Wavelet Methods for Curve Estimation

A. ANTONIADIS, G. GREGOIRE and I. W. MCKEAGUE*

The theory of wavelets is a developing branch of mathematics with a wide range of potential applications. Compactly supported wavelets are particularly interesting because of their natural ability to represent data with intrinsically local properties. They are useful for the detection of edges and singularities in image and sound analysis and for data compression. But most of the wavelet-based procedures currently available do not explicitly account for the presence of noise in the data. A discussion of how this can be done in the setting of some simple nonparametric curve estimation problems is given. Wavelet analogies of some familiar kernel and orthogonal series estimators are introduced, and their finite sample and asymptotic properties are studied. We discover that there is a fundamental instability in the asymptotic variance of wavelet estimators caused by the lack of translation invariance of the wavelet transform. This is related to the properties of certain lacunary sequences. The practical consequences of this instability are assessed.

KEY WORDS: Delta sequences; Hazard rate; Kernel smoothing; Multiresolution analysis; Nonparametric regression; Orthogonal series.

1. INTRODUCTION

Wavelet theory has the potential to provide statisticians with powerful new techniques for nonparametric inference. It combines recent advances in approximation theory with insights gained from applied signal analysis (for a recent survey on the use of wavelets in signal processing, see Rioul and Vetterli 1991; for a recent discussion connecting wavelets with problems in nonparametric statistical inference, see Wegman 1991). The mathematical side of wavelet theory has been developed by Yves Meyer (1990) and his coworkers in a long series of papers (see, for example, Daubechies 1990, Mallat 1989, or, for a concise survey, Strang 1989).

Consider the following standard nonparametric regression model involving an unknown regression function r:

$$Y_i = r(X_i) + \varepsilon_i, \qquad i = 1, \ldots, n.$$

Two versions of this model are distinguished in the literature:

1. The *fixed design* model, in which the X_i's are nonrandom design points (in this case the X_i's are denoted by t_i and taken to be ordered $0 \le t_1 \le \cdots \le t_n \le 1$), with the observation errors ε_i iid with mean zero and variance σ^2
2. the *random design* model, in which the (X_i, Y_i)'s are independent and distributed as (X, Y), with $r(x) = \mathbb{E}(Y \mid X = x)$ and $\varepsilon_i = Y_i - r(X_i)$.

In each case the problem is to estimate the regression function $r(t)$ for $0 < t < 1$. We shall introduce wavelet versions of the most frequently used kernel and orthogonal series estimators for these models, as well as for the problem of hazard rate estimation in survival analysis. Our estimators are delta sequence smoothers based on wavelet kernels $E_m(\cdot, \cdot)$, as defined by Meyer (1990). These kernels represent integral operators E_m that project onto closed subspaces V_m of $L^2(\mathbb{R})$. The increasing sequence of subspaces V_m form a so-called *multiresolution analysis* of $L^2(\mathbb{R})$. The basic idea (to be discussed at greater length in Sec. 2) is that the V_m provide successive approximations, with details being added as m increases. Thus m acts as a tuning parameter, much as the bandwidth does for standard kernel smoothers. A key aspect of wavelet estimators is that the tuning parameter ranges over a much more limited set of values than is common with other nonparametric regression techniques. In practice only a small number of values of m (say three or four) need to be considered. Despite this lack of control through a tuning parameter, which is in fact an advantage when it comes to cross-validation, we shall see that wavelet estimators can compete effectively.

For the fixed design model, we propose the estimator

$$\hat{r}(t) = \sum_{i=1}^{n} Y_i \int_{A_i} E_m(t, s)\, ds,$$

where the A_i are intervals that partition [0, 1] with $t_i \in A_i$. This is a wavelet version of Gasser and Müller's (1979) (convolution) kernel estimator or of Härdle's (1990, p. 51) orthogonal series estimator. For the random design model, we propose

$$\tilde{r}(t) = n^{-1} \sum_{i=1}^{n} Y_i E_m(t, X_i)/\tilde{f}(t),$$

where $\tilde{f}$ is a wavelet estimator of the density of X given by

$$\tilde{f}(t) = n^{-1} \sum_{i=1}^{n} E_m(t, X_i).$$

A standard kernel density estimator could be used in place of $\tilde{f}$. The estimator $\tilde{r}$ is a wavelet version of the (evaluation) kernel estimator proposed by Nadaraya (1990) and Watson (1964). It can also be viewed as a wavelet version of an orthogonal series estimator studied by Härdle (1984). Antoniadis and Carmona (1990) introduced density estimators of the form $\tilde{f}$. In all these estimators the tuning parameter $m = m(n)$ needs to be chosen appropriately. A recent study of the relative merits of the convolution and evaluation kernel approaches to nonparametric regression has been made by Chu and Marron (1991).

* Anestis Antoniadis is Professor and Gérard Gregoire is Associate Professor, the University of Grenoble, Cedex 9, France. Both these authors also are members of Laboratory LMC, IMAG, BP 53, 38041 Grenoble Cedex 9, France. I. W. McKeague is Professor, Department of Statistics, Florida State University, Tallahassee, FL 32306. The work of I. W. McKeague was partially supported by U.S. Air Force Office of Scientific Research Grant AFOSR91-0048.
I. W. McKeague thanks the Department of Statistics, Université Joseph Fourier, for their warm hospitality during the writing of this article.

Journal of the American Statistical Association
December 1994, Vol. 89, No. 428, Theory and Methods

Like wavelet estimators, orthogonal series estimators use projections onto closed subspaces of $L^2(\mathbb{R})$ to represent successive approximations. But the projections used by orthogonal series estimators are finite dimensional, whereas the projections used by the wavelet estimators are infinite dimensional. Wavelet estimators cannot be seen as location-adaptive kernel estimators either (cf. Breiman, Meisel, and Purcell 1977). In fact, wavelet estimators are properly regarded as delta sequence estimators (see Walter and Blum 1979): $\hat{r}$ is a special type of the delta sequence estimator studied recently by Isogai (1990); $\tilde{r}$ is a special case of the estimator considered by Collomb (1981) and studied recently by Doukhan (1990). We shall obtain consistency of $\hat{r}$ and $\tilde{r}$ and $\hat{r}$ by applying a result of Isogai.

We are also able to establish rate of convergence results for $\hat{r}$ and asymptotic normality results for suitably modified versions of $\hat{r}$ and $\tilde{r}$. For $\hat{r}$, we do this by adapting some techniques that were originally developed for kernel estimators by Gasser and Müller (1979).

Eubank and Speckman (1991) have studied rates of convergence for a least squares orthogonal series estimator for r. They used trigonometric series and their method of proof is heavily dependent on the special properties of these systems. To avoid the need for periodic boundary conditions on the derivatives of r, they added appropriate polynomial terms to the orthogonal series. By using a least squares estimator constructed from an orthonormal wavelet basis of $L^2([0, 1])$, we show that the rates obtained by Eubank and Speckman hold without the need for more than just linear correction to deal with the boundary behavior of r.

Most delta sequence estimators in statistics have a wavelet version that can be studied using techniques similar to those developed in this article. We have focused our attention on the fixed-design wavelet estimator $\hat{r}$. The article is organized as follows. Some background on wavelet theory is reviewed in Section 2. Wavelet estimators are discussed for nonparametric regression in Section 3 and for hazard rates in Section 4. A discussion of applications to real data and a comparison of kernel and wavelet estimators is presented in Section 5. Proofs are collected in Section 6.

2. SOME BACKGROUND ON WAVELETS

This section is devoted to a brief introduction to the theory of wavelets that will be used in the sequel. We limit ourselves to the basic definitions and the main properties of wavelets. (For more information, including proofs of the theorems in full generality and more extensive discussion and examples, see Chui 1992, Daubechies 1990, Mallat 1989, and Meyer 1990.)

Computing with wavelets requires a description of two basic functions: the *scaling function* $\varphi(x)$ and the *primary wavelet* $\psi(x)$. The function $\varphi(x)$ is a solution of a two-scale difference equation

$$\varphi(x) = \sum_{k\in\mathbb{Z}} c_k \varphi(2x - k) \tag{1}$$

with normalization

$$\int_{\mathbb{R}} \varphi(x)\, dx = 1.$$

The function $\psi(x)$ is defined by

$$\psi(x) = \sum_{k\in\mathbb{Z}} (-1)^k c_{k+1} \varphi(2x + k). \tag{2}$$

The coefficients c_k are called the *filter coefficients,* and it is from careful choice of these that wavelet functions with desirable properties can be constructed. The condition

$$\sum_k c_k = 2$$

ensures the existence of a unique $L^1(\mathbb{R})$ solution to (1) (see Daubechies and Lagarias 1988a, thm. 2.1, p. 8). A wavelet system is the infinite collection of translated and scaled versions of φ and ψ defined by

$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k), \qquad j, k \in \mathbb{Z}$$

and

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \qquad j, k \in \mathbb{Z}.$$

An additional condition on the filter coefficients,

$$\sum_k c_k c_{k+2l} = 2 \quad \text{if } l = 0,$$
$$= 0 \quad \text{if } l \in \mathbb{Z},\ l \neq 0,$$

together with some other regularity conditions, implies that $\{\psi_{j,k}, j, k \in \mathbb{Z}\}$ is an orthonormal basis of $L^2(\mathbb{R})$, and $\{\varphi_{j,k}, k \in \mathbb{Z}\}$ is an orthonormal system in $L^2(\mathbb{R})$ for each $j \in \mathbb{Z}$ (see Daubechies 1990, lem. 3.4, p. 958). A key observation of Daubechies (1990, sec. 4) is that it is possible to construct finite-length sequences of filter coefficients satisfying all of these conditions, resulting in compactly supported φ and ψ.

The simplest example of a wavelet system is the Haar system, defined by setting $c_0 = c_1 = 1$ and all other $c_k = 0$. In this case both the scaling function and the primary wavelet are supported by the interval [0, 1], and the resulting system is an orthonormal basis of $L^2(\mathbb{R})$. But if instead of a general function in $L^2(\mathbb{R})$, one wants to analyze a function with much less or much more regularity, then the expansion given by the Haar system is inappropriate, the reason being that the coefficients either do not make any sense or their decay at infinity is bad. Replacing the scaling function in the Haar system by a more regular function produces a system with a much better behavior with respect to spaces of smooth functions. The regularity of the scaling function φ is defined in the following sense.

Definition 2.1. A scaling function φ is q-regular ($q \in \mathbb{N}$) if for any $l \leq q$, and for any integer k, one has

$$\left|\frac{d^l\varphi}{dx^l}\right| \leq C_k(1 + |x|)^{-k},$$

where C_k is a generic constant depending only on k.

We assume throughout that φ is q regular for some $q \in \mathbb{N}$. Of course the primary wavelet inherits the regularity of the scaling function. Moreover, if ψ is regular enough, then the resulting wavelet orthonormal basis provides unconditional bases for most of the usual function spaces, (see Meyer 1990). To obtain such a result, Mallat (1989) introduced the notion of a multiresolution analysis, the definition of which we recall here.

Definition 2.2. A multiresolution analysis of $L^2(\mathbb{R})$ consists of an increasing sequence of closed subspaces $V_j, j \in \mathbb{Z}$, of $L^2(\mathbb{R})$ such that

(a) $\cap V_j = \{0\}$;
(b) $\overline{\cup V_j} = L^2(\mathbb{R})$;
(c) there exists a scaling function $\varphi \in V_0$ such that $\{\varphi(\cdot - k), k \in \mathbb{Z}\}$ is an orthonormal basis of V_0;

and for all $h \in L^2(\mathbb{R})$,

(d) for all $k \in \mathbb{Z}$, $h(\cdot) \in V_0 \Leftrightarrow h(\cdot - k) \in V_0$, and
(e) $h(\cdot) \in V_j \Leftrightarrow h(2\cdot) \in V_{j+1}$.

The intuitive meaning of (e) is that in passing from V_j to V_{j+1}, the resolution of the approximation is doubled. Mallat (1989) has shown that given any multiresolution analysis, it is possible to derive a function ψ (the primary wavelet) such that the family $\{\psi_{j,k}, k \in \mathbb{Z}\}$ is an orthonormal basis of the orthogonal complement W_j of V_j in V_{j+1}, so that $\{\psi_{j,k}, j, k \in \mathbb{Z}\}$ is an orthonormal basis of $L^2(\mathbb{R})$. Conversely, the compactly supported wavelet systems mentioned earlier give rise to multiresolution analyses of $L^2(\mathbb{R})$ (see Daubechies 1990, thm. 3.6). When the scaling function is q regular, the corresponding multiresolution analysis is said to be q regular.

Let us now introduce the following projector and its associated integral kernel:

$$h \to E_j(h) = \int_{\mathbb{R}} E_j(\cdot, y) h(y)\, dy = \text{projection of } h \text{ onto } V_j.$$

It is easy to see that $E_j(x, y) = 2^j E_0(2^j x, 2^j y)$ and that $E_0(x + k, y + k) = E_0(x, y)$ for $k \in \mathbb{Z}$. Obviously, E_0 is not a convolution kernel, but the regularity of φ and ψ implies that it is bounded above by a convolution kernel; that is, $|E_0(x, y)| \leq K(x - y)$, where K is some positive, bounded, integrable function satisfying moment conditions (see Meyer 1990, p. 33). This remark will be exploited in the following sections. In particular, the bound $\sup_{x,y} |E_j(x, y)| = O(2^j)$ is often needed. We also mention some other useful properties. For any polynomial p of degree $\leq q$, one has

$$E_j(p) = p \tag{3}$$

(see Meyer 1990, p. 38). By (3) applied to $p(x) \equiv 1$ and part (c) of the definition of a multiresolution analysis, we see that

$$\sum_{k \in \mathbb{Z}} \varphi(x - k) = 1. \tag{4}$$

If a function h belongs to the Sobolev space $H^\nu = H^\nu(\mathbb{R})$, then the sequence $E_j(h)$ converges strongly to h in H^ν for $|\nu| \leq q$ and

$$\|h - E_j(h)\|_\nu = o(2^{-j\nu}) \tag{5}$$

for $0 < \nu \leq q$, by Mallat (1989, thm. 3), where $\|\cdot\|_\nu$ denotes the norm associated with H^ν. The Sobolev space $H^\nu(\mathbb{R}^d)$, $\nu \in \mathbb{R}$, $d \geq 1$, is defined to be the space of tempered distributions whose Fourier transforms are square integrable with respect to the measure $(1 + |x|^2)^\nu dx$ on $\mathbb{R}^d$ (see Hörmander 1989, p. 240).

Compactly supported wavelets are partitioned by the *wavelet number* N into families whose scaling functions have supports of equal size. N is defined as $(K_{\max} - K_{\min} + 1)/2$, where $K_{\min}$ is the greatest even integer and $K_{\min}$ is the least odd integer, such that $c_k \neq 0 \Rightarrow K_{\min} \leq k \leq K_{\max}$. Thus N is generically one-half the number of nonzero filter coefficients. The support of φ is the interval $[K_{\min}, K_{\max}]$, and the support of ψ is the interval $[1 - N, N]$; note that both support widths are $2N - 1$ unit intervals long. The examples constructed by Daubechies have the property that their support widths increase linearly with their regularity. This is illustrated by Figure 1. Daubechies shows that there exists $\nu > 0$ such that ${}_N\varphi, {}_N\psi \in C^{\nu N}$, where $\varphi \in C^{n+\gamma}$ if $\varphi \in C^n$ and $\varphi^{(n)}$ is Hölder continuous with exponent γ $(0 \geq \gamma < 1)$. More precisely, Daubechies and Lagarias (1988, p. 62) obtained

$${}_2\varphi \in C^{.5500} \qquad {}_3\varphi \in C^{1.0878\ldots} \qquad {}_4\varphi \in C^{1.6179\ldots}.$$

An algorithm described by Daubechies and Lagarias (1988, p. 17), the cascade algorithm, allows us to construct the orthogonal compactly supported wavelets as limits of step functions that are finer and finer scale approximations of ${}_N\varphi$. The algorithm is easy to implement on a computer and converges very rapidly. Given a finite sequence of filter coefficients, $c_0, \ldots, c_N$, define the linear operator A by

$$(Aa)_n = \sum_{k \in \mathbb{Z}} c_{n-2k} a_k, \qquad a = (a_k)_{k \in \mathbb{Z}},$$

where it is understood that $c_k \equiv 0$ if $k < 0$ or $k > N$. Define $a^j = A^j a^0$, where $(a^0)_0 = 1$ and $(a^0)_k = 0$ for $k \neq 0$. Set

$$\varphi_j(x) = \sum_{k \in \mathbb{Z}} a_k^j \chi(2^j x - k), \tag{6}$$

where χ is the indicator function of the interval $[-\frac{1}{2}, \frac{1}{2}[$. Under certain conditions (see Daubechies 1990, p. 951), the sequence of functions φ_j converges pointwise to a limit function φ that satisfies the two-scale difference equation (1).

3. NONPARAMETRIC REGRESSION

In this section we establish consistency of $\hat{r}$ using a theorem of Isogai (1990). Also, under conditions on the regression function r that are weaker than the usual smoothness assumptions, we give asymptotic bounds for the bias and variance of $\hat{r}$ and establish asymptotic normality for a modified version of $\hat{r}$. This modified version of $\hat{r}$ is an approximation that agrees with $\hat{r}$ at dyadic points of the form $k2^{-m}$; it is needed to stabilize the variance. At the nondyadic points, the variance of $\hat{r}$ itself is unstable because of irregularity in the wavelet kernels. In practice, the "optimal" bandwidth can be selected by cross-validation (see Sec. 5.2 for further discussion). This usually amounts to a choice between at most three or four values of m. This small range of possibly optimal resolution is very desirable, because the computational demands for $\hat{r}$ can be large.

The cascade algorithm described in Section 2 gives a simple method to calculate the estimator $\hat{r}$ and $\tilde{r}$. Note that the delta sequence E_m used in $\hat{r}$ and $\tilde{r}$ can be written as

$$E_m(t, s) = 2^m \sum_{k \in \mathbb{Z}} \varphi(2^m t - k)\varphi(2^m s - k).$$

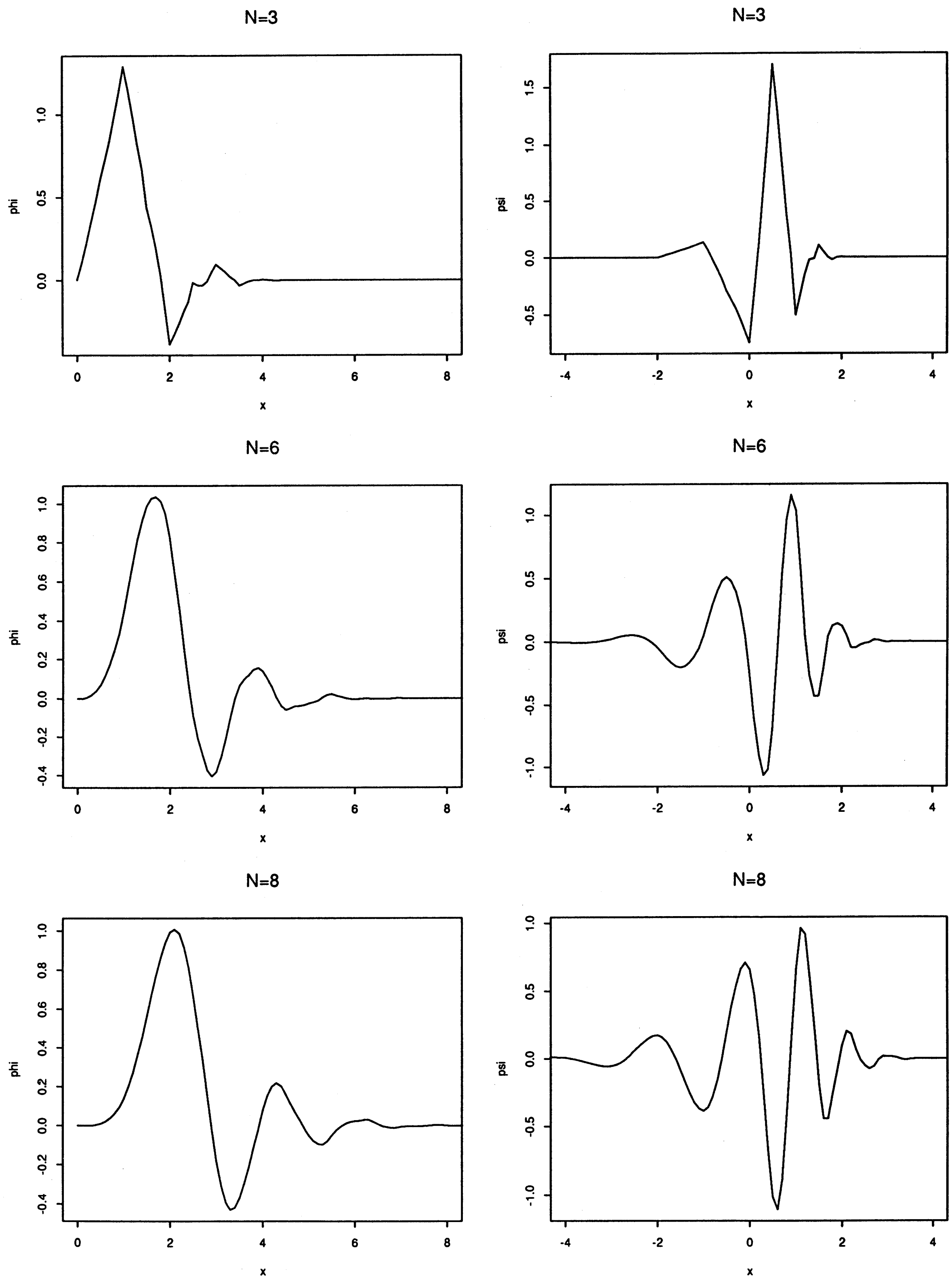

Figure 1. The Scaling Function $_N\varphi$ (Left Column) and the Corresponding Wavelet $_N\psi$ (Right Column) for N = 3, 6, and 8. Note that the support widths increase with the regularity.

When φ has compact support, this is a finite sum, each term of which can be evaluated by the cascade algorithm. To evaluate the weights $\int_{A_i} E_m(t, s)\, ds$ used in $\hat{r}(t)$, we use an integrated version of (6):

$$\int_u^v \varphi_j(x)\, dx = \sum_{k\in\mathbb{Z}} a_k^j \int_u^v \chi(2^j x - k)\, dx.$$

The sequence $\int_u^v \varphi_j(x)\, dx$ converges to $\int_u^v \varphi(x)\, dx$ for each $u < v$.

Some plots of $E_m(t, s)$ for the scaling function ${}_6\varphi$ are given in Figure 2. Note that the wavelet kernels are dyadic translation invariant: $E_m(t + u, \cdot) = E_m(t, \cdot - u)$ for all dyadic rationals u of the form $k/2^m$ but *not* for general real numbers u. Also note the substantial variation in the form of the wavelet kernel as one passes between the dyadic points. This is more than just a variation in the local bandwidth—compare the curves corresponding to $t = .2$ and $t = .5$ in Figure 2b. It appears that this feature of the wavelet kernel allows wavelet estimators to adapt automatically to local features of the regression function. An unfortunate side effect is that the asymptotic variance of wavelet estimators is unstable.

Another reasonable estimator of r is

$$\hat{r}_c(t) = \sum_{i=1}^n Y_i \int_{A_i} E_m(0, s - t)\, ds,$$

a convolution kernel estimator based on the kernel $K(t) = E_0(0, -t)$ and having bandwidth 2^{-m}. A similar change can be made to $\tilde{r}$. Note that $\hat{r}$ and $\hat{r}_c$ agree at dyadic rationals of the form $k/2^m$. Asymptotic results for this estimator are special cases of those given by Gasser and Müller (1979), although by using this special kernel K we can relax the smoothness conditions on r. But a finite sample comparison between $\hat{r}$ and $\hat{r}_c$ that examines their integrated mean squared errors for various values of m shows that $\hat{r}$ is superior (see Sec. 5.1). This is explained by the global approximation property (5) of the projection operator E_m used in $\hat{r}$. Such a property is not available for $\hat{r}_c$. A general bandwidth might improve the performance of $\hat{r}_c$, which only uses bandwidth of the form 2^{-m}. But the heavy computational demands for such an estimator make any bandwidth cross-validation selection procedure impractical.

Our first result gives consistency of $\hat{r}$.

Theorem 3.1. If r is continuous at t, $m \to \infty$ and $\max_i |t_i - t_{i-1}| = o(2^{-m})$, then $\hat{r}(t)$ is mean square consistent.

Strong consistency of $\hat{r}(t)$ can be obtained under a more refined condition on the rate of increase of m, using Isogai's theorem 3.2. To obtain deeper results, we need the regression function r and the density f (in the random design case) to satisfy the following conditions:

1. $r, f, rf \in H^\nu$, for some $\nu > \frac{1}{2}$.
2. r and f are Lipschitz of order $\gamma > 0$.
3. f does not vanish on $]0, 1[$.

Functions belonging to H^ν for $\nu > \frac{3}{2}$ are continuously differentiable (see Treves 1967, p. 331), so condition 2 is redundant when $\nu > \frac{3}{2}$. We also need some additional assumptions on the scaling function φ:

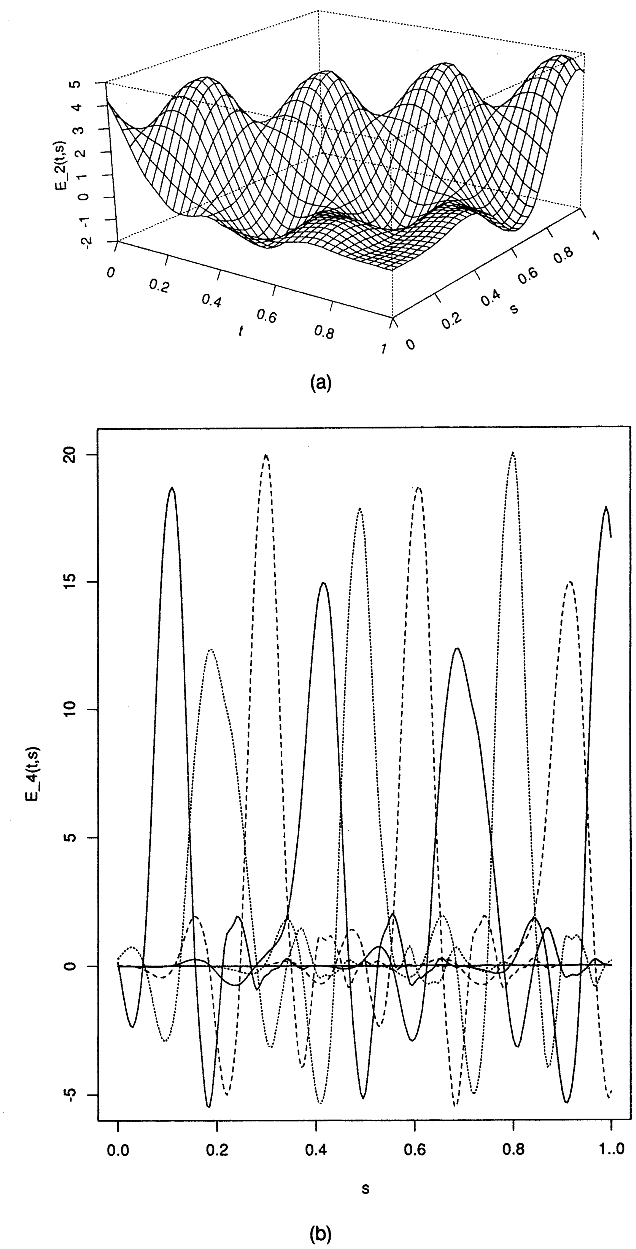

Figure 2. The Wavelet Kernel $E_m(t, s)$ for the Scaling Function ${}_6\varphi$. (a) Perspective plot of $E_2(t, s)$; (b) $E_4(t, \cdot)$ for ten different values (.1, .2, . . . , 1.0) of t. Note the translation invariance $E_m(t + u, \cdot) = E_m(t, \cdot - u)$ for dyadic rationals u of the form $k/2^m$.

4. φ has compact support.
5. φ is Lipschitz.
6. $|\hat{\varphi}(\xi) - 1| = O(|\xi|)$ as $\xi \to 0$.

Here $\hat{\varphi}$ denotes the Fourier transform of φ. The compactly supported scaling functions ${}_N\varphi$, $N \geq 3$, satisfy all of these conditions; in particular, condition 6 holds by Daubechies (1990, p. 963). For our asymptotic normality results, we will need φ to be regular of order $q \geq 1$. But to obtain good rates of convergence for the mean square error of $\hat{r}$, we need to adapt the regularity of φ to the smoothness of r:

7. φ is regular of order $q \geq \nu$.

A disadvantage of more regular wavelets is that their support is larger and thus their boundary effects are more pronounced. But wavelet estimators based on more regular compactly supported wavelets are unbiased away from the boundary for higher-order polynomials; see (3).

Following Gasser and Müller (1979) for the fixed design case, to study the mean squared error of $\hat{r}$, we assume that

$$\max_i |t_i - t_{i-1}| = O(n^{-1}). \tag{7}$$

We shall also assume that for some Lipschitz function $\kappa(\cdot)$,

$$\rho(n) \equiv \max_i \left| s_i - s_{i-1} - \frac{\kappa(s_i)}{n} \right| = o(n^{-1}), \tag{8}$$

where $A_i = [s_{i-1}, s_i)$. This is a standard assumption for the fixed design model but is somewhat weaker than the "asymptotic equidistance" assumption of Gasser and Müller (1979) in which $\kappa(t) \equiv 1$ and $\rho(n) = O(n^{-\delta})$ for some $\delta > 1$.

The next result gives an asymptotic bound for the bias of $\hat{r}$.

Theorem 3.2.

$$\mathbb{E}\hat{r}(t) - r(t) = O(n^{-\gamma}) + O(\eta_m),$$

where

$$\begin{aligned}\eta_m &= (1/2^m)^{\nu-1/2} \quad \text{if } \tfrac{1}{2} < \nu < \tfrac{3}{2},\\ &= \sqrt{m}/2^m \quad \text{if } \nu = \tfrac{3}{2},\\ &= 1/2^m \quad \text{if } \nu > \tfrac{3}{2}.\end{aligned}$$

To obtain an asymptotic expansion of the variance and an asymptotic normality result, we need to consider an approximation to $\hat{r}$ based on its values at dyadic points of order m. That is, define

$$\hat{r}_d(t) = \hat{r}(t^{(m)}),$$

where $t^{(m)} = [2^m t]/2^m$. Thus $\hat{r}_d$ is the piecewise-constant approximation to $\hat{r}$ at resolution 2^{-m}. The piecewise-constant feature of $\hat{r}_d$ makes it an unattractive alternative to the unmodified estimator $\hat{r}$ (at least for small m). In particular, the bias is increased by an additional term of order $O(2^{-m\gamma})$. But if one tries to obtain a precise asymptotic expansion of the variance of $\hat{r}(t)$, then a difficulty arises in that the variance is unstable as a function of t. This problem is avoided with $\hat{r}_d$.

Theorem 3.3.

$$\operatorname{var}(\hat{r}_d(t)) = \frac{\sigma^2 2^m}{n} \kappa(t)(w_0^2 + o(1)) + O(2^m \rho(n)) + O\left(\frac{2^{2m}}{n^2}\right),$$

where $w_0^2 = \sum_{k\in\mathbb{Z}} \varphi^2(k)$. The variance of $\hat{r}(t)$ has this form except that for general (nondyadic) t, the leading term is $O(2^m/n)$.

From the proof of this theorem, it can be seen that the leading term of the variance of $\hat{r}(t)$ is $\sigma^2 2^m n^{-1} \kappa(t) w^2(t_m)$, where $t_m = 2^m t - [2^m t]$ and w^2 is the function defined by

$$w^2(u) = \int_{\mathbb{R}} E_0^2(u, v)\, dv.$$

Notice that for dyadic t and m sufficiently large, $t_m = 0$, so the variance of $\hat{r}(t)$ is asymptotically stable. But if t is nondyadic, then the sequence t_m wanders around the unit interval and fails to converge. For example, at $t = \frac{1}{3}$, it oscillates between $\frac{1}{3}$ (m even) and $\frac{2}{3}$ (m odd), so the variance oscillates between $w^2(\frac{1}{3})$ and $w^2(\frac{2}{3})$. The sequence t_m belongs to the class of exponential lacunary sequences studied in ergodic theory. It is known that except for at most countably many t, the sequence t_m has infinitely many accumulation points (see Rauzy 1976, p. 67, cor. 2.2). It is also interesting to note that for irrational t's, the sequence is eventually confined to the interval $[\frac{1}{3}, \frac{2}{3}]$ (see Rauzy 1976, p. 69).

Plots of w^2 for the Daubechies scaling functions $\varphi = {}_N\varphi$, $N = 3, 5, 8$, are displayed in Figure 3. It can be seen that the variance of $\hat{r}(t)$ at nondyadic t can vary approximately by a factor of 3 for $N = 3$ and by a factor of $\frac{5}{3}$ for $N = 5$ and 8. The variance of $\hat{r}$ is inflated over the variance of $\hat{r}_d$ by a factor of at most 1.75 for $N = 5$ and 1.19 for $N = 8$. Taking the larger bias of $\hat{r}_d$ into account, it appears that the unmodified estimator $\hat{r}$ is at least as efficient as $\hat{r}_d$, and it is $\hat{r}$ that we recommend in practice. Generally, higher regularity of the wavelet basis reduces instability in the asymptotic variance of $\hat{r}(t)$, although this comes at the expense of larger bias (the support of the scaling function increases with the regularity).

For $N = 3, 5, 8$, the constants $w_0^2 = w^2(0)$ are 1.81, .72, and 1.05. This suggests that ${}_5\varphi$ is more suitable than ${}_3\varphi$ or ${}_8\varphi$ when used in connection with $\hat{r}_d$. When used in connection with $\hat{r}$ there is little difference between ${}_3\varphi$ and ${}_8\varphi$.

3.1 Optimal Rates

To give a rate of convergence for the mean squared error of their estimates, Gasser and Müller (1979) assumed that r is k-times continuously differentiable and use a kernel of order $k \geq 2$. They found that the best rate of convergence for the mean squared error is $n^{-2k/(2k+1)}$. An analogous result holds in our case. Assume that r is $k = q + 1$ times continuously differentiable, where q is the regularity of the scaling

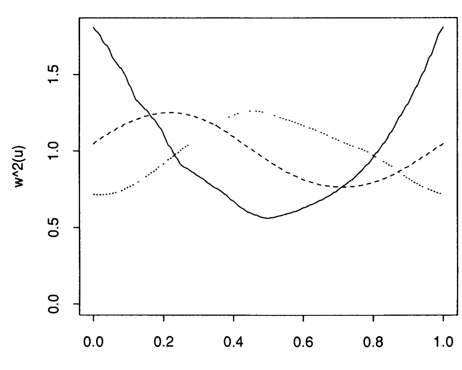

Figure 3. The Function w^2 for ${}_3\varphi$ (Solid Line), ${}_5\varphi$ (Dotted Line), and ${}_8\varphi$ (Dashed Line).

function. Because polynomials of degree $\leq q$ are invariant under $E_m(t, s)$ [see (3)], we get, using a Taylor expansion of r, that the best rate of convergence for the mean squared error of $\hat{r}$ at dyadic points is the same as for the kernel estimator and is attained by $m = \log_2 n/(2k + 1)$. It is worth stressing that the wavelet approach allows us to obtain rates under much weaker assumptions on r than second-order differentiability. For example, the triangular function having Fourier transform $\sin^2(\xi/2)/(\xi/2)^2$ belongs to H^1 and is Lipschitz of order 1, so it satisfies our conditions 1 and 2 but is not differentiable. The mean squared error of $\hat{r}_d$ is of order $O(2^m/n) + O(2^{-m(2\nu-1)}) + O(2^{-2m\gamma})$. The best rate is $n^{-2\nu^*/(2\nu^*+1)}$, which is attained by $m = \log_2 n/(2\nu^* + 1)$, where $\nu^* = \min(\frac{3}{2}, \nu, \gamma + \frac{1}{2}) - \varepsilon$ and $\varepsilon = 0$ for $\nu \neq \frac{3}{2}$, $\varepsilon > 0$ for $\nu = \frac{3}{2}$.

Our next result concerns asymptotic normality of $\hat{r}_d$. It can be applied to the unmodified estimator $\hat{r}$ at dyadic points.

Theorem 3.4. If $n2^{-m} \to \infty$ and $n2^{-2m\nu^*} \to 0$, then $\sqrt{n2^{-m}}(\hat{r}_d(t) - r(t))$ is asymptotically normal with mean zero and variance $\sigma^2 w_0^2 \kappa(t)$.

We now turn to the estimator $\tilde{r}$ used in the random design model. Much of the above discussion carries over to this case. The following result gives consistency of $\tilde{r}$.

Theorem 3.5. If $m \to \infty$ and $n2^{-m} \to \infty$, then $\tilde{f}(t)$ is consistent and, if in addition $\mathbb{E}(Y^2 | X = x)$ is bounded for x belonging to a neighborhood of t, then $\tilde{r}(t)$ is consistent.

A result of Doukhan (1990, thm. 1) dealing with general delta sequence estimates can be used to establish uniform strong consistency of $\tilde{r}$, but under more stringent conditions on the rate of increase of m. Conditions 1–6 of Doukhan's paper are easily checked along the lines that we check Isogai's conditions in the proof of Theorem 3.1 and using (11). As for the fixed design model, to obtain an asymptotic distribution result (at all t), we need to consider the piecewise constant approximation $\tilde{r}_d(t) = \tilde{r}(t^{(m)})$ instead of $\tilde{r}$.

Theorem 3.6. Suppose that for some $\varepsilon > 0$, we have $\mathbb{E}(|Y|^{2+\varepsilon} | X = x)$ bounded for x belonging to a neighborhood of t, $n2^{-m} \to \infty$ and $n2^{-2m\nu^*} \to 0$. Then $\sqrt{n2^{-m}}(\tilde{r}_d(t) - r(t))$ is asymptotically normal with mean zero and variance $\operatorname{var}(Y | X = t) w_0^2 / f(t)$.

3.2 Symmetrized Wavelet Estimators

Inspecting Figure 3, it can be seen that there is a lack of symmetry in the wavelet kernels $E_m(t, s)$ about the point t, as inherited from the asymmetry in the scaling functions; see Figure 1. This is somewhat unnatural from a statistical point of view, because a time-reversal in the data produces a different estimate from the time-reversed $\hat{r}$ (denoted $\hat{r}_{\text{rev}}$). Unfortunately, except for the Haar basis, there exists *no* compactly supported wavelet basis in which the scaling function is symmetric around any axis (see Daubechies 1990). Another difficulty is caused by the excessive weight placed at points far to the left of t, resulting in a pronounced edge effect at the lower limit of the design interval (see the discussion concerning the voltage data example in Sec. 5.3). A simple way of correcting these flaws in $\hat{r}$ is to use a weighted average of $\hat{r}$ and $\hat{r}_{\text{rev}}$ with weights depending on the evaluation point:

$$\hat{r}_{\text{sym}}(t) = t\hat{r}(t) + (1 - t)\hat{r}_{\text{rev}}(t).$$

It is easily seen that this estimator inherits the properties of $\hat{r}$ proved earlier. A similar modification can be made to any of the wavelet estimators considered in this article.

3.3 Confidence Intervals

To use our asymptotic normality result to obtain confidence intervals for $r(t)$ at a given t, one needs to consistently estimate the noise variance. In the fixed design case, the noise variance is σ^2. We suggest using the following estimate of Müller (1985):

$$\hat{\sigma}^2 = \frac{2}{3(n-2)} \sum_{i=2}^{n-1} \left[Y_i - \frac{1}{2}(Y_{i-1} + Y_{i+1}) \right]^2,$$

obtained by fitting constants to successive triples of the data. Müller's lemma 1 shows that if the regression function is Hölder continuous of order 1, then $\hat{\sigma}^2$ is almost surely consistent and

$$|\hat{\sigma}^2 - \sigma^2| = O\left(\frac{(\log n)^{1/2+\varepsilon}}{n^{1/2}} \right)$$

a.s. as $n \to \infty$ for any $\varepsilon > 0$. In practice, to obtain a good impression of the errors involved in the point estimates $\hat{r}(t)$ of $r(t)$, it would be enough to provide confidence intervals at the 2^m dyadic points of the design region. For $m = 4$, this would give 16 confidence intervals.

3.4 Least Squares Wavelet Regression

Orthogonal series used for least squares regression should form a basis of the L^2 space on the design region; that is, $L^2([0, 1])$. The wavelets described up to now form an orthonormal basis of $L^2(\mathbb{R})$ and are not appropriate. Instead, we shall use a wavelet orthonormal basis $\{\psi_{j,k}, j \geq 1, k \in S_j\}$ of $L^2([0, 1])$ constructed by Jaffard and Meyer (1989). Here S_j is a subset of $\mathbb{Z}$, defined as R_j by Jaffard and Meyer (1989, p. 95). For some integer j_0 depending on q, the set S_j is empty for $j \leq j_0$. These wavelets belong to the space C^{2q-2}, where $q \geq 2$ and the subscript 0 indicates support within $]0, 1[$. They are defined through a multiresolution analysis of $L^2([0, 1])$ and form unconditional bases of H_0^ν, $0 < \nu < 2q - 2$. Assume that $r(0) = r(1) = 0$ and $r \in H_0^\nu$. This is a weaker assumption than condition (ii) of theorem 1 of Eubank and Speckman (1991), but the boundary condition $r(0) = r(1) = 0$ is still rather restrictive. It can be removed by adding a linear function to the regression analysis (cf. Eubank and Speckman (1991).

We shall obtain a rate of convergence for the mean squared error,

$$R(\hat{r}_{\text{ls}}) = n^{-1} \sum_{i=1}^{n} \mathbb{E}(r(X_i) - \hat{r}_{\text{ls}}(X_i))^2,$$

of the least squares wavelet estimator $\hat{r}_{\text{ls}}$ given by

$$\hat{r}_{\rm ls}(t) = \sum_{j=1}^{m} \sum_{k \in S_j} d_{j,k} \psi_{j,k}(t),$$

where the $d_{j,k}$'s are obtained by least squares. The number $D_m = \sum_{j=j_0}^{m} |S_j|$ of functions $\psi_{j,k}$ used in the regression is bounded above by $\frac{3}{4} 2^m$. We assume that the observation errors have constant variance σ^2. Let G_n denote the empirical distribution function of the design points X_i and assume that $\delta_n = \sup_t |G_n(t) - G(t)| \to 0$, where G is some distribution function that is absolutely continuous with density bounded away from zero and infinity. Typically δ_n is of order $O(n^{-1})$ in the fixed design case and of order $O(n^{-1/2} \log \log n)$ in the random design case (see Eubank and Speckman 1991 for further discussion).

Theorem 3.7. If $r \in H_0^\nu$, where $\nu \geq 1$ is an integer, then

$$R(\hat{r}_{\rm ls}) \leq O(2^{-2m\nu}) + \sigma^2 D_m / n + O(\delta_n 2^{-m(\nu-1)}).$$

This rate of convergence essentially agrees with the rate given in theorem 1 (iii) of Eubank and Speckman (1991).

4. HAZARD RATE ESTIMATION

In this section we study a wavelet version of Ramlau-Hansen's (1983) estimator of a hazard rate function. It turns out that most of the wavelet techniques we have used for nonparametric regression carry over to this setting. Since the work of Aalen (1978), it is well known that hazard rate estimation can be viewed in the context of inference for a counting process multiplicative intensity model given by $\lambda(t) = \alpha(t)Y(t)$, where $Y(t)$ is a nonnegative observed process. In the usual survival analysis or reliability application, a portion $\tilde{T} = \min(T, C)$ of an individual's lifetime T is observed, where C is a censoring time (assumed to be independent of T). Data are available on n individuals with corresponding (T_i, C_i) being independent and distributed as (T, C). Suppose that T has hazard rate function α and that the distribution function H of $\tilde{T}$ is such that $H(1) < 1$. Then the counting process $N_n(t) = \sum_{i=1}^n I\{T_i \leq t, C_i \geq T_i\}$ has intensity $\alpha(t)Y_n(t)$, where $Y_n(t) = \sum_{i=1}^n I\{T_i \geq t, C_i \geq t\}$ is the number of individuals at risk at time t^-. This is a special case of Aalen's multiplicative intensity model. In what follows, the notation is essentially the same as Ramlau-Hansen's.

Our wavelet estimator for the hazard function α is defined by

$$\hat{\alpha}(t) = \int_0^1 E_m(t, s)\, d\hat{\beta}(s), \tag{9}$$

where $\hat{\beta}$ is the Nelson–Aalen estimator,

$$\hat{\beta}(t) = \int_0^t \frac{J(s)}{Y(s)}\, dN(s),$$

$J(s) = I\{Y(s) > 0\}$, and $J(s)/Y(s)$ is defined to be 0 when $Y(s) = 0$. To obtain asymptotic results, we index the processes N, J, and Y by n. We use the same assumptions on α as were used for r in the regression case. We also assume that there exists a positive function τ bounded away from zero and infinity such that $\mathbb{E}[\sup_{0<s<1} |nJ_n(s)/Y_n(s) - 1/\tau(s)|] \to 0$ as $n \to \infty$. This condition is easily checked in the survival analysis case described earlier.

Define $\delta_n = \sup_{0<s<1} \mathbb{E}(1 - J_n(s))$. Our first result implies that the wavelet estimate is asymptotically unbiased.

Theorem 4.1.

$$\mathbb{E}\hat{\alpha}(t) - \alpha(t) = O(\eta_m) + 2^{m/2} O(\delta_n^{1/2}),$$

where η_m is defined in the statement of Theorem 3.2.

As in the regression case, it is convenient to approximate $\hat{\alpha}$ by an estimator based on the values of $\hat{\alpha}$ at dyadic points; $\hat{\alpha}_d(t) = \hat{\alpha}(t^{(m)})$, where $t^{(m)} = [2^m t]/2^m$. Observe that Theorem 4.1 holds for $\hat{\alpha}_d(t)$, provided that we add $O(2^{-m\gamma})$ to the asymptotic expansion of the bias.

Theorem 4.2.

$$\mathbb{E}(\hat{\alpha}_d(t) - \alpha(t))^2 = \frac{2^m}{n} \frac{\alpha(t)}{\tau(t)} w_0^2 + 2^m o(n^{-1}) + O(\eta_m^2) + 2^m O(\delta_n) + O(2^{-2m\gamma}).$$

The mean squared error of $\hat{\alpha}(t)$ has the same form, except that for general (nondyadic) t, the leading term is $O(2^m/n)$.

Under $n2^{-m} \to \infty$, $n2^{-2m\nu^*} \to 0$, and $\delta_n = o(n^{-1})$, we have L^2 consistency of $\hat{\alpha}_d(t)$. The leading term in the asymptotic expansion of the mean squared error is then of order $O(2^m/n)$. If $\hat{\alpha}$ is used instead of $\hat{\alpha}_d$, then the Lipschitz condition on α is unnecessary.

Theorem 4.3. If $n2^{-m} \to \infty$, $n2^{-2m\nu^*} \to 0$ and $\delta_n = o(n^{-1})$, then $\sqrt{n2^{-m}}(\hat{\alpha}_d(t) - \alpha(t))$ is asymptotically normal with zero mean and variance $\alpha(t)w_0^2/\tau(t)$.

5. PRACTICAL APPLICATION AND DISCUSSION

5.1 Finite Sample Comparisons

So far we have only considered the asymptotic behavior of our estimators. However, as long as one deals with linear estimates and is interested in the mean squared error or the integrated mean squared error of these estimates for finite samples, numerical calculations are possible that approximate these quantities to any desired degree of accuracy when the true regression function, the error variance and the weights are known (other properties of the error probability law being irrelevant). The method that we are going to describe has been used by Gasser and Müller (1984) for a finite sample comparison between cubic smoothing splines and various types of kernel estimates.

The method applies to linear estimates of the form $\hat{r}(t) = \sum_{i=1}^n w_i(t) Y_i$. For such estimates the bias at t is $\sum_{i=1}^n w_i(t)(r(t_i) - r(t))$ and the variance is $\sigma^2 \sum_{i=1}^n w_i^2(t)$. The integrated mean squared error is obtained by numerically integrating variance + bias^2 over a fine grid of t's.

We used the same underlying function as did Gasser and Müller; that is,

$$r(t) = 2 - 2t + 3\exp(-(t - .5)^2/.01), \qquad t \in [0, 1],$$

and compared our wavelet estimator with a kernel estimate. The residual variance was taken as $\sigma^2 = .2$ and the sample size $n = 25$. The results are presented in Table 1. The inte-

Table 1. The Performance of the Wavelet Estimator $\hat{f}$ for Various Scaling Functions φ when $\sigma^2 = .2$

φ	Optimal MSE	Optimal m	Integrated bias²	Integrated variance
$_3\varphi$	2.64×10^{-2}	3	1.07×10^{-2}	1.56×10^{-2}
$_5\varphi$	3.09×10^{-2}	4	3.54×10^{-4}	3.05×10^{-2}
$_6\varphi$	3.08×10^{-2}	4	2.04×10^{-4}	3.06×10^{-2}
$_8\varphi$	3.07×10^{-2}	3	1.51×10^{-2}	1.56×10^{-2}

NOTE: For comparison, the integrated mean square error of the convolution kernel (Epanechnikov) estimate with optimal bandwidth .065, is 2.35×10^{-2}.

grated mean squared error was evaluated using a grid of 200 equidistant points between .25 and .75. We restricted attention to an interval smaller than [0, 1] to avoid possible boundary effects. The wavelet kernels corresponding to four different scaling functions ($_N\varphi$ for $N = 3, 5, 6$, and 8) were used. They were compared with an Epanechnikov kernel having optimal bandwidth. Although the results are not reported here, we also examined the performance of the wavelet convolution estimator $\hat{r}_c$. The integrated mean square error was significantly larger, due mainly to a larger variance.

The convolution kernel estimate does slightly better than the wavelet estimate, but this is not unexpected, because the optimal bandwidth is chosen from a continuum of possible values, whereas the tuning parameter m is discrete.

5.2 Cross-validation

Any nonparametric regression method is highly dependent on the tuning parameter, so it is desirable to select such parameters automatically. The problem of selecting m is rather easier than the bandwidth selection problem for kernel estimators (see, e.g., Härdle and Marron 1985 in the regression case and Grégoire 1991 in the survival analysis case), because the bandwidth is essentially reduced to being of the form 2^{-m}, where $m < \frac{1}{2}\log_2 n$. A commonly used selection rule adapted to our setting is to choose m as the minimizer of the cross-validation function

$$CV(m) = n^{-1} \sum_{i=1}^{n} (Y_i - \hat{r}_{(i)}(t_i))^2,$$

where $\hat{r}_{(i)}(t)$ is the leave-one-out estimator obtained by evaluating $\hat{r}$ (as a function of m and t) with the ith data point removed. This gives reasonable results when applied to real and simulated data. In practice, for sample sizes between 100 and 200, we have found that it suffices to examine only $m = 3$, 4, and 5.

5.3 Examples

To illustrate the techniques given so far, and to add to the earlier discussion, we now consider two real examples. The first example concerns the motorcycle impact data given by Härdle (1990) and presented in Figure 4. The observations consist of accelerometer readings taken through time in an experiment on the efficacy of crash helmets. This particular data set was also analyzed by Silverman (1980) by spline smoothing techniques. For several reasons, the time points are not regularly spaced. It is of interest both to discern the general shape of the underlying acceleration curve and to draw inferences about its minimum and maximum values. Obviously, the observations are correlated and their variance is not constant, but for illustrative purposes we shall assume that the fixed design model holds.

We plotted the estimate $\hat{r}$ for various values of m, using the wavelet kernel based on $_8\varphi$. This choice of scaling function is reasonable according to the discussion following the statement of 3.3. We tried $_3\varphi$, which the finite sample comparisons suggested as being even better than $_8\varphi$, but obtained a very poor fit. This poor performance of $_3\varphi$ is probably due to the greater instability of the variance; see Figure 2. Cross-validation selected the curve $m = 4$ as giving the best fit: the function $CV(m)$ was found to be 534 at $m = 3$, 432 at $m = 4$, and 497 at $m = 5$. Inspecting Figure 4, one notices that $\hat{r}$ is considerably biased for $m = 3$; for $m = 5$, it detects the sharp drop in acceleration around 15 milliseconds but has undesirable oscillations. The $m = 4$ estimate is clearly the best—it captures the general features of the underlying curve, except for a positive bias around 12 milliseconds.

Another example is presented in Figure 5. The data set, discussed in example 3.4.5 of Eubank (1988, p. 82), represents the voltage drop in the battery of a guided missile motor during flight. In this example the assumptions of the fixed design model are much more reasonable. We find that there is an undesirable boundary effect in $\hat{r}$ at time 0. The time reversed $\hat{r}$ has a similar problem at the right end of the design interval. But the symmetrized estimator $\hat{r}_{\text{sym}}$ discussed in Section 3 produces an acceptable fit. In fact, considering that $\hat{r}_{\text{sym}}$ uses a tuning parameter setting chosen from among only three different values ($m = 3$, 4, and 5), it gives an outstanding result compared with other nonparametric regression estimates.

6. PROOFS

6.1 Proof of Theorem 3.1

We apply Theorem 3.1 of Isogai (1990) with 2^m in the role of m and E_m in place of δ_m. We need to check that following conditions hold for each $x \in [0, 1]$:

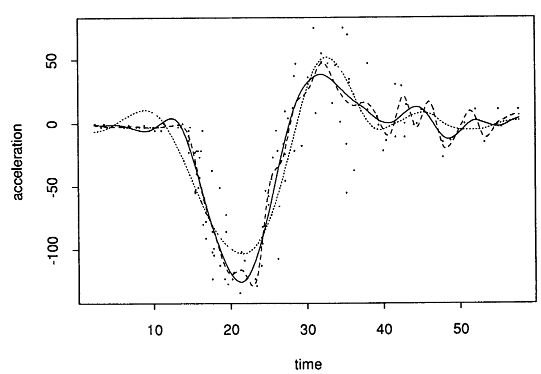

Figure 4. Plot of the Motorcycle Impact Data Together with the Wavelet Regression Estimates $\hat{r}$ Based on the Scaling Function $_8\varphi$ for $m = 3$ (Dotted Line), $m = 4$ (Solid Line), and $m = 5$ (Dashed Line). Cross-validation selected the curve $m = 4$ as giving the best fit.

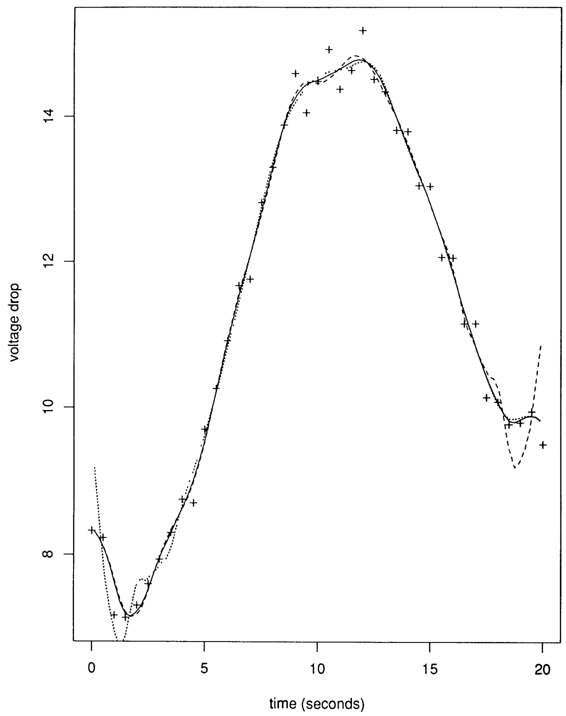

Figure 5. Plot of the Voltage Drop Data Together With the Symmetrized Wavelet Regression Estimate $\hat{f}_{sym}$ (Solid Line), $\hat{f}$ (Dotted Line) and $\hat{f}_{rev}$ (Dashed Line) for m = 4 and Scaling Function ${}_6\varphi$. Note that symmetrization has improved the estimate at the boundaries.

(a) $\sup_{m\geq 1} \int_0^1 |E_m(x, y)|\, dy < \infty$;
(b) $\int_0^1 E_m(x, y)\, dy \to 1$;
(c) $\int_0^1 |E_m(x, y)| I(|x - y| > \varepsilon)\, dy \to 0$ for all $\varepsilon > 0$; and
(d) $\sup_{y\in[0,1]} |E_m(x, y)| = O(2^m)$.

Using the assumption that φ is 0 regular, we have

$$\int_0^1 |E_m(x,y)|\, dy \leq C_2 2^m \int_0^1 (1 + 2^m|x-y|)^{-2}\, dy, \quad (10)$$

so (a) holds. Condition (b) follows by setting $f \equiv 1$ in equation (33) of Mallat (1989). Using the indicator to control the integrand in (10), we see that the expression in (c) is of order $O(2^{-m}) \to 0$. Condition (d) is immediate from the properties of E_m discussed in Section 2.

6.2 Proof of Theorem 3.2

Arguing along the lines of Gasser and Müller (1979, app. 1), and using the Lipschitz condition (2) on r, it can be seen that

$$\mathbb{E}\hat{r}(t) = \int_0^1 E_m(t, s) r(s)\, ds + O(n^{-\gamma}).$$

To complete the proof, it suffices to show that

$$\int_0^1 E_m(t, s) r(s)\, ds = r(t) + O(\eta_m). \quad (11)$$

This is demonstrated by applying an extension of a result of Schomburg (1990) to the function $g(x, y) = E_0(x, y)$; see Theorem A.1 in the Appendix. In Lemma A.2 we check that this function satisfies the conditions of Theorem A.1. First, note that

$$\int_0^1 E_m(t, s) r(s)\, ds = (E_m r)(t)$$

for m sufficiently large, because t is in the interior of [0, 1] and φ has compact support. Next, denoting the delta distribution centered at t by δ_t and the duality between H^ν and $H^{-\nu}$ by $\langle \cdot, \cdot \rangle$ (see Treves 1967, p. 331), one has

$$\begin{aligned} |r(t) - (E_m r)(t)| &= |\langle r, \delta_t \rangle - \langle E_m r, \delta_t \rangle| \\ &= |\langle r, \delta_t - E_m \delta_t \rangle| \\ &\leq \|r\|_\nu \|\delta_t - E_m(\cdot, t)\|_{-\nu}. \quad (12) \end{aligned}$$

Here we have used the fact that E_m can be defined $H^{-\nu}$ and is a projection operator (see Meyer 1990, p. 43). Applying Theorem A.1 with 2^m in the role of n now gives the result.

6.3 Proof of Theorem 3.3

Following Gasser and Müller (1979, app. 2),

$$\begin{aligned} &\left| \operatorname{var}(\hat{r}(t)) - \frac{\sigma^2}{n} \int_0^1 E_m^2(t, s)\kappa(s)\, ds \right| \\ &= \sigma^2 \left| \sum_{i=1}^n \left(\int_{A_i} E_m(t, s) ds \right)^2 - \frac{1}{n} \int_0^1 E_m^2(t, s)\kappa(s)\, ds \right| \\ &\leq \sigma^2 \sum_{i=1}^n \Big| (s_i - s_{i-1})^2 E_m^2(t, u_i) \\ &\qquad - \frac{1}{n}(s_i - s_{i-1}) E_m^2(t, v_i)\kappa(v_i) \Big| \end{aligned}$$

(where u_i and v_i belong to A_i)

$$\begin{aligned} &= O\left(\frac{1}{n}\right) \sum_{i=1}^n \left| \left(s_i - s_{i-1} - \frac{\kappa(s_i)}{n} \right) E_m^2(t, u_i) \right. \\ &\qquad \left. - \frac{1}{n}(E_m^2(t, v_i)\kappa(v_i) - E_m^2(t, u_i)\kappa(s_i)) \right|. \end{aligned}$$

From (8) the number of terms contributing to the above sum is of order $O(n2^{-m})$. Hence, using (7), the bound $\sup_{t,s} E_m^2(t, s) \leq 2^{2m}$, and the Lipschitz property of κ (which implies $\kappa(v_i) = \kappa(s_i) + O(1/n)$), the last displayed quantity is bounded by

$$\begin{aligned} &O\left(\frac{1}{n}\right) O(n2^{-m}) \left(\rho(n) 2^{2m} + \frac{1}{n} 2^{2m} \frac{1}{n} + \frac{1}{n} 2^{2m} \right. \\ &\qquad \left. \times \sup_i |E_0(2^m t, 2^m v_i) - E_0(2^m t, 2^m u_i)| \right). \end{aligned}$$

Using the compact support and Lipschitz properties of φ, one can show that $E_0(t, \cdot)$ is Lipschitz (uniformly in t), so that

$$\sup_i |E_0(2^m t, 2^m v_i) - E_0(2^m t, 2^m u_i)| = O\left(\frac{2^m}{n}\right).$$

Simplifying, we obtain

$$\left|\operatorname{var}(\hat{r}(t)) - \frac{\sigma^2}{n}\int_0^1 E_m^2(t, s)\kappa(s)\,ds\right| = O(2^m\rho(n)) + O\left(\frac{2^{2m}}{n^2}\right).$$

The proof is completed by appealing to the following lemma.

Lemma 6.1. a. If $h : \mathbb{R} \to \mathbb{R}$ is continuous at t, then

$$\lim_{m\to\infty} 2^{-m}\int_{\mathbb{R}} E_m^2(t^{(m)}, s)h(s)\,ds = h(t)w_0^2.$$

b. If $h : \mathbb{R} \to \mathbb{R}$ is bounded in a neighborhood of t, then

$$\int_{\mathbb{R}} E_m^2(t, s)h(s)\,ds = O(2^m).$$

Proof. Because $2^m t^{(m)} = [2^m t]$ and $E_0(x + k, y + k) = E_0(x, y)$ for all $k \in \mathbb{Z}$,

$$\begin{aligned} 2^{-m}\int_{\mathbb{R}} E_m^2(t^{(m)}, s)h(s)\,ds &= 2^m\int_{\mathbb{R}} E_0^2(2^m t^{(m)}, 2^m s)h(s)\,ds \\ &= 2^m\int_{\mathbb{R}} E_0^2(0, 2^m s - [2^m t])h(s)\,ds \\ &= \int_{\mathbb{R}} E_0^2(0, u)h(t^{(m)} + u2^{-m})\,du \to h(t)\int_{\mathbb{R}} E_0^2(0, u)\,du \end{aligned}$$

as $m \to \infty$. Here we have used the continuity of h at t and the compact support assumption for φ, which implies that $E_0(0, \cdot)$ has compact support. This assumption and the fact that $\{\varphi(\cdot - k) : k \in \mathbb{Z}\}$ is an orthonormal system in $L^2(\mathbb{R})$ give

$$\int_{\mathbb{R}} E_0^2(v, u)\,du = \sum_{k\in\mathbb{Z}} \varphi^2(v - k),$$

so that $\int_{\mathbb{R}} E_0^2(0, u)\,du = w_0^2$, completing the proof of part a. The proof of part b is similar.

6.4 Proof of Theorem 3.4

The Lipschitz condition on r gives

$$r(t) = r(t^{(m)}) + O(2^{-m\gamma}),$$

so by Theorem 3.2 we have $\sqrt{n2^{-m}}(\mathbb{E}\hat{r}_d(t) - r(t)) \to 0$. Write $\hat{r}_d(t) - \mathbb{E}\hat{r}_d(t)$ in the form $\sum_{i=1}^n w_i\varepsilon_i$, where $w_i = w_{in} = \int_{A_i} E_m(t^{(m)}, s)\,ds$. We shall appeal to a central limit theorem for weighted sums of iid random variables (see Eicker 1963) to obtain

$$\frac{\hat{r}_d(t) - \mathbb{E}r_d(t)}{\sqrt{\operatorname{var}(\hat{r}_d(t))}} = \frac{\sum_{i=1}^n w_i\varepsilon_i}{\sigma(\sum_{i=1}^n w_i^2)^{1/2}} \xrightarrow{\mathcal{D}} N(0, 1).$$

To complete the proof, we need to check the Lindeberg-type condition

$$\max_{1\le i\le n} |w_i|^2/\operatorname{var}(\hat{r}_d(t)) \to 0$$

and show that

$$\operatorname{var}(\hat{r}_d(t)) \sim 2^m\sigma^2 w_0^2\kappa(t)/n.$$

From Theorem 3.3 and $\rho(n) = o(1/n)$, we have

$$n2^{-m}\operatorname{var}(\hat{r}_d(t)) = \sigma^2 w_0^2\kappa(t) + o(1) + O(n\rho(n)) + O\left(\frac{2^m}{n}\right) \to \sigma^2 w_0^2\kappa(t).$$

Also using $\max_{1\le i\le n}|w_i|^2 = O(2^{2m}/n^2)$, we have

$$\max_{1\le i\le n}|w_i|^2/\operatorname{var}(\hat{r}_d(t)) = \frac{O(2^m/n)}{n2^{-m}\operatorname{var}(\hat{r}(t))} \to 0,$$

so the Lindeberg condition holds.

6.5 Proof of Theorem 3.5

$\operatorname{Var}(\tilde{f}(t)) = \operatorname{var}(E_m(t, X))/n$ is bounded by

$$\frac{1}{n}\int_{-\infty}^{\infty} E_m^2(t, x)f(x)\,dx = O(2^m/n) \to 0,$$

by Lemma 6.1b. The bias of $\tilde{f}(t)$ is $(E_m f)(t) - f(t)$, which tends to zero by the same argument that was applied to r at the end of the proof of Theorem 3.2. Thus $\tilde{f}$ is pointwise consistent. Denote $g = rf$ and $\tilde{g}(t) = \sum_{i=1}^n E_m(t, X_i)Y_i/n$, so that $\tilde{r} = \tilde{g}/\tilde{f}$. It can be shown, along the lines in which $\operatorname{var}(\tilde{f}(t))$ was handled, except using the conditional variance formula, that $\operatorname{var}(\tilde{g}(t)) = \operatorname{var}(E_m(t, X)Y)/n = O(2^m/n)$. Finally, the bias of $\tilde{g}(t)$ is $(E_m g)(t) - g(t) \to 0$, and we conclude that $\tilde{g}(t)$ is pointwise consistent.

6.6 Proof of Theorem 3.6

Replacing t by $t^{(m)}$ in the proof of consistency of $\tilde{f}(t)$ and using continuity of f at t, we have that $\tilde{f}_d(t)$ consistently estimates $f(t)$. Thus, by

$$\tilde{r}_d - r = (\tilde{g}_d - r\tilde{f}_d)/\tilde{f}_d,$$

we can reduce to considering $\sqrt{n2^{-m}}(\tilde{g}_d - r\tilde{f}_d)(t)$, which can be expressed as

$$\sqrt{\frac{1}{n2^m}}\sum_{i=1}^n (Z_{ni} - \mathbb{E}Z_{ni}) + \sqrt{n2^{-m}}\mathbb{E}Z_{n1}, \tag{13}$$

where $Z_{ni} = E_m(t^{(m)}, X_i)(Y_i - r(t))$. But

$$\begin{aligned} \mathbb{E}Z_{n1} &= (E_m g)(t^{(m)}) - r(t)(E_m f)(t^{(m)}) \\ &= (E_m g)(t^{(m)}) - g(t^{(m)}) - [g(t) - g(t^{(m)})] \\ &\quad - r(t)\{(E_m f)(t^{(m)}) \\ &\qquad - f(t^{(m)}) - [f(t) - f(t^{(m)})]\}, \end{aligned}$$

so the last term in (13) is of order $\sqrt{n2^{-m}}(O(\eta_m) + O(2^{-m\gamma}))$, where η_m is given in Theorem 3.2, and we have used the Lipschitz conditions on r and f (which imply that g is Lipschitz of the same order) to bound the terms inside the square brackets. It follows that $\mathbb{E}Z_{n1} \to 0$ by $n2^{-2m\nu^*} \to 0$. To complete the proof, we apply the Lindeberg–Feller theorem to the first term in (13). First, note that

$$2^{-m}\text{var}(\mathbb{E}(Z_{n1}\,|\,X_1))$$
$$= 2^{-m}\int_{\mathbb{R}} E_m^2(t^{(m)}, x)(r(x) - r(t))^2 f(x)\,dx - [\mathbb{E}Z_{n1}]^2,$$

which tends to zero by Lemma 6.1. Next,

$$2^{-m}\mathbb{E}(\text{var}(Z_{n1}\,|\,X_1))$$
$$= 2^{-m}\int_{\mathbb{R}} E_m^2(t^{(m)}, x)\text{var}(Y\,|\,X = x)f(x)\,dx \to \text{var}(Y\,|\,X = t)f(t)w_0^2,$$

again by Lemma 6.1. Thus, by the conditional variance formula

$$\text{var}(Z_{n1}) = \mathbb{E}(\text{var}(Z_{n1}\,|\,X_1)) + \text{var}(\mathbb{E}(Z_{n1}\,|\,X_1)),$$

we see that the variance of the first term in (13) tends to $\text{var}(Y\,|\,X = t)f(t)w_0^2$. It remains to check the Lindeberg condition, which amounts to showing that

$$\mathbb{E}(U_n^2 I(|U_n| > \delta\sqrt{n})) \to 0, \qquad \forall\,\delta > 0,$$

where $U_n = (Z_{n1} - \mathbb{E}Z_{n1})/\sqrt{\text{var}(Z_{n1})}$. Suppose that $\mathbb{E}(Y^4\,|\,X = x)$ is bounded in a neighborhood of t; the general case of a bounded conditional moment of order $2 + \varepsilon$ is similar. Then, by the Cauchy–Schwarz and Chebyshev inequalities,

$$\mathbb{E}(U_n^2 I(|U_n| > \delta\sqrt{n})) \le [\mathbb{E}U_n^4]^{1/2}(n\delta^2)^{-1/2}.$$

Using the compact support property of φ,

$$\mathbb{E}U_n^4 = O(2^{-2m})\mathbb{E}Z_{n1}^4$$
$$= O(2^{-2m})O(2^{4m})\int_{\mathbb{R}} I(|t - x| < C2^{-m})$$
$$\times [\mathbb{E}(Y^4\,|\,X = x) + C]f(x)\,dx = O(2^m),$$

where C is a generic positive constant. Thus $\mathbb{E}(U_n^2 I(|U_n| > \delta\sqrt{n})) = O(\sqrt{2^m/n}) = 0$, as required.

6.7 Proof of Theorem 3.7

The reader should have a copy of the paper of Eubank and Speckman (1991) on hand before attempting this proof. Using the inequality (Jaffard and Meyer 1989, p. 104)

$$|\partial^l \psi_{j,k}(x)| \le C_1 2^{jl} 2^{j/2}\exp(-C_2 2^j|x - k2^{-j}|),$$
$$x \in \mathbb{R}, \qquad k \in S_j, \quad \text{and} \quad l \le 2q - 2,$$

where C_1 and C_2 are generic constants independent of k, the conclusion of lemma 2 of Eubank and Speckman (1991) becomes

$$\|r' - (T_{mg}r)'\| \le \|r' - (T_m r)'\|$$
$$+ (C_1/\sqrt{2C_2})2^m C_3\|r - (T_m r)\|.$$

The theorem now follows from Eubank and Speckman (1991) by applying the inequality

$$\sum_{j=1}^{\infty}\sum_{k\in S_j} 2^{2j\nu}\langle r, \psi_{j,k}\rangle^2 < \infty$$

for $r \in H_0^\nu$ (theorem 2 of Jaffard and Meyer).

6.8 Proof of Theorem 4.1

From (9), we get the following expansion:

$$\hat{\alpha}(t) = \int_0^1 E_m(t, s)\alpha(s)J_n(s)\,ds$$
$$+ \int_0^1 E_m(t, s)\frac{J_n(s)}{Y_n(s)}\,dM_n(s). \quad (14)$$

Because the last integral is a mean zero martingale evaluated at 1, we have

$$\mathbb{E}\hat{\alpha}(t) = \mathbb{E}\int_0^1 E_m(t, s)\alpha(s)J_n(s)\,ds$$
$$= \mathbb{E}\left[\int_0^1 E_m(t, s)\alpha(s)(J_n(s) - 1)\,ds\right]$$
$$+ \int_0^1 E_m(t, s)\alpha(s)\,ds. \quad (15)$$

The last term in (15) is the same as (11), with r replaced by α. Using the Cauchy–Schwarz inequality and Lemma 6.1, the first term in (15) is seen to be of order $2^{m/2}O(\delta_n^{1/2})$.

6.9 Proof of Theorem 4.2

First, note that $\hat{\alpha}_d(t) - \alpha(t)$ can be written as

$$\int_0^1 E_m(t^{(m)}, s)\alpha(s)(J_n(s) - 1)\,ds$$
$$+ \left[\int_0^1 E_m(t^{(m)}, s)\alpha(s)\,ds - \alpha(t^{(m)})\right] + (\alpha(t^{(m)})$$
$$- \alpha(t)) + \int_0^2 E_m(t^{(m)}, s)\frac{J_n(s)}{Y_n(s)}\,dM_n(s). \quad (16)$$

Along the lines of the previous proof, we see that the first term in (16) is of order $2^m O(\delta_n)$ and the second term is of order $O(\eta_m)$. The third term is of order $O(2^{-m\gamma})$, because α is Lipschitz of order γ. The second moment of the stochastic integral is $n^{-1}\int_0^1 E_m^2(t^{(m)}, s)\alpha(s)\tau_n(s)\,ds$, where $\tau_n(s) = n\mathbb{E}[J_n(s)/Y_n(s)]$. It follows along the lines of the proof of Lemma 6.1(a), using $\alpha\tau_n/n$ in place of h, and by our assumptions on (τ_n) and τ, that

$$\mathbb{E}\left[\int_0^1 E_m(t^{(m)}, s)\frac{J_n(s)}{Y_n(s)}\,dM_n(s)\right]^2$$
$$= \frac{2^m}{n}\frac{\alpha(t)}{\tau(t)}w_0^2 + 2^m o(n^{-1}).$$

The second part of the theorem is proved in a similar fashion, except using part (b) of Lemma 6.1.

6.10 Proof of Theorem 4.3

By our assumptions, $\sqrt{n2^{-m}}(\hat{\alpha}_d(t) - \alpha(t))$ is asymptotically equivalent to $\sqrt{n2^{-m}}\int_0^1 E_m(t^{(m)}, s)J_n(s)/Y_n(s)\,dM_n(s)$, which is the value at 1 of a martingale having quadratic variation asymptotically equivalent to $2^{-m}\int_0^1 E_m^2(t^{(m)}, s)\alpha(s)/\tau(s)\,ds$ at 1. The previous proof gives that the latter quantity tends to $\alpha(t)w_0^2/\tau(t)$. The result follows using

Rebolledo's martingale central limit theorem (cf. Ramlau-Hansen 1983).

APPENDIX: AN EXTENSION OF SCHOMBURG'S THEOREM AND ITS WAVELET APPLICATION

Schomburg's (1990) original result gives the rate of convergence of certain sequences of type δ to the delta distribution centered at the origin in $\mathbb{R}^d$. We need to extend this result to deal with approximations for δ_t, the delta distribution centered at $t \in \mathbb{R}^d$. Following Schomburg, we allow $d \geq 1$, although we really only need $d = 1$. The sign of ν is reversed from Condition 1. Let $\nu < -d/2$, $g(\cdot, t) \in H^\nu(\mathbb{R}^d)$ for all t and define the sequence $(g_n(\cdot, t))_{n\geq 1} \subset H^\nu(\mathbb{R}^d)$ for each t by

$$\langle g_n(\cdot, t), \phi \rangle = \left\langle g(\cdot, nt), \phi\left(\frac{\cdot}{n}\right)\right\rangle \quad \text{for} \quad \phi \in \mathcal{S}(\mathbb{R}^d),$$

where $\mathcal{S}(\mathbb{R}^d)$ is the space of rapidly decreasing test functions (see Hörmander 1989, p. 160). For a classical function g, one has $g_n(s, t) = n^d g(ns, nt)$. The Fourier transform of a function $h \in L^1(\mathbb{R}^d)$ is defined by $\hat{h}(\xi) = \int_{\mathbb{R}^d} e^{-ix\cdot\xi} h(x)\,dx$, $\xi \in \mathbb{R}^d$. In this Appendix $\hat{h}$ denotes the Fourier transform rather than an estimator of h.

Theorem A.1. Suppose that

$$\sup_t \| g(\cdot, t)\|_\nu < \infty \tag{A.1}$$

and for some $\gamma > 0$ there exists a neighborhood U of 0 in $\mathbb{R}^d$ such that

$$\xi \mapsto |\xi|^{-\gamma}(\hat{g}(\xi, t) - e^{-i\xi\cdot t}) \tag{A.2}$$

belongs to $L^\infty(U)$ for each $t \in \mathbb{R}^d$. Then

$$\begin{aligned} \| g_n(\cdot, t) - \delta_t\|_\nu &= O(n^{\nu+(d/2)}) \quad \text{if} -\nu < \gamma + \frac{d}{2}, \\ &= O(n^{-\gamma}\sqrt{\log n}) \quad \text{if} -\nu = \gamma + \frac{d}{2}, \\ &= O(n^{-\gamma}) \quad \text{if} -\nu > \gamma + \frac{d}{2} \end{aligned}$$

as $n \to \infty$.

Proof. Clearly, one may take U as the unit ball in $\mathbb{R}^d$. Noting that $\hat{g}_n(\xi, t) = \hat{g}(\xi/n, nt)$, we have

$$\begin{aligned} \| g_n(\cdot, t) - \delta_t\|^2 &= \int_{\mathbb{R}^d} (1 + |\xi|^2)^\nu |\hat{g}_n(\xi, t) - e^{-i\xi\cdot t}|^2\, d\xi \\ &= \int_{\mathbb{R}^d} (1 + |\xi|^2)^\nu |\hat{g}(\xi/n, nt) - e^{-i\xi\cdot t}|^2\, d\xi \\ &= n^d \int_{\mathbb{R}^d} (1 + n^2|\tau|^2)^\nu |\hat{g}_n(\tau, nt) - e^{-in\tau\cdot t}|^2\, d\tau. \end{aligned}$$

Now split the integration into

$$\begin{aligned} &n^d \int_{|\tau|\geq 1} (1 + n^2|\tau|^2)^\nu |\hat{g}_n(\tau, nt) - e^{-in\tau\cdot t}|^2\, d\tau \\ &\leq 2n^{2\nu+d} \int_{|\tau|\geq 1} |\tau|^{2\nu} |\hat{g}(\tau, nt)|^2\, d\tau + 2n^{2\nu+d} \int_{|\tau|\geq 1} |\tau|^{2\nu}\, d\tau \\ &= O(n^{2\nu+d}) \end{aligned}$$

by (A.1), and

$$\begin{aligned} &n^d \int_{|\tau|\leq 1} (1 + n^2|\tau|^2)^\nu |\hat{g}_n(\tau, nt) - e^{-in\tau\cdot t}|^2\, d\tau \\ &\qquad \leq Cn^d \int_{|\tau|\leq 1} (1 + n^2|\tau|^2)^\nu |\tau|^{2\gamma}\, d\gamma \end{aligned}$$

by (A.2), with $\xi = \tau$ and t set to nt. The remainder of the proof is routine integration (see Schomburg 1990 for details).

Lemma A.2. The function $g(x, y) = E_0(x, y)$ satisfies (A.1) and (A.2) of Theorem A.1.

Proof. Noting that

$$E_0(s, t) = \sum_{k\in\mathbf{Z}} \varphi(s - k)\varphi(t - k),$$

we have

$$\begin{aligned} \hat{E}_0(\xi, t) &= \int_{\mathbb{R}} E_0(s, t) e^{-i\xi s}\, ds = \sum_{k\in\mathbf{Z}} \varphi(t - k) \int_{\mathbb{R}} \varphi(s - k) e^{-i\xi s}\, ds \\ &= \hat{\varphi}(\xi) \sum_{k\in\mathbf{Z}} \varphi(t - k) e^{-ik\xi} = \hat{\varphi}(\xi) \sum_{k\in\mathbf{Z}} \varphi(t + k) e^{ik\xi}. \end{aligned}$$

Changing k to $k - [t]$ and setting $u = t - [t]$, we have

$$\sum_{k\in\mathbf{Z}} \varphi(t + k) e^{ik\xi} = e^{-i\xi t} e^{i\xi u} \sum_{k\in\mathbf{Z}} \varphi(u + k) e^{ik\xi}.$$

Because φ has compact support, (A.1) holds if $\varphi \in H^\nu$ for $\nu < 0$. But $\varphi \in L^2(\mathbb{R}) \subset H^\nu$ for $\nu < 0$, as required. Next, by (4) and again because φ has compact support, we have

$$\sum_{k\in\mathbf{Z}} \varphi(u + k) e^{ik\xi} = 1 + \sum_{k\in\mathbf{Z}} \varphi(u + k)(e^{ik\xi} - 1) = 1 + O(|\xi|)$$

as $\xi \to 0$. Thus, by Condition 6,

$$\hat{E}_0(\xi, t) = (1 + O(|\xi|)) e^{-i\xi t} e^{i\xi u}(1 + O(|\xi|)) = e^{-i\xi t}(1 + O(|\xi|)),$$

so (A.2) holds.

[Received May 1992.]

REFERENCES

Aalen, O. O. (1978), "Nonparametric Inference for a Family of Counting Processes," *The Annals of Statistics,* 6, 701–726.

Antoniadis, A., and Carmona, R. (1990), "Multiresolution Analyses and Wavelets for Density Estimation," technical report, University of California at Irvine.

Breiman, L., Meisel, W., and Purcell, E. (1977), "Variable Kernel Regression Estimates of Probability Densities," *Technometrics,* 19, 135–144.

Chu, C.-K., and Marron, J. S. (1991), "Choosing a Kernel Regression Estimator," *Statistical Science,* 6, 404–436.

Chui, K. (1992), *Wavelets: A Tutorial in Theory and Applications,* Boston: Academic Press.

Collomb, G. (1981), "Estimation Nonparamétrique de la Régression: Revue Bibliographique," *International Statistical Reviews,* 49, 75–93.

Daubechies, I. (1990), "Orthonormal Bases of Compactly Supported Wavelets," *Communications on Pure and Applied Mathematics,* 49, 906–996.

Daubechies, I., and Lagarias, J. C. (1988a), "Two-Scale Difference Equations: I. Global Regularity of Solutions," technical report, Bell Laboratories.

——— (1988b), "Two-Scale Difference Equations II. Local Regularity, Infinity Products of Matrices and Fractals," technical report, Bell Laboratories.

Doukhan, P. (1991), "Consistency of Delta-Sequence Estimates of a Density or of a Regression Function for a Weakly Dependent Stationary Sequence," Séminaire de Statistique d'Orsay, Université Paris Sud.

Eicker, F. (1963), "Central Limit Theorems for Families of Sequences of Random Variables," *Annals of Mathematical Statistics,* 34, 439–446.

Eubank, R. (1988), *Spline Smoothing and Nonparametric Regression,* New York: Marcel Dekker.

Eubank, R., and Speckman, P. (1991), "Convergence Rates for Trigonometric and Polynomial-Trigonometric Regression Estimators," *Statistical and Probability Letters,* 11, 119–124.

Gasser, T., and Müller, H. (1979), *Curve Estimation,* Heidelberg: Springer-Verlag.

——— (1984), "Estimating Regression Functions and Their Derivatives by the Kernel Method," *Scandinavian Journal of Statistics,* 11, 171–185.

Grégoire, G. (1991), "Bandwidth Selection for the Intensity of Point Processes. Application to Hazard Rate Models for Censored Data," technical report, Université Joseph Fourier.

Härdle, W. (1984), "Robust Regression Function Estimation," *Journal of Multivariate Analysis,* 14, 169–180.

——— (1990), *Applied Nonparametric Regression,* Cambridge, U.K.: Cambridge University Press.

Härdle, W., and Marron, J. S. (1985), "Optimal Bandwidth Selection in Nonparametric Regression Function Estimation," *The Annals of Statistics,* 13, 1465–1481.

Hörmander, L. (1989), *The Analysis of Linear Partial Differential Operators,* Vol. 2, London: Springer-Verlag.

Isogai, E. (1990), "Nonparametric Estimation of a Regression Function by Delta Sequences," *Ann. Inst. Statist. Math.,* 42, 699–708.

Jaffard, S., and Meyer, Y. (1989), "Bases d'ondelettes dans des ouverts de $\mathbb{R}^n$," *Journal Mathematique Pures et Appliqués,* 68, 95–108.

Mallat, S. (1989), "Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\mathbb{R})$," *Transactions of the American Mathematical Society,* 315, 69–87.

Müller, H. G. (1985), "Empirical Bandwidth Choice for Nonparametric Kernel Regression by Means of Pilot Estimators," *Statistics & Decisions,* Supplementary Issue 2, 193–206.

Meyer, Y. (1990), *Ondelettes et Opérateurs I: Ondelettes,* Paris: Hermann.

Nadaraya, E. A. (1964), "On Estimating Regression," *Theory of Probability and Its Applications,* 9, 141–142.

Ramlau-Hansen, H. (1983), "Smoothing Counting Processes by Means of Kernel Functions," *The Annals of Statistics,* 11, 453–466.

Rauzy, G. (1976), *Propriétés Statistiques de Suites Arithmétiques,* Paris: Presses Universitaires de France.

Rioul, O., and Vetterli, M. (1991), "Wavelets and Signal Processing," *IEEE Signal Processing Magazine,* 8, 14–38.

Schomburg, B. (1990), "On the Approximation of the Delta Distribution in Sobolev Spaces of Negative Order," *Applicable Analysis,* 36, 89–93.

Silverman, B. W. (1985), "Some Aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting" (with discussion), *Journal of the Royal Statistical Society,* Ser. B, 47, 1–52.

Strang, G. (1989), "Wavelets and Dilation Equations: A Brief Introduction, *SIAM Review,* 31, 614–627.

Treves, F. (1967), *Topological Vector Spaces, Distributions and Kernels,* New York: Academic Press.

Walter, G., and Blum, J. (1979), "Probability Density Estimation Using Delta Sequences," *The Annals of Statistics,* 7, 328–340.

Watson, G. S. (1964), "Smooth Regression Analysis," *Sankhya,* Ser. A, 26, 359–372.

Wegman, E. J. (1991), "The Straight Scoop on Wavelets and Nonparametric Function Estimation," in *Proceedings of the 37th Conference on Design of Experiments in Army, Research, Development and Testing.*