume without loss of generality that $0 < j_1 - i < j_2 - i$ and take $v = (0, j_1 - i, j_2 - i)$. Then

$$\begin{aligned} E\{t_1(X_i, X_{j_1}) t_1(X_i, X_{j_2})\} = h^2 \iiint & t_1(x_1, x_1 - hx_2) t_1(x_1, x_1 - hx_3) \\ & \times f_v(x_1, x_1 - hx_2, x_1 - hx_3)\, dx_1\, dx_2\, dx_3. \end{aligned} \tag{3.8}$$

Approximate f_v using the expansion at (3.1) for arbitrarily large ν. Also, expand $t_1(x_1, x_1 - hx_j)$ (for $j = 2, 3$) using (3.7), and multiply out these parts of the integrand at (3.8). Take the absolute value of the resulting expression for the right-hand side of (3.8), and sum over distinct values of i, j_1, j_2. Two of these three series each contribute a factor n to a bound for the triple series, while the third contributes a factor γ, by virtue of the properties: (a) the right-hand side of (3.8) vanishes if we replace f_v there by $f_{v,0}$, and (b) all other terms in the expansion (3.1) of f_v involve at least one factor ρ_i. Therefore, the triple series on the right-hand side of (3.5) equals $O(n^2 \gamma h^2)$.

The same argument, applied to bound the fourfold series on the right-hand side of (3.5), gives the bound $O(n^3 \gamma h^2)$. However, this can be improved upon as follows. If the vector $v_+(i_1, \dots, j_\ell)$ at (3.1) has four components, then there are at least two distinct terms $\rho(j_s - i_s)$ in the corresponding product at (3.1), and as a result, when summing over all four distinct indices in the quadruple series at (3.5), two (rather than just one) of the factors n may be replaced by γ. Therefore, the total contribution of terms such as these equals $O(n^2 \gamma^2 h^2)$ rather than $O(n^3 \gamma h^2)$. In the remaining case, where $v_+(i_1, \dots, j_\ell)$ at (3.1) has three or fewer components, the vector $v_-(i_1, \dots, j_\ell)$ must have one or more components (since v has itself exactly four components). Multiplying $t_1(x_1, x_2) t_1(x_3, x_4)$ by the marginal density of this component (i.e., the marginal density of one of $x_1, \dots, x_4$), and integrating over that component, we obtain zero, since the definition of $t_1(x_1, x_2)$ ensures that $E\{t_1(X, x)\} = E\{t_1(x, X)\} = 0$. There is an exceptional case, corresponding to (d) in the account in Section 3.1 of terms in the expansion (3.1). However, since $\rho_i = O(i^{-\varepsilon})$ for some $\varepsilon > 0$, then, provided $\nu > 1/\varepsilon$, the exceptional case contributes a term of order $O\{n^2 (\sum |\rho_i|^\nu)^2 h^2\} = O(n^2 h^2)$.

Combining these bounds we deduce from (3.5) that

$$(n^2 h)^2 E\{T_1(h)^2\} = O(n^2 h + n^2 \gamma h^2 + n^2 \gamma^2 h^2 + n^2 h^2) = o(n^2 \gamma h),$$

the latter identity holding provided $\gamma \to \infty$ and $\gamma h \to 0$. Hence, if these conditions hold,

$$T_1(h) = o_p\{(\gamma / n^2 h)^{1/2}\}. \tag{3.9}$$

3.3.3. *Limit theorem for $T_2(h)$.* We use the method of moments to prove that T_2 is asymptotically Normally distributed. Note that

$$k_5(x) = -\mu_2 h^2 f''(x) + o(h^2)$$

uniformly in x. If we define $t_3(x) = f''(x) + \int (f')^2$ and $T_3(h) = n^{-1} \sum_i t_3(X_i)$, then the mth moment of $T_2(h)$ will equal that of $-\mu_2 h^2 T_3(h)$, up to terms that are of smaller order than $\{h^2(\gamma/n)^{1/2}\}^m$.

Consider writing $E\{nT_3(h)\}^m$ as the sum, over $1 \le r \le m$, of series of the form

$$\sum_{\substack{i_1,\dots,i_m \text{ such that} \\ \text{exactly } r \text{ are distinct}}} \cdots \sum \int \cdots \int t_3(x_{i_1}) \cdots t_3(x_{i_m}) f_v(x)\, dx, \tag{3.10}$$

where, if $1 = k_0 + 1 < \cdots < k_{r-1} + 1$ are the distinct elements of the sequence $i_1, \dots, i_{2m}$, then $v = (0, k_1, \dots, k_{r-1})$. Expanding f_v as in (3.1), and interchanging the order of the summation and integration in the integral immediately above, we deduce that the series equals $O(n^r)$ if $r < m/2$, and equals $O(n^{m/2}\gamma_1 \cdots \gamma_{m/2})$ if $r > m/2$ and m is even, where each γ_j has the form $\sum_{i \le n} |\rho_i|^s$ for some integer $s = s(j) \ge 1$, and $s(j) \ge 2$ for at least one j. If $r \ge (m+1)/2$ and $m-1$ is even, then the series equals $O(n^{(m-1)/2}\gamma_1 \cdots \gamma_{(m-1)/2})$, for the same interpretation of γ_j. Define

$$q_1 = E[f'''\{a(N)\}a'(N)]/E\{a'(N)\}.$$

If m is even and $r = m/2$, then standard arguments in a proof of the central limit theorem by the method of moments may be used to show that the series in (3.10) is asymptotic to $n^{m/2} q_2^m E(N^m)$, where the random variable N has a standard Normal distribution,

$$\begin{aligned} q_2^2 &= 2\sum_{i=1}^{n} \operatorname{cov}\{t_3(X_1), t_3(X_{i+1})\} = 2\sum_{i=1}^{n} \tau_i \big[E\{t_3(a)\}'(N)\big]^2 + o(\gamma) \\ &= 2\gamma \frac{[E\{t_3(a)\}'(N)]^2}{\{Ea'(N)\}^2} + o(\gamma) = 2\gamma q_1^2 + o(\gamma), \end{aligned}$$

and we have used (2.6) to derive asymptotic formulae for correlations. Hence,

$$h^{-2}(n/\gamma)^{1/2} T_2(h) \to N(0, 2\mu_2^2 q_1^2) \quad \text{in distribution.} \tag{3.11}$$

3.3.4. *Decomposition of $\widehat{M}''(h)$.* Put

$$\widehat{M}_2(h) \equiv \tfrac{1}{2} h^2 \widehat{M}''(h) = \int \{(\hat{f}_h - \hat{g}_h)^2 + (\hat{f}_h - f)(2\hat{f}_h - \hat{g}_h - \hat{e}_h)\}, \tag{3.12}$$

where $\hat{e}_h(x) = (nh)^{-1} \sum_i K_5\{(x - X_i)/h\}$ and $K_5(x) = -xK_1'(x)$. This formula is analogous to that for $\widehat{M}_2(h)$ at (3.3), and in fact the methods of Sections 3.3.1–3.3.3 may be used to show that, for all integer $k \ge 1$,

$$E\{\widehat{M}_2(h) - E\widehat{M}_2(h)\}^{2k} = O\{(\gamma n^{-9/5})^k\},$$

uniformly in $h \in \mathcal{H}(C_1, C_2) \equiv [n^{-1/5}C_1, n^{-1/5}C_2]$ for any $0 < C_1 < C_2 < \infty$. Furthermore,

$$E\{\widehat{M}_2(h)\} = M_2(h) + O(n^{-1})$$

uniformly in $h \in \mathcal{H}$, where $M_2(h) = (nh)^{-1}\kappa + \frac{3}{2}h^4\mu_2^2 I_2$. Therefore, by Markov's inequality and the fact that $\gamma = O(n^{(1/5)-\varepsilon})$ for some $\varepsilon > 0$,

$$\sup_{h\in\mathcal{H}} P\{|\widehat{M}_2(h) - M_2(h)| > \varepsilon n^{-4/5}\} = O(n^{-\lambda})$$

for all ε, $\lambda > 0$. We may deduce from this result and the Hölder continuity of K'' that

$$\sup_{h\in\mathcal{H}} |\widehat{M}_2(h) - M_2(h)| = o(n^{-4/5}) \tag{3.13}$$

with probability 1.

3.3.5. *Completion of proof of* (2.16). (It is notationally simpler if we confine attention to the case $q_1 \neq 0$, which we do below.) A subsidiary argument may be used to prove that, for constants $0 < C_1 < C_2 < \infty$, $P\{\hat{h}^{\mathrm{opt}} \in \mathcal{H}(C_1, C_2)\} \to 1$ as $n \to \infty$. Without loss of generality, C_1 and C_2 also have the property that $h^{\mathrm{opt}} \in \mathcal{H}(C_1, C_2)$ for all sufficiently large n. By the Taylor expansion,

$$0 = \widehat{M}'(\hat{h}^{\mathrm{opt}}) = \widehat{M}'(h^{\mathrm{opt}}) + (\hat{h}^{\mathrm{opt}} - h^{\mathrm{opt}})\widehat{M}''(h^*), \tag{3.14}$$

where h^* lies between h^{opt} and $\hat{h}^{\mathrm{opt}}$. Therefore the probability that $h^* \in \mathcal{H}(C_1, C_2)$ converges to 1, and so, by (3.13),

$$\widehat{M}''(h^*) = \{1 + o_p(1)\}2(h^*)^{-2}M_2(h^*). \tag{3.15}$$

Now, it may be deduced from (3.4), (3.9) and the property $E\{\widehat{M}'(h^{\mathrm{opt}})\} = M'(h^{\mathrm{opt}}) = 0$ that

$$\widehat{M}'(h^{\mathrm{opt}}) = 2^{3/2}\mu_2 q_1 h^{\mathrm{opt}}(\gamma/n)^{1/2} Z_1, \tag{3.16}$$

where for each j, Z_j will denote an asymptotically standard Normal random variable. Combining (3.14)–(3.16) we deduce that

$$(\hat{h}^{\mathrm{opt}} - h^{\mathrm{opt}})2(h^*)^{-2}M_2(h^*) = 2^{3/2}\mu_2 q_1 h^{\mathrm{opt}}(\gamma/n)^{1/2} Z_2. \tag{3.17}$$

In particular this implies that $\hat{h}^{\mathrm{opt}} - h^{\mathrm{opt}} = O_p\{(\gamma/n^{3/5})^{1/2}\}$, and so $h^* = \{1 + o_p(1)\}h^{\mathrm{opt}}$. Therefore, by (3.17) again,

$$\begin{aligned}\hat{h}^{\mathrm{opt}} - h^{\mathrm{opt}} &= 2^{1/2}\mu_2 q_1 (h^{\mathrm{opt}})^3\{M_2(h^{\mathrm{opt}})\}^{-1}(\gamma/n)^{1/2} Z_3 \\ &= 2^{3/2}\mu_2 q_1 h^{\mathrm{opt}}\{M''(h^{\mathrm{opt}})\}^{-1}(\gamma/n)^{1/2} Z_4.\end{aligned}$$

This implies (2.16).

3.3.6. *Outline proof of* (2.15). Let f_{ij} denote the joint density of X_1 and X_j. When (2.10) holds, a slight elaboration of the argument in Section 3.3.2 shows that

$$\begin{aligned}(n^2h)^2 \operatorname{var}(T_1) &= 4 \sum_{i<j}\sum E\{t_1(X_i, X_j)^2\} + O(n^2h^2) \\ &\sim 4h \sum_{i<j}\sum \iint K_4(x_2)^2 f_{ij}(x_1, x_1)\,dx_1\,dx_2 \\ &\sim 2n^2h\left(\int f^2\right)\left(\int K_4^2\right).\end{aligned}$$

Furthermore,

$$\operatorname{var}(T_2) \sim (\mu_2 h^2)^2 \operatorname{var}\left\{n^{-1}\sum_{i=1}^{n} f''(X_i)\right\} \sim n^{-1}(\mu_2 h^2)^2 s_2^2,$$

and $\operatorname{cov}(T_1, T_2) = o\{(n^2h)^{-1} + n^{-1}h^4\}$. Therefore,

$$n^{7/5} \operatorname{var}\{\widehat{M}'(h^{\mathrm{opt}})\} \sim 2c_0^{-3}\left(\int f^2\right)\left(\int K_4^2\right) + c_0^2\mu_2^2 s_2^2 = (c_0 c_3 \sigma_1/2)^2.$$

A longer argument will show that all the moments of $n^{7/10}\widehat{M}'(h^{\mathrm{opt}})$ converge to the respective moments of a Normal random variable with zero mean and variance $(c_0c_3\sigma_1/2)^2$. Therefore, $n^{7/10}\widehat{M}'(h^{\mathrm{opt}})$ has this limit distribution, and so, by an argument similar to that in Section 3.3.5,

$$(\hat{h}^{\mathrm{opt}} - h^{\mathrm{opt}})c_3 n^{-2/5} = n^{-7/10} c_0 c_3 \sigma_1 Z_1,$$

where Z_1 is asymptotically Normal $N(0, 1)$. This result is equivalent to (2.15).

3.4. *Proof of Theorem* 2.4. Define

$$\widehat{D}(h) = \frac{1}{nh}\sum_{i=1}^{n}\int K\left(\frac{x - X_i}{h}\right) f(x)\,dx - \frac{1}{nn_\ell h}\sum_{i=1}^{n}\sum_{j:\,|i-j|>\ell} K\left(\frac{X_i - X_j}{h}\right),$$

and put $K_6 = K - K_1$, $k_6(x) = E[K_6\{(x - X_i)/h\}]$, $\kappa_6 = E\{k_6(X_i)\}$,

$$N_i = \#\{j\colon |i - j| > \ell \text{ and } 1 \le j \le n\},$$

$$T_3(h) = \frac{1}{nn_\ell h}\sum_{i=1}^{n}\sum_{j:\,|i-j|>\ell}\left\{K_6\left(\frac{X_i - X_j}{h}\right) - k_6(X_i) - k_6(X_j) + \kappa_6\right\},$$

$$T_4(h) = \frac{1}{nh}\sum_{i=1}^{n}\{k_6(X_i) - \kappa_6\}\left\{\frac{2N_i}{n_\ell} - 1\right\}.$$

In this notation,

$$\widehat{D}_1(h) \equiv -h\widehat{D}'(h) = T_3(h) + T_4(h).$$

The terms T_3 and T_4 are analogous, in the present setting, to T_1 and T_2 introduced in Section 3.3.1.

Assume first that (2.10) holds. Note that $\sup_i |(2N_i n_\ell^{-1} - 1) - 1| \to 0$ and

$$h^{-1} k_6(x) = \int K_6(u) f(x - hu)\, du = -\mu_2 h^2 f''(x) + o(h^2).$$

Using these properties we may derive asymptotic formulae for the variances of T_3 and T_4, and their covariance, as was done in Section 3.3.6 for T_1 and T_2. In this manner we may show that

$$\operatorname{var}\{\widehat{D}_1(h)\} \sim 2(n^2 h)^{-1} \left(\int f^2 \right) \left(\int K_6^2 \right) + n^{-1} h^4 \mu_2^2 s_2^2,$$

and thence that

$$n^{7/5} \operatorname{var}\{\widehat{D}'(h^{\mathrm{opt}})\} \sim 2c_0^{-3} \left(\int f^2 \right) \left(\int K_6^2 \right) + c_0^2 \mu_2^2 s_2^2 = (c_0 c_3 \sigma_5 / 2)^2.$$

As in Section 3.3.6, a longer argument based on the method of moments will show that the distribution of $n^{7/10} \widehat{D}'(h^{\mathrm{opt}})$ is asymptotically Normal with zero mean and variance $(c_0 c_3 \sigma_5 / 2)^2$.

Put $V = \widehat{M} - M$. Subsidiary arguments may be used to prove that $\hat{h}^{\mathrm{cv}} / h^{\mathrm{opt}} \to 1$ in probability and $|V''(h)| + |\widehat{D}''(h)| = o_p(n^{-2/5})$, uniformly in $h \in \mathcal{H}(C_1, C_2)$ for any $0 < C_1 < C_2 < \infty$. (The argument leading to the latter property is similar to that in Section 3.3.4.) Note too that $M''(h^{\mathrm{opt}}) \sim c_3 n^{-2/5}$ and

$$CV + \int f^2 = M + V + 2\widehat{D}.$$

Therefore,

$$\begin{aligned} 0 = CV'(\hat{h}^{\mathrm{cv}}) &= M'(\hat{h}^{\mathrm{cv}}) + V'(\hat{h}^{\mathrm{cv}}) + 2\widehat{D}'(\hat{h}^{\mathrm{cv}}) \\ &= (\hat{h}^{\mathrm{cv}} - h^{\mathrm{opt}}) M''(\hat{h}_1^*) + V'(h^{\mathrm{opt}}) + (\hat{h}^{\mathrm{cv}} - h^{\mathrm{opt}}) V''(\hat{h}_2^*) \\ &\quad + 2\widehat{D}'(h^{\mathrm{opt}}) + 2(\hat{h}^{\mathrm{cv}} - h^{\mathrm{opt}}) \widehat{D}''(\hat{h}_3^*) \\ &= \{1 + o_p(1)\} (\hat{h}^{\mathrm{cv}} - h^{\mathrm{opt}}) c_3 n^{-2/5} + V'(h^{\mathrm{opt}}) + 2\widehat{D}'(h^{\mathrm{opt}}). \end{aligned}$$

A similar argument gives

$$0 = \{1 + o_p(1)\} (\hat{h}^{\mathrm{opt}} - h^{\mathrm{opt}}) c_3 n^{-2/5} + V'(h^{\mathrm{opt}});$$

compare (3.14). Subtracting, and noting that $\hat{h}^{\mathrm{opt}} - h^{\mathrm{opt}} = O_p(n^{-3/10})$, by (2.4), we deduce that

$$(\hat{h}^{\mathrm{cv}} - h^{\mathrm{opt}}) c_3 n^{-2/5} = -2\{1 + o_p(1)\} \widehat{D}'(h^{\mathrm{opt}}) + o_p(n^{-7/10}). \tag{3.18}$$

This formula, and the limit result given in the previous paragraph, imply (2.19).

Now suppose (2.10) fails. Then, arguing as in Sections 3.3.1–3.3.5, we may show that the conclusion reached in the second paragraph of the present section changes to

$$\widehat{D}'(h^{\mathrm{opt}}) = 2^{1/2}\mu_2 q_1 h^{\mathrm{opt}}(\gamma/n)^{1/2} Z, \tag{3.19}$$

where Z is asymptotically Normal $N(0,1)$. Furthermore, the argument leading to (3.18) now gives instead

$$(\hat{h}^{\mathrm{cv}} - h^{\mathrm{opt}})c_3 n^{-2/5} = -2\{1 + o_p(1)\}\widehat{D}'(h^{\mathrm{opt}}) + o_p(n^{-7/10}\gamma^{1/2}). \tag{3.20}$$

Result (2.21) follows from (3.19) and (3.20).

Tracing the origins of the dominant terms in $\widehat{D}'(h^{\mathrm{opt}})$ and $\widehat{M}'(h^{\mathrm{opt}})$ in Sections 3.3 and 3.4 we may show that, when (2.10) fails, $\widehat{D}'(h^{\mathrm{opt}}) = -\widehat{M}'(h^{\mathrm{opt}}) + o_p(n^{-7/10}\gamma^{1/2})$. Result (2.23) follows from this property, (3.14) and (3.20). Results (2.20) and (2.22) follow via a simple Taylor expansion argument.

3.5. *Proof of Theorem* 2.5. We shall use the method of moments. Put $\Delta = n\{\hat{f}(x) - E\hat{f}(x)\}$, $u_0 = E\{\hat{f}(x)\}$ and $u_1(y) = h^{-1}K\{(x-y)/h\} - u_0$, and let $m \ge 1$ be an integer. Then

$$E(\Delta^m) = \sum_{r=1}^{m} \mathop{\sum\cdots\sum}_{\substack{p_1,\dots,p_m \text{ such that} \\ \text{exactly } r \text{ are distinct}}} E\{u_1(X_{p_1})\cdots u_1(X_{p_m})\}. \tag{3.21}$$

If $r < m/2$, then the expected value on the right-hand side equals $O(h^{r-m})$, uniformly in sequences $i_1,\dots,i_m$ with just r distinct components, and so the contribution to the right-hand side equals

$$O(n^r h^{r-m}) = O\{(nh)^r h^{-m}\} = o\{(nh)^{m/2} h^{-m}\}.$$

To treat the case $r > m/2$ we expand the joint density of the r distinct components of $(X_{p_1},\dots,X_{p_m})$, as in (3.1). Taking a particular term in this expansion, say

$$u_3(y) = \rho(j_1 - i_1)\cdots\rho(j_\ell - i_\ell)\psi_{v_+(i_1,\dots,j_\ell)}(y) f_{v_-(i_1,\dots,j_\ell),0}(y), \tag{3.22}$$

where $y = (y_1,\dots,y_r)$, we write the contribution to $E\{n\hat{f}(x)\}^m$ as

$$\begin{aligned} &\int\cdots\int u_1(y_1)^{\xi_1}\cdots u_1(y_r)^{\xi_r} u_3(y)\,dy \\ &\quad = h^{r-m}\int\cdots\int \{K(y_1) - hu_0\}^{\xi_1}\cdots\{K(y_r) - hu_0\}^{\xi_r} \\ &\qquad\qquad \times u_3(x - hy_1,\dots,x - hy_r)\,dy_1\cdots dy_r, \end{aligned}$$

where each $\xi_i \ge 1$ and $\xi_1 + \cdots + \xi_r = m$. Performing, for this contribution to $E(\Delta^m)$ and this value of r, the summation over $p_1,\dots,p_m$ at (3.21), and noting that at least $r-(m/2)$ (for r even) or $r-\{(m-1)/2\}$ (for r odd) of the summations are effectively over one of the indices in one of the components $\rho(j_k - i_k)$ in the product at (3.22), and so produce a quantity that equals $O(\gamma)$ rather than $O(n)$, we deduce that the contribution to $E(\Delta^m)$ equals

$$O(h^{r-m}n^{m/2}\gamma^{r-(m/2)}) = O\{(n/h)^{m/2}(h\gamma)^{r-(m/2)}\}$$

in the case of even r, and

$$O(h^{r-m}n^{(m-1)/2}\gamma^{r-\{(m-1)/2\}}) = O\{(nh)^{(m-1)/2}h^{-m}(h\gamma)^{r-\{(m-1)/2\}}\}$$

for odd r. Combining this result with that derived in the previous paragraph, and noting that $h\gamma \to 0$, we deduce that, for even m,

$$E(\Delta^m) = \sum_{\substack{p_1,\dots,p_m \text{ such that} \\ \text{exactly } m/2 \text{ are distinct}}} \!\!\cdots\!\! \sum E\{u_1(X_{p_1})\cdots u_1(X_{p_m})\} + o\{(n/h)^{m/2}\}, \tag{3.23}$$

while for odd m,

$$E(\Delta^m) = o\{(n/h)^{m/2}\}. \tag{3.24}$$

Take $m = 2r$ for an integer $r \ge 1$, and suppose there are exactly r distinct values among $p_1, \dots, p_m$. Let these be $q_1 < \cdots < q_r$, put $q = (q_1, \dots, q_r)$ and let f_q denote the joint density of $X_{q_1}, \dots, X_{q_r}$. Then

$$\begin{aligned} E\{u_1(X_{p_1})\cdots u_1(X_{p_m})\} &= E\{u_1(X_{q_1})^2\cdots u_1(X_{q_r})^2\} \\ &= h^{-r}\int \{K(y_1) - hu_0\}^2\cdots\{K(y_r) - hu_0\}^2 \\ &\qquad \times f_q(x - hy_1, \dots, x - hy_r)\,dy_1\cdots dy_r \\ &\sim \left(h^{-1}\int K^2\right)^r f_q(x, \dots, x). \end{aligned}$$

Hence, by (3.23),

$$E(\Delta^m) = \left(h^{-1}\int K^2\right)^r \sum_{\substack{p_1,\dots,p_m \text{ such that} \\ \text{exactly } r \text{ are distinct}}} \!\!\cdots\!\! \sum f_q(x, \dots, x) + o\{(n/h)^r\}. \tag{3.25}$$

Still considering the case $m = 2r$, let $f_0(x, \dots, x) = f(x)^r$ denote the value that $f_q(x, \dots, x)$ would take if the components of the time series $\mathcal{X}$ were totally independent. Given $\ell \ge 1$, let $\mathcal{Q}(\ell)$ be the set of vectors q, among all those involved in the summation in (3.25), for which none of the absolute values of the lags between any two components is strictly less than ℓ. For each fixed ℓ,

$$\{\#\mathcal{Q}(\ell)\}/\{\#\mathcal{Q}(1)\} \to 1$$

as $n \to \infty$, and

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} \sup_{q\in\mathcal{Q}(\ell)} \left| f_q(x,\dots,x) - f_0(x,\dots,x) \right| \to 0.$$

These two results imply the following simplification of (3.25):

$$E(\Delta^m) = \left\{ h^{-1} f(x) \int K^2 \right\}^r \sum_{\substack{p_1,\dots,p_m \text{ such that} \\ \text{exactly } r \text{ are distinct}}} \cdots \sum 1 + o\{(n/h)^r\}. \tag{3.26}$$

The expansion on the right-hand side of (3.26) does not involve the dependence structure of $\mathcal{X}$, and in the case of total independence, where (3.26) is also valid, it is of course known that the mth moment of Δ^m is asymptotic to that of its Normal limiting distribution:

$$E(\Delta^m) \sim \left\{ (n/h) f(x) \int K^2 \right\}^r E(N^m),$$

where N is a standard Normal random variable. Hence, (3.26) implies that for a general time series $\mathcal{X}$ satisfying our regularity conditions, and for even integers $m \ge 2$,

$$E(\Delta^m) = \left\{ (n/h) f(x) \int K^2 \right\}^{m/2} E(N^m) + o\{(n/h)^{m/2}\}.$$

This result and (3.24) imply that all the moments of $(h/n)^{1/2}\Delta$ converge to those of a Normal $N\{0, \sigma(x)^2\}$ random variable, and so the distribution of $(h/n)^{1/2}\Delta$ converges to this limit.

Acknowledgment. We are grateful for the helpful comments of four readers of our manuscript.

REFERENCES

AHMAD, I. A. (1982). Integrated mean square properties of density estimation by orthogonal series methods for dependent variables. *Ann. Inst. Statist. Math.* **34** 339–350.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, London.

CASTELLANA, J. V. (1989). Integrated consistency of smoothed probability density estimators for stationary sequences. *Stochastic Process. Appl.* **33** 335–346.

CASTELLANA, J. V. and LEADBETTER, M. R. (1986). On smoothed probability density estimation for stationary processes. *Stochastic Process. Appl.* **21** 179–193.

CHENG, B. and ROBINSON, P. M. (1991). Density estimation in strongly dependent nonlinear time series. *Statist. Sinica* **1** 335–359.

CSÖRGŐ, S. and MIELNICZUK, J. (1995). Density estimation under long-range dependence. *Ann. Statist.* **23** 990–999.

HALL, P. (1997). Defining and measuring long-range dependence. In *Nonlinear Dynamics and Time Series: Building a Bridge between the Natural and Statistical Sciences* (C. D. Cutler and D. T. Kaplan, eds.) 153–160. Amer. Math. Soc., Providence, RI.

HALL, P. and HART, J.D. (1990). Nonparametric regression with long-range dependence. *Stochastic Process. Appl.* **36** 339–351.

HALL, P. and JOHNSTONE, I. (1992). Empirical functionals and efficient smoothing parameter selection (with discussion). *J. Roy. Statist. Soc. Ser. B* **54** 475–531.

HALL, P., LAHIRI, S. N. and TRUONG, Y. K. (1995). On bandwidth choice for density estimation with dependent data. *Ann. Statist.* **23** 2241–2263.

HALL, P. and MARRON, J. S. (1987). Extent to which least-squares cross-validation minimizes integrated square error in nonparametric density estimation. *Probab. Theory Related Fields* **74** 567–581.

HALL, P. and MINNOTTE, M. C. (2000). High-order data sharpening for density estimation. Manuscript.

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

HÄRDLE, W., LÜTKEPOHL, H. and CHEN, R. (1997). A review of nonparametric time series analysis. *Internat. Statist. Rev.* **65** 49–72.

HART, J. D. (1996). Some automated methods of smoothing time-dependent data. *J. Nonparametr. Statist.* **6** 115–142.

HART, J. D. and VIEU, P. (1990). Data-driven bandwidth choice for density estimation based on dependent data. *Ann. Statist.* **18** 873–890.

KIM, T. Y. and COX, D. D. (1997). A study of bandwidth selection in density estimation under dependence. *J. Multivariate Anal.* **62** 190–203.

MASRY, E. and FAN, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scand. J. Statist.* **24** 165–179.

NGUYEN, H. T. (1979). Density estimation in a continuous-time stationary Markov process. *Ann. Statist.* **7** 341–348.

PRAKASA RAO, B. L. S. (1978). Density estimation for Markov processes using delta sequences. *Ann. Inst. Statist. Math.* **30** 321–328.

ROBINSON, P. M. (1991). Nonparametric function estimation for long memory time series. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the 5th International Symposium in Economic Theory and Econometrics* (W. W. Barnett, J. Powell and G. E. Tauchen, eds.) 437–457. Cambridge Univ. Press.

ROSENBLATT, M. (1970). Density estimates and Markov sequences. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 199–213. Cambridge Univ. Press.

ROUSSAS, G. (1969). Nonparametric estimation in Markov processes. *Ann. Inst. Statist. Math.* **21** 73–87.

ROUSSAS, G. (1990). Asymptotic normality of the kernel estimate under dependence conditions: application to hazard rate. *J. Statist. Plann. Inference* **25** 81–104.

ROUSSAS, G. and IOANNIDES, D. A. (1987). Note on the uniform convergence of density estimates for mixing random variables. *Statist. Probab. Lett.* **5** 279–285.

TAQQU, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287–302.

TRAN, L. T. (1989). The L^1 convergence of kernel density estimates under dependence. *Canad. J. Statist.* **17** 197–208.

TRAN, L. T. (1990). Kernel density estimation under dependence. *Statist. Probab. Lett.* **10** 193–201.

DEPARTMENT OF STATISTICS
TEXAS A&M UNIVERSITY
3143 TAMU
COLLEGE STATION, TEXAS 77843-3143

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA, ACT 0200
AUSTRALIA
E-MAIL: halpstat@pretty.anu.edu.au