The Annals of Statistics
1999, Vol. 27, No. 6, 2033–2053

ADAPTIVE WAVELET ESTIMATOR FOR NONPARAMETRIC DENSITY DECONVOLUTION

BY MARIANNA PENSKY AND BRANI VIDAKOVIC

University of Central Florida and Duke University

The problem of estimating a density g based on a sample $X_1, X_2, \dots, X_n$ from $p = q * g$ is considered. Linear and nonlinear wavelet estimators based on Meyer-type wavelets are constructed. The estimators are asymptotically optimal and adaptive if g belongs to the Sobolev space H^α. Moreover, the estimators considered in this paper adjust automatically to the situation when g is supersmooth.

1. Introduction. Let θ and ε be independent random variables with density functions g and q, respectively, where g is unknown and q is known. One observes a sample of random variables,

$$X_i = \theta_i + \varepsilon_i, \qquad i = 1, 2, \dots, n. \tag{1.1}$$

The objective is to estimate the density function g. In this situation the density function p of X_i, $i = 1, ..., n$, is the convolution of q and g,

$$p(x) = \int_{-\infty}^{\infty} q(x-\theta) g(\theta)\, d\theta. \tag{1.2}$$

Hence the problem of estimating g in (1.2) is called a deconvolution problem. The problem arises in many applications [see, e.g., Desouza (1991), Louis (1991), Zhang (1992)] and, therefore, it was studied extensively in the last decade. The most popular approach to the problem was to estimate $p(x)$ by a kernel estimator and then solve equation (1.2) using a Fourier transform [see Carroll and Hall (1988), Devroye (1989), Diggle and Hall (1993), Efromovich(1997), Fan (1991a, c), Liu and Taylor (1989), Masry (1991, 1993a, b), Stefansky (1990), Stefansky and Carroll (1990), Taylor and Zhang (1990), Zhang (1990)]. Fan (1991a, 1993) proved that the estimators of $g(\theta)$ are asymptotically optimal pointwise and globally, if the kernel has a limited bandwidth, that is, the Fourier transform of the kernel has bounded support. The estimators based on the deconvolution of kernel estimators and similar methods were studied in many different contexts: the asymptotic normality was established [see, e.g., Fan (1991b), Piterbarg and Penskaya (1993), Masry (1993a)]; the case of dependent ε_i was examined [Masry (1991, 1993b)], etc.

This present paper deals with the estimation of a deconvolution density using a wavelet decomposition. The underlying idea is to present $g(\theta)$ via a

Received April 1998; revised September 1999.
AMS 1991 *subject classifications*. Primary 62G05; secondary 62G07.
Key words and phrases. Mixing distribution, wavelet transformation, Sobolev space, Meyer wavelet.

wavelet expansion and then to estimate the coefficients using a deconvolution algorithm. The proposed approach is based on orthogonal series methods for the estimation of a prior density [see Walter (1981), Penskaya (1985)], and also on modern developments of wavelet techniques in curve estimation [see Antoniadis, Gregoire and McKeague (1994), Abramovich and Silverman (1997), Donoho and Johnstone (1995), Hall and Patil (1995), Hall, Penev, Kerkyacharian and Picard (1997), Hall, Kerkyacharian and Picard (1998), Kerkyacharian and Picard (1992), Masry (1994) and Walter (1994), among others].

Estimation of the density $g(\theta)$ is conducted in the well-known Sobolev space H^α which describes the level of smoothness of a deconvolution density in terms of its characteristic function $\tilde{g}$. Estimation of $g(\theta)$ splits into two different cases: the case when the distribution of the error ε is supersmooth, that is, the Fourier transform $\tilde{q}$ of q has exponential descent, and the case when $\tilde{q}$ has polynomial descent. In the first case, even when α is unknown, the linear wavelet estimator proposed in the paper allows an adaptive choice of parameters that ensures the optimal convergence rate of the estimator. In the case where $\tilde{q}$ has polynomial descent, the linear wavelet estimator fails to provide the optimal convergence rate if α is unspecified. In this case a nonlinear adaptive wavelet estimator is constructed which achieves the optimal convergence rate.

The estimators proposed in this paper are based on Meyer-type wavelets rather than on wavelets with bounded support. Meyer-type wavelets form a subset of the set of band-limited wavelets that allow immediate deconvolution. It should be noted that the nonlinear wavelet estimator constructed in this paper is based on a "global thresholding" which is somewhat different from the "block thresholding" suggested by Hall, Penev, Kerkyacharian and Picard (1997): in the "global thresholding" procedure all coefficients of the same level are thresholded simultaneously, while "block thresholding" groups together only a finite number of coefficients.

The estimators based on Meyer-type wavelets are asymptotically optimal in the sense that for $g \in H^\alpha$ the rates of convergence of the mean integrated squared error (MISE) cannot be improved [see Fan (1993)]. Moreover, the estimators obtained in this paper adjust automatically to the situation when $g(\theta)$ is supersmooth. In this case, without any change of parameters, both the linear and the nonlinear wavelet estimators achieve better convergence rates. Namely, if both $g(\theta)$ and $q(x)$ are supersmooth, then the linear wavelet estimator has a polynomial rate of convergence which is better than the logarithmic rate of convergence that can be attained for $g \in H^\alpha$. If $g(\theta)$ is supersmooth and $\tilde{q}$ has polynomial descent, then the MISE of the nonlinear wavelet estimator is $O(n^{-1} \ln^\nu n)$ with $\nu > 0$ as $n \to \infty$.

This article is organized in the following way. In Section 2 we give a brief description of Meyer-type wavelets and derive the linear and nonlinear wavelet estimators of $g(\theta)$. In Section 3 we investigate asymptotic behavior of the estimators when $g(\theta) \in H^\alpha$. The case of supersmooth $g(\theta)$ is considered in Section 4. In Section 5 we illustrate the theory by examples. Section 6 concludes the paper with discussion. Section 7 contains proofs of the theorems.

2. Wavelet estimation of a deconvolution density based on Meyer-type wavelets. Throughout this paper we use the notation $\mathscr{F}[f](\omega)$ or $\tilde{f}(\omega)$ for the Fourier transform $\int_{-\infty}^{\infty} \exp(-i\omega x) f(x)\,dx$ of a function $f(x)$ and $\mathscr{F}^{-1}[\tilde{f}](x)$ for the inverse Fourier transform of $\tilde{f}(\omega)$. Let $\|f\|_c = \sup_y |f(y)|$ for any continuous function f and $\|f\|_{L_k} = \{\int_{-\infty}^{\infty} |f(x)|^k\,dx\}^{1/k}$. Assume that $g(\theta)$ is square integrable and that $\tilde{q}(\omega)$ does not vanish for real ω.

If $\varphi(\theta)$ and $\psi(\theta)$, respectively, are a scaling function and a wavelet generated by an orthonormal multiresolution decomposition of $L^2(-\infty,\infty)$, then for any integer m the density function $g(\theta)$ allows the following representation:

$$g(\theta) = \sum_{k\in\mathbf{Z}} a_{m,k}\varphi_{m,k}(\theta) + \sum_{k\in\mathbf{Z}}\sum_{j=m}^{\infty} b_{j,k}\psi_{j,k}(\theta), \tag{2.1}$$

where $\varphi_{m,k}(\theta) = 2^{m/2}\varphi(2^m\theta - k)$ and $\psi_{j,k}(\theta) = 2^{j/2}\psi(2^j\theta - k)$, and the coefficients $a_{m,k}$ and $b_{j,k}$ have the forms

$$a_{m,k} = \int_{-\infty}^{\infty} \varphi_{m,k}(\theta) g(\theta)\,d\theta, \qquad b_{j,k} = \int_{-\infty}^{\infty} \psi_{j,k}(\theta) g(\theta)\,d\theta, \tag{2.2}$$

respectively.

A special class of wavelets are band-limited wavelets, the Fourier transform of which have bounded support [see Hernández and Weiss (1996)]. In this article, we shall use a particular type of band-limited wavelet, a Meyer-type wavelet [see Walter (1994), Zayed and Walter (1996)]. Let P be a probability measure with support in $[-\pi/3, \pi/3]$. Define the scaling function $\varphi(x)$ and the wavelet function $\psi(x)$ as the functions whose Fourier transforms are

$$\tilde{\varphi}(\omega) = \left[\int_{\omega-\pi}^{\omega+\pi} dP\right]^{1/2}, \qquad \tilde{\psi}(\omega) = \exp(-i\omega/2)\left[\int_{|\omega|/2-\pi}^{|\omega|-\pi} dP\right]^{1/2}, \tag{2.3}$$

the nonnegative square roots of the integrals. Then $\tilde{\varphi}(\omega)$ and $\tilde{\psi}(\omega)$ both have bounded support: $\operatorname{supp}\tilde{\varphi} \subset [-4\pi/3, 4\pi/3]$ and $\operatorname{supp}\tilde{\psi} \subset \Omega_1 \cup \Omega_2$ with

$$\Omega_1 = [-8\pi/3, -2\pi/3], \qquad \Omega_2 = [2\pi/3, 8\pi/3]. \tag{2.4}$$

Moreover, $\tilde{\varphi}(\omega) = 1$ if $|\omega| < 2\pi/3$. In order to ensure that $\varphi(x)$ and $\psi(x)$ have sufficient rates of descent as $|x| \to \infty$, we choose P to be smooth, so that the functions $\tilde{\varphi}(\omega)$ and $\tilde{\psi}(\omega)$ are $s \geq 2$ times continuously differentiable on $(-\infty,\infty)$. Since $\tilde{\varphi}(\omega)$ and $\tilde{\psi}(\omega)$ both have bounded support, this implies that

$$C_\varphi = \sup_x \left[|\varphi(x)|\,(|x|^s + 1)\right] < \infty, \qquad C_\psi = \sup_x \left[|\psi(x)|\,(|x|^s + 1)\right] < \infty. \tag{2.5}$$

The coefficients $a_{m,k}$ and $b_{j,k}$ can be viewed as mathematical expectations of the functions $u_{m,k}$ and $v_{j,k}$

$$a_{m,k} = \int_{-\infty}^{\infty} u_{m,k}(x) p(x)\,dx, \qquad b_{j,k} = \int_{-\infty}^{\infty} v_{j,k}(x) p(x)\,dx, \tag{2.6}$$

provided that $u_{m,k}(x)$ and $v_{j,k}(x)$ are solutions of the following equations:

$$\int_{-\infty}^{\infty} q(x-\theta) u_{m,k}(x)\,dx = \varphi_{m,k}(\theta), \qquad \int_{-\infty}^{\infty} q(x-\theta) v_{j,k}(x)\,dx = \psi_{j,k}(\theta).$$

Taking the Fourier transform of both sides, we obtain $u_{m,k}(x) = 2^{m/2} \cdot U_m(2^m x - k)$, $v_{j,k}(x) = 2^{j/2} V_j(2^j x - k)$, where $U_m(\cdot)$ and $V_j(\cdot)$ are the inverse Fourier transforms of the functions

$$\tilde{U}_m(\omega) = \tilde{\varphi}(\omega)/\tilde{q}(-2^m \omega), \qquad \tilde{V}_j(\omega) = \tilde{\psi}(\omega)/\tilde{q}(-2^j \omega), \tag{2.7}$$

respectively. Therefore, estimating $a_{m,k}$ and $b_{j,k}$ by

$$\hat{a}_{m,k} = n^{-1} \sum_{l=1}^{n} 2^{m/2} U_m(2^m X_l - k), \qquad \hat{b}_{j,k} = n^{-1} \sum_{l=1}^{n} 2^{j/2} V_j(2^j X_l - k) \tag{2.8}$$

and truncating the series (2.1), we obtain a linear wavelet estimator

$$\hat{g}_n^{(L)}(\theta) = \sum_{k \in \mathbf{Z}} \hat{a}_{m,k}\, \varphi_{m,k}(\theta), \tag{2.9}$$

and a nonlinear wavelet estimator of $g(\theta)$,

$$\hat{g}_n^{(N)}(\theta) = \sum_{k \in \mathbf{Z}} \hat{a}_{m,k} \varphi_{m,k}(\theta) + \sum_{j=m}^{m+r} \left[\sum_{k \in \mathbf{Z}} \hat{b}_{j,k} \psi_{j,k}(\theta) \right] I\left(\sum_{k \in \mathbf{Z}} \hat{b}_{j,k}^2 > \delta_{j,n}^2 \right). \tag{2.10}$$

Note that the estimator (2.10) has the block thresholding which is different from the block thresholding used by Hall, Penev, Kerkyacharian and Picard (1997) and Hall, Kerkyacharian and Picard (1998), who dealt with the estimation of a density function based on direct observations by wavelets with bounded support. Hall, Penev, Kerkyacharian and Picard (1997), and Hall, Kerkyacharian and Picard (1998), partitioned coefficients $b_{j,k}$ into blocks $\mathscr{B} = \{b_{j,k}\colon (j-1)l < k < jl\}$ of the length l and then thresholded all the coefficients of a block simultaneously. In the present paper all the coefficients $b_{j,k}$, $k \in \mathbf{Z}$, are thresholded together.

At first glance, the estimators (2.9) and (2.10) seem computationally intractable since their constructions involve the calculation of infinite series. However, under very nonrestrictive conditions, the infinite series estimators (2.9) and (2.10) can be replaced by finite series estimators,

$$\hat{g}_n^{(LF)}(\theta) = \sum_{|k| \le K_n} \hat{a}_{m,k} \varphi_{m,k}(\theta), \tag{2.11}$$

$$\begin{aligned} \hat{g}_n^{(NF)}(\theta) = & \sum_{|k| \le M_n} \hat{a}_{m,k} \varphi_{m,k}(\theta) \\ & + \sum_{j=m}^{m+r} \left[\sum_{|k| \le L_n} \hat{b}_{j,k} \psi_{j,k}(\theta) \right] I\left(\sum_{|k| \le L_n} \hat{b}_{j,k}^2 > \delta_{j,n}^2 \right) \end{aligned} \tag{2.12}$$

without any loss in the rate of convergence.

3. Asymptotic behavior of wavelet estimators. To investigate asymptotic properties of the estimators (2.9) and (2.10), we assume that the density g belongs to the following class:

$$\mathscr{S}_\alpha(A_\alpha) = \{g \in H^\alpha\colon\ \|g\|_\alpha \le A_\alpha,\ \alpha > 0\} \tag{3.1}$$

where $\|g\|_\alpha$ is the norm in the Sobolev space H^α,

$$\|g\|_\alpha = \left\{\int_{-\infty}^{\infty} |\tilde{g}(\omega)|^2(\omega^2+1)^\alpha\, dw\right\}^{1/2} < \infty.$$

We shall measure the performance of an estimator $g_n(\theta)$ by

$$\text{MISE}\ (g_n) = \mathbf{E}\int_{-\infty}^{\infty} \big(g_n(\theta) - g(\theta)\big)^2\, d\theta. \tag{3.2}$$

Let

$$\begin{aligned} \Delta_1(m) &= \int_{-\infty}^{\infty} |\tilde{\varphi}(\omega)|^2 |\tilde{q}(2^m\omega)|^{-2}\, d\omega,\\ \Delta_k(j) &= \int_{-\infty}^{\infty} |\tilde{\psi}(\omega)|^k\ |\tilde{q}(2^j\omega)|^{-k}\, d\omega, \qquad k = 2, 4. \end{aligned} \tag{3.3}$$

The following theorem establishes the upper bound for the MISE of the linear wavelet estimator (2.9) uniformly over the class $\mathscr{S}_\alpha(A_\alpha)$ defined in (3.1).

THEOREM 1.

$$\sup_{g\in\mathscr{S}_\alpha} \text{MISE}\left(\hat{g}_n^{(L)}\right) \le 2\,\pi^{-1}(2\pi/3)^{-2\alpha}\|\psi\|_c^2\, A_\alpha^2\, 2^{-2m\alpha} + n^{-1}2^{m+1}\Delta_1(m). \tag{3.4}$$

COROLLARY 1. *If* $|\tilde{q}(\omega)| \ge A_0(w^2+1)^{-\gamma/2}\exp\{-B|\omega|^\beta\}$ *and* m *is such that*

$$2^m = \begin{cases} n^{1/(2\alpha+2\gamma+1)}, & \text{if } B = 0,\\ \left[\left(2B\left(\dfrac{4\pi}{3}\right)^\beta + \Lambda\right)^{-1} \ln n\right]^{1/\beta}, & \text{if } B > 0, \end{cases} \tag{3.5}$$

with $\Lambda > 0$, *then*

$$\sup_{g\in\mathscr{S}_\alpha} \text{MISE}\left(\hat{g}_n^{(L)}\right) = \begin{cases} O\big(n^{-2\alpha/(2\alpha+2\gamma+1)}\big), & \text{if } B = 0,\\ O\big((\ln n)^{-2\alpha/\beta}\big), & \text{if } B > 0. \end{cases} \tag{3.6}$$

Observe that the rates of convergence in (3.6) coincide with the optimal rate of convergence [see Fan (1993)]. Also, in the case of exponential descent of $\tilde{q}(\omega)$, the linear wavelet estimator is adaptive, that is, the choice of the parameter m does not depend on the unknown smoothness α of the density $g(\theta)$. However, in the case of polynomial descent, the estimator (2.9) fails to provide the optimal convergence rate when α is unknown. This difficulty can be overcome by using the nonlinear estimator (2.10).

THEOREM 2. *Suppose* $|\tilde{q}(\omega)| \geq A_0(w^2+1)^{-\gamma/2}$. *Let* $\hat{g}_n^{(N)}$ *be the estimator* (2.10) *with* $m = (2+\varepsilon)\log_2(\ln n)$ *where* $\varepsilon > 0$, $m + r = (2\gamma+1)^{-1}\log_2 n$ *and* $\delta_{j,n} = 2^{j(\gamma+0.5)}\delta_n$. *If* $\delta_n = \delta_0 n^{-1/2}$ *with* $\delta_0 \geq 2\sqrt{2}K_2$ *and* $\Delta_4(j)/\Delta_2^2(j) \leq C_0$ *for any* j, *then*

$$\sup_{g\in\mathscr{S}_\alpha} \mathrm{MISE}\big(\hat{g}_n^{(N)}\big) = O\big(n^{-2\alpha/(2\alpha+2\gamma+1)}\big). \tag{3.7}$$

Here K_2 *is an absolute constant* [*see* (A.1)].

The reasoning behind Theorem 2 is as follows. If the value of α were known, then the best choice of m in the linear estimator (2.9) would be $m_{\mathrm{opt}} \sim (2\alpha + 2\gamma+1)^{-1}\log_2 n$. Since α is unknown, we can only tell that for any λ, the optimal value of m lies between $\lambda\log_2(\ln n)$ and $(2\gamma+1)^{-1}\log_2 n$. Thus we construct the nonlinear estimator (2.10) with $m = (2+\varepsilon)\log_2(\ln n)$, which is smaller than the optimal value m_{opt} and $m + r = (2\gamma+1)^{-1}\log_2 n$. By doing this, we include all terms with $j \leq (2+\varepsilon)\log_2(\ln n)$ and exclude the terms with $j > (2\gamma+1)^{-1}\log_2 n$. The terms with $(2+\varepsilon)\log_2(\ln n) < j \leq (2\gamma+1)^{-1}\log_2 n$ are included only if $\sum_{k\in\mathbf{Z}} \hat{b}_{j,k}^2 \geq \delta_{j,n}^2$, where $\delta_{j,n}^2 \sim n^{-1}\sum_{k\in\mathbf{Z}} \mathrm{Var}\,\hat{b}_{j,k}$. It enables one to include only terms whose variance does not exceed $O(n^{-2\alpha/(2\alpha+2\gamma+1)})$ and, therefore, to ensure the optimal convergence rate. Note that in order that $m < m_{\mathrm{opt}}$ for finite values of n, the value of ε should not be large, say, $\varepsilon < \varepsilon_0$, where ε_0 is chosen in advance.

In order to replace the estimators (2.9) and (2.10) by their finite series counterparts, we assume that g has a certain rate of descent as $|\theta| \to \infty$. Namely, let

$$\mathscr{S}_\alpha^*(A_\alpha, A_g) = \Big\{g\colon\ g \in \mathscr{S}_\alpha(A_\alpha),\ \sup_\theta[|\theta| g(\theta)] \leq A_g\Big\}, \tag{3.8}$$

where $\mathscr{S}_\alpha(A_\alpha)$ is defined in (3.1) and assume that $g \in \mathscr{S}_\alpha^*(A_\alpha, A_g)$. Note that the condition $\sup[|\theta| g(\theta)] < \infty$ is very nonrestrictive and holds for every familiar p.d.f. The following theorem shows that the rate of convergence of (2.11) and (2.12) uniformly over $\mathscr{S}_\alpha^*(A_\alpha, A_g)$ is the same as the rate of convergence of (2.9) and (2.10), respectively, uniformly over $\mathscr{S}_\alpha(A_\alpha)$.

THEOREM 3. *Assume that the assumptions of Corollary* 1 *and Theorem* 2 *are valid and* K_n, M_n *and* L_n *are such that*

$$\lim_{n\to\infty} nK_n^{-1} = 0, \qquad \lim_{n\to\infty} nM_n^{-1} = 0, \qquad \lim_{n\to\infty} n^{(2\gamma+2)/(2\gamma+1)}L_n^{-1} = 0.$$

Then the estimators (2.11) *and* (2.12), *with the same choice of parameters* m, r, *and* $\delta_{j,n}$ *as in Corollary* 1 *and Theorem* 2, *have the following rates of convergence uniformly over* $\mathscr{S}_\alpha^*(A_\alpha, A_g)$:

$$\sup_{g\in\mathscr{S}_\alpha^*} \mathrm{MISE}\big(\hat{g}_n^{(LF)}\big) = O\big((\ln n)^{-2\alpha/\beta}\big) \quad \textit{if } B > 0, \tag{3.9}$$

$$\sup_{g\in\mathcal{S}_\alpha^*} \mathrm{MISE}\big(\hat{g}_n^{(LF)}\big) \sim \sup_{g\in\mathcal{S}_\alpha^*} \mathrm{MISE}\big(\hat{g}_n^{(NF)}\big) = O\big(n^{-2\alpha/(2\alpha+2\gamma+1)}\big) \quad \textit{if } B = 0. \tag{3.10}$$

4. Estimation in the case of a supersmooth $g(\theta)$. The asymptotic results provided by Theorems 1–3 are not very optimistic: if $q(x)$ is supersmooth, the estimator has a logarithmic convergence rate. Nevertheless, it is the best we can do if $g \in H^\alpha$. But is the situation always so gloomy? One can immediately guess that a better rate of convergence can be achieved provided that g belongs to a subset of H^α; for example, if g is supersmooth itself. Let

$$\mathcal{SS}_{\alpha,\nu,\varrho}(A_\alpha) = \left\{g\colon \int_{-\infty}^{\infty} |\tilde{g}(\omega)|^2(\omega^2+1)^\alpha \exp\{2\varrho|\omega|^\nu\}\,dw \le A_\alpha\right\}, \tag{4.1}$$

$$\mathcal{SS}^*_{\alpha,\nu,\varrho}(A_\alpha, A_g) = \left\{g\colon\ g \in \mathcal{SS}_{\alpha,\nu,\varrho}(A_\alpha),\ \sup_\theta[|\theta|g(\theta)] \le A_g\right\} \tag{4.2}$$

and assume that $g \in \mathcal{SS}^*_{\alpha,\nu,\varrho}(A_\alpha, A_g)$ with positive ϱ and ν. Observe that $\mathcal{SS}_{\alpha,\nu,\varrho}(A_\alpha) \subseteq \mathcal{S}_\alpha(A_\alpha)$, $\mathcal{SS}^*_{\alpha,\nu,\varrho}(A_\alpha, A_g) \subseteq \mathcal{S}^*_\alpha(A_\alpha, A_g)$, and for $\varrho = 0$ the sets coincide: $\mathcal{SS}_{\alpha,\nu,\varrho}(A_\alpha) = \mathcal{S}_\alpha(A_\alpha)$, $\mathcal{SS}^*_{\alpha,\nu,\varrho}(A_\alpha, A_g) = \mathcal{S}^*_\alpha(A_\alpha, A_g)$.

The advantage of Meyer-type wavelet estimators is that they adjust automatically to the degree of smoothness of $g(\theta)$. It means that the estimators (2.11) and (2.12), with the same choice of parameters m, r and $\delta_{j,n}$ as before, achieve better convergence rates if $g(\theta)$ is supersmooth.

THEOREM 4. *If the conditions of Corollary* 1 *are valid, then the estimator* (2.11) *with m given by* (3.5) *attains the following convergence rate:*

$$\sup_{g\in\mathcal{SS}^*_{\alpha,\nu,\varrho}} \mathrm{MISE}\big(\hat{g}_n^{(LF)}\big) = \begin{cases} O\Big(n^{-1}(\ln n)^{(2\gamma+1)/\nu}\Big), & \textit{if } B = 0,\\ O\Big(n^{-\eta}\ln^{\xi} n\Big), & \textit{if } B > 0 \textit{ and } \nu \ge \beta,\\ O\Big((\ln n)^{-2\alpha/\beta}\exp\{-\zeta(\ln n)^{\nu/\beta}\}\Big), & \textit{if } B > 0 \textit{ and } \nu < \beta, \end{cases} \tag{4.3}$$

provided $\lim_{n\to\infty} K_n^{-1}n^\mu = 0$. *Here ν and ϱ are positive, $\mu = 1$ if $B > 0$ and $\mu = (2\gamma+1)^{-1}(2\gamma+2)$ if $B = 0$; $\zeta = [2B(4\pi/3)^\beta + \Lambda]^{-1}2\varrho(2\pi/3)^\nu$. If $\beta = \nu$, then $\xi = \beta^{-1}(2\gamma+1)I(\Lambda \ge 2\varrho(2\pi/3)^\beta) - 2\alpha\beta^{-1}I(\Lambda < 2\varrho(2\pi/3)^\beta)$ and $\eta = [2B(4\pi/3)^\beta + \Lambda]^{-1}\min(\Lambda, 2\varrho(2\pi/3)^\beta)$. If $\beta < \nu$, then $\xi = \beta^{-1}(2\gamma+1)$ and $\eta = [2B(4\pi/3)^\beta + \Lambda]^{-1}\Lambda$.*

THEOREM 5. *Let the assumptions of Theorem* 2 *hold and let L_n and M_n be such that*

$$\lim_{n\to\infty} nM_n^{-1} = 0, \qquad \lim_{n\to\infty} n^{(2\gamma+2)/(2\gamma+1)}L_n^{-1} = 0.$$

Then

$$\sup_{g\in \mathcal{SS}^*_{\alpha,\nu,\varrho}} \mathrm{MISE}\big(\hat{g}_n^{(NF)}\big) = O\Big((\ln n)^{\kappa(2\gamma+1)}\, n^{-1}\Big), \tag{4.4}$$

with $\kappa = \nu^{-1}$ *if* $\nu < 0.5$ *and* $\kappa = 2+\varepsilon$ *if* $\nu \geq 0.5$. *Here* ε *is an arbitrary constant.*

5. Examples. Let us now consider some examples of applications of Meyer-type wavelets to nonparametric density deconvolution.

EXAMPLE 1. Let $q(x) = (\sqrt{2\pi}\sigma)^{-1}\exp\{-0.5\, x^2\, \sigma^{-2}\}$ be the normal p.d.f. Then $\tilde{q}(\omega) = \exp\{-0.5\,\omega^2\sigma^2\}$ so that $\beta = 2$, $B = 0.5\sigma^2$, $\gamma = 0$. Since $q(x)$ is supersmooth, we use the linear wavelet estimator $\hat{g}_n^{(LF)}$. If $g \in \mathcal{S}^*_\alpha(A_\alpha, A_g)$, then $\hat{g}_n^{(LF)}$ has the optimal rate of convergence MISE $(\hat{g}_n^{(LF)}) = O((\ln n)^{-\alpha})$. Moreover, if $g \in \mathcal{SS}^*_{\alpha,\nu,\varrho}(A_\alpha, A_g)$ with $\varrho > 0$ and $\nu < \beta$, then MISE $(\hat{g}_n^{(LF)}) = O((\ln n)^{-2\alpha/\beta}\exp\{-\zeta\,(\ln n)^{\nu/\beta}\})$ with ζ given by Theorem 4. For example, if $g(\theta)$ is the Cauchy p.d.f. $g(\theta) = [\pi(\theta^2+1)]^{-1}$, then $\tilde{g}(\omega) = 0.5\exp\{-|\omega|\}$. Hence, applying Theorem 4 with $\nu = \varrho = 1$ and $\alpha = 0$, we obtain MISE $(\hat{g}_n^{(LF)}) = O((\ln n)^{-\alpha}\exp\{-\zeta\sqrt{\ln n}\})$. Note that in this case MISE $(\hat{g}_n^{(LF)})$ is $\exp\{-\zeta\sqrt{\ln n}\}$ times smaller than in the case of $g \in H^\alpha$. Here $\exp\{-\zeta\sqrt{\ln n}\} = o((\ln n)^{-\tau})$ for any positive τ as $n \to \infty$.

If $\nu = \beta$, then MISE $(\hat{g}_n^{(LF)}) = O(n^{-\eta}\ln^{\xi} n)$ with ξ and η given by Theorem 4. In particular, if $g(\theta)$ is also the p.d.f. of a normal distribution with variance σ_0^2, then formula (4.3) is valid with $\eta = [\Lambda + (4\pi\sigma/3)^2]^{-1}\min(\Lambda, (2\pi\sigma_0/3)^2)$; $\xi = 0.5$ if $\Lambda \geq \sigma_0^2(2\pi/3)^2$ and $\xi = 0$ otherwise. Note that in this case $\hat{g}_n^{(LF)}(\theta)$ has a polynomial rate of convergence, which is significantly better than the logarithmic rate of convergence in the case of $g \in \mathcal{S}^*_\alpha(A_\alpha, A_g)$.

EXAMPLE 2. Let $q(x) = 0.5\sigma\exp(-\sigma\,|x|)$, the p.d.f. of a double-exponential distribution. Then $\tilde{q}(\omega) = (1+\sigma^2\,\omega^2)^{-1}$, that is, $\gamma = 2$. Hence, Theorems 3–5 yield

$$\sup_{g\in \mathcal{SS}^*_{\alpha,\nu,\varrho}} \mathrm{MISE}\left(\hat{g}_n^{(LF)}\right) = \begin{cases} O\Big(n^{-2\alpha/(2\alpha+5)}\Big), & \text{if } \varrho = 0,\\ O\Big(n^{-1}(\ln n)^{5/\nu}\Big), & \text{if } \varrho > 0,\end{cases}$$

$$\sup_{g\in \mathcal{SS}^*_{\alpha,\nu,\varrho}} \mathrm{MISE}\left(\hat{g}_n^{(NF)}\right) = \begin{cases} O\Big(n^{-2\alpha/(2\alpha+5)}\Big), & \text{if } \varrho = 0,\\ O\Big(n^{-1}(\ln n)^{5\kappa}\Big), & \text{if } \varrho > 0,\end{cases}$$

where κ is defined in Theorem 5. Although the estimator $\hat{g}_n^{(LF)}(\theta)$ has better convergence rates than $\hat{g}_n^{(NF)}(\theta)$, it must be noted that the first estimator is constructed under the assumption that α, ϱ and ν are known, while the second estimator is adaptive and does not assume any knowledge of α, ϱ and ν.

For instance, if $g(\theta)$ is the p.d.f. of a normal distribution, then $\nu = 2$ and MISE $(\hat{g}_n^{(LF)}) = O(n^{-1}(\ln n)^{5/2})$, while MISE $(\hat{g}_n^{(NF)}) = O(n^{-1}(\ln n)^{10+\varepsilon})$. On the other hand, in the situation when $g(\theta)$ is the p.d.f. of a double-exponential distribution, MISE $(\hat{g}_n^{(LF)}) \sim$ MISE $(\hat{g}_n^{(NF)}) \sim O(n^{-2\alpha/(2\alpha+5)})$.

In what follows we conduct a numerical study of the construction of the deconvolution density when $q(x)$ is the p.d.f. of a double-exponential distribution

$$q(x) = 0.5\sigma \exp(-\sigma|x|), \qquad \sigma = 0.1. \tag{5.1}$$

We investigate the two cases when $g(\theta)$ is the standard normal p.d.f. and $g(\theta)$ is the standard double-exponential p.d.f. $g(\theta) = 0.5 \exp(-|\theta|)$. We compare the Meyer-type wavelet estimators $\hat{g}_n^{(NF)}(\theta)$ of the form (2.12) with the kernel deconvolution estimators based on the kernel $K(x)$,

$$\hat{g}_n^{(K)}(\theta) = \frac{1}{nh}\sum_{l=1}^{n} L\left(\frac{\theta - X_l}{h}, h\right) \quad \text{with } \tilde{L}(\omega, h) = [\tilde{q}(h^{-1}\omega)]^{-1}\, \tilde{K}(\omega). \tag{5.2}$$

Here $L(\cdot, h)$ is the inverse Fourier transform of $\tilde{L}(\omega, h)$. We consider two types of kernels, $K_1(x) = \varphi(x)$ where $\varphi(x)$ is the Meyer-type scale function and $K_2(x) = (\sqrt{2\pi})^{-1} \exp(-0.5\, x^2)$. We denote the kernel deconvolution estimators based on kernels K_1 and K_2 by $\hat{g}_n^{(K_1)}(\theta)$ and $\hat{g}_n^{(K_2)}(\theta)$, respectively. It is easy to show that for $q(x)$ given by (5.1),

$$\begin{aligned} U_m(x) &= \varphi(x) + 2^{2m}\sigma^2\varphi''(x), \qquad V_j(x) = \psi(x) + 2^{2j}\sigma^2\psi''(x), \\ L_1(x, h) &= \varphi(x) + h^{-2}\sigma^2\varphi''(x), \\ L_2(x) &= (\sqrt{2\pi})^{-1}\exp(-0.5x^2)[1 + h^{-2}\sigma^2(x^2 - 1)]. \end{aligned} \tag{5.3}$$

Therefore, the coefficients $\hat{a}_{m,k}$ and $\hat{b}_{j,k}$ of the wavelet estimator $\hat{g}_n^{(NF)}(\theta)$ are calculated according to (2.8) with U_m and V_j given by (5.3). For practical purposes, we use the approximations $\varphi(x) =$ MeyerPhi[s, x, 20] and $\psi(x) =$ MeyerPsi[s, x, 20] included with the Mathematica Wavelet Package with $s = 2$. All simulations are conducted with $n = 500$.

Figure 1a–c presents the simulation study when $g(\theta)$ is the p.d.f. of the standard normal distribution. The estimators $\hat{g}_n^{(NF)}(\theta)$, $\hat{g}_n^{(K_1)}(\theta)$ and $\hat{g}_n^{(K_2)}(\theta)$ are based on the same sample of $n = 500$ observations. Panel (a) shows $\hat{g}_n^{(NF)}(\theta)$ with $m = -1$, $r = 0$, $M_n = L_n = 10$ and $\delta_{-1,n} = 0.003$. Observe that the sum $\sum_{j=m}^{m+r}$ contains one term with $j = -1$, so that the estimator is nonlinear. Panels (b) and (c) depict $\hat{g}_n^{(K_1)}(\theta)$ and $\hat{g}_n^{(K_2)}(\theta)$, respectively, with $h = 0.5$. In all three panels, the exact density $g(\theta)$ is shown in dashed lines.

Figure 2a–c presents an investigation, analagous to that in Figure 1a–c, when $g(\theta)$ is the p.d.f. of the standard double-exponential distribution. Since the p.d.f. $g(\theta)$ is "unknown" we conduct estimation with the same values of the parameters m, r, M_n, L_n $\delta_{-1,n}$ and h as in the case of the standard normal $g(\theta)$ and with the same sample size $n = 500$.

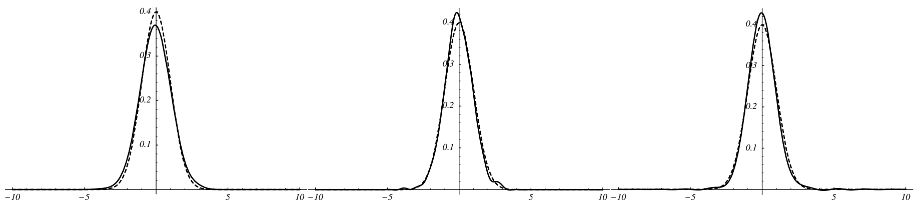

FIG. 1. *Deconvolutions with normal prior and double-exponential conditional distribution.* (a) *Estimation with Meyer-type wavelets.* (b) *Estimation with Meyer-type kernels.* (c) *Estimation with the standard normal kernels.*

6. Discussion. In this paper we constructed the linear and the nonlinear estimators of a deconvolution density $g(\theta)$ based on Meyer-type wavelets. We showed that the estimators are asymptotically optimal for $g \in H^\alpha$. Moreover, we demonstrated that, for $B > 0$, the linear wavelet estimator, and, for $B = 0$, the nonlinear wavelet estimator are globally adaptive. That is, the choice of parameters is independent of the unknown parameter α.

Another merit of the estimators (2.11) and (2.12) is that they can adjust even to a supersmooth deconvolution density. It is easy to see that the estimator (2.11) provides a better convergence rate if $g(\theta)$ is supersmooth. Namely, if $q(x)$ is also supersmooth and $\nu < \beta$, then $\sup_{g \in \mathscr{S}^*_{\alpha,\nu,\varrho}} \mathrm{MISE}\,(\hat{g}_n^{(LF)}) = o((\ln n)^{-\tau})$ for any positive τ as $n \to \infty$. If $\nu \geq \beta$, then $\hat{g}_n^{(LF)}(\theta)$ has a polynomial rate of convergence [see (4.3)]. The rate of convergence is governed by the parameter Λ. A large value of Λ ensures that the estimator has a high convergence rate when $g(\theta)$ is supersmooth ($\varrho > 0$) without affecting the convergence rate when $g(\theta)$ has a finite degree of smoothness ($\varrho = 0$). However, increasing Λ immediately leads to the increase of a constant in front of $(\ln n)^{-2\alpha/\beta}$ in (3.9). Therefore, there is an obvious trade-off between one's desire to accommodate the case of supersmooth $g(\theta)$ and reluctance to slow down the convergence provided $g(\theta)$ is not supersmooth.

In the case $B = 0$, the linear wavelet estimator (2.11) has a convergence rate close to $O(n^{-1})$ when the values of α, ϱ and ν are known. If (and it is usually the case) they are unknown, the nonlinear wavelet estimator $\hat{g}_n^{(NF)}(\theta)$ attains

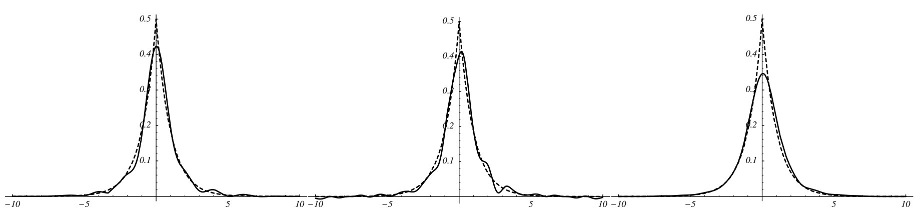

FIG. 2. *Deconvolutions with a double-exponential prior and double-exponential conditional distribution.* (a) *Estimation with Meyer-type wavelets.* (b) *Estimation with Meyer-type kernels.* (c) *Estimation with the standard normal kernels.*

a convergence rate which either coincides with the convergence rate of the linear estimator (if $\nu < 0.5$) or is $(\ln n)^{(2\gamma+1)(2+\varepsilon-\nu^{-1})}$ times greater (if $\nu \geq 0.5$). It should be noted that an estimator based on a wavelet with bounded support fails to provide a convergence rate better than $O(n^{-2s/(2s+1)})$, where s is the degree of regularity of the wavelet. The fact that Meyer-type wavelets work better in the situation of a supersmooth $g(\theta)$ seems completely natural: Meyer-type wavelets are supersmooth and, consequently, are suitable to estimate a supersmooth density function.

Section 5 provides finite sample size simulations study of the performance of the Meyer-type wavelet estimators versus the kernel deconvolution estimators. In the case when $g(\theta)$ is the standard normal p.d.f., the Meyer-type wavelet estimator $\hat{g}_n^{(NF)}(\theta)$ is more precise than $\hat{g}_n^{(K_1)}(\theta)$ and is as precise as $\hat{g}_n^{(K_2)}(\theta)$ (see Figure 1a–c). In the situation where $g(\theta) = 0.5\exp(-|\theta|)$, $\hat{g}_n^{(NF)}(\theta)$ is closer to $g(\theta)$ than its kernel deconvolution counterparts (see Figures 2a–c). However, the simulation sudy conducted in the present paper is very limited. The detailed simulation study of the performance of the method is the authors' ongoing research project.

Let us also make a remark about the use of Meyer-type wavelet estimators for the estimation of $p(x)$. Without loss of generality, let us consider the linear wavelet estimator $\hat{g}_n^{(L)}(\theta)$. It is easy to see that the estimator $\hat{p}_n^{(L)}(x) = \int_{-\infty}^{\infty} q(x-\theta)\hat{g}_n^{(L)}(\theta)\, d\theta$ can be written as

$$\hat{p}_n^{(L)}(x) = \sum_{k\in\mathbf{Z}} \hat{a}_{m,k}\Phi_{m,k}(x), \tag{6.1}$$

where the coefficients $\hat{a}_{m,k}$ are given by (2.8) and $\Phi_{m,k}(x) = 2^{m/2}\Phi_m^*(2^m x - k)$ with $\Phi_m^*(y) = \int_{-\infty}^{\infty} q(2^{-m}y - z)\varphi(2^m z)\, dz$. The estimator (6.1) does not have desirable properties: the variances of the coefficients $\hat{a}_{m,k}$ are high since $\hat{a}_{m,k}$ are based on a "deconvolved" sample while the functions $2^{m/2}\Phi_m^*(2^m x - k)$ are not orthonormal unlike $2^{m/2}\varphi(2^m x - k)$. Also, Φ_m^* depends on m. Therefore, $\hat{p}_n^{(L)}(x)$ cannot be recommended as the estimator of $p(x)$.

On the other hand, the conditional expectation $\hat{g}_n^* = \mathbf{E}(\hat{g}_n^{(L)}|\theta_1, \theta_2, \dots, \theta_n)$ is an adequate approximation of $g(\theta)$. It is easy to show that

$$\hat{g}_n^*(\theta) = \sum_{k\in\mathbf{Z}} \hat{a}_{m,k}^*\varphi_{m,k}(\theta) = n^{-1}\sum_{l=1}^{n}\sum_{k\in\mathbf{Z}} \varphi_{m,k}(\theta_l)\varphi_{m,k}(\theta), \tag{6.2}$$

that is, $\hat{g}_n^*(\theta)$ is a linear wavelet estimator of $g(\theta)$ based on $\theta_1, \theta_2, \dots, \theta_n \sim g(\theta)$. The properties of the estimator (6.2) were studied by several authors [see, e.g., Walter (1994), Pensky (1999)]. It is easy to show that

$$\left|\mathbf{E}\hat{g}_n^*(\theta) - g(\theta)\right| = O(2^{-m\alpha}), \quad \text{Var } \hat{g}_n^*(\theta) = O(n^{-1}2^m).$$

Therefore, the precision of (6.2) depends on the choice of m. If m has the form (3.5), then $\hat{g}_n^*(\theta)$ has the same convergence rate as $\hat{g}_n^{(L)}(\theta)$. However, if m is such that $2^m \sim n^{1/(2\alpha+1)}$, then MISE $(\hat{g}_n^*)$ has the optimal rate of convergence in H^α: MISE $(\hat{g}_n^*) = O(n^{-2\alpha/(2\alpha+1)})$.

APPENDIX

The proofs of all major statements are based on several auxiliary lemmas.

LEMMA 1 [Talagrand (1994)]. *Let $X_1, \dots, X_n$ be independent and identically distributed random variables, let $\varepsilon_1, \dots, \varepsilon_n$ be independent Rademacher variables, also independent of $X_1, \dots, X_n$ and let $\mathcal{F}$ be a class of functions uniformly bounded by T. If*

$$\sup_{f\in\mathcal{F}} \operatorname{Var} f(X) \le v \qquad \textit{and} \qquad \mathbf{E}\{\sup_{f\in\mathcal{F}} \sum_{l=1}^{n} \varepsilon_l f(X_l)\} \le nH,$$

then there exist universal constants K_1, K_2 such that for $\nu_n(f) = n^{-1}\sum_{l=1}^{n} f(X_l) - \mathbf{E} f(X)$ and any $t > 0$ we have

$$P\left\{\sup_{f\in\mathcal{F}} \nu_n(f) \ge t + K_2 H\right\} \le \exp\left\{-nK_1 \min\left(\frac{t^2}{v}, \frac{t}{T}\right)\right\}. \tag{A.1}$$

LEMMA 2. *If $\varrho \ge 0$ and $\nu \ge 0$, then*

$$\sup_{g\in\mathcal{SS}_{\alpha,\nu,\varrho}} \sum_{k\in\mathbf{Z}} b_{j,k}^2 \le C_b 2^{-2j\alpha} \exp\{-2\varrho(2\pi/3)^{\nu} 2^{j\nu}\}\,\varepsilon_j,$$

with $C_b = \pi^{-1}\|\psi\|_c^2\,(2\pi/3)^{-2\alpha}$ and $\sum_{j\in\mathbf{Z}} \varepsilon_j \le 2A_\alpha^2$. Here $\mathcal{SS}_{\alpha,\nu,\varrho}(A_\alpha)$ is defined in (4.1).

PROOF. The proof of Lemma 2 is similar to the proof of Theorem 3.4 by Zayed and Walter (1996). We represent the coefficients $b_{j,k}$ as $b_{j,k} = b_{j,k}^{(1)} + b_{j,k}^{(2)}$, where

$$b_{j,k}^{(l)} = (2\pi)^{-1} 2^{j/2} \int_{\Omega_l} \exp(ik\omega)\, \tilde{g}(2^j\omega)\tilde{\psi}(\omega)\, d\omega, \qquad l = 1, 2,$$

with Ω_1 and Ω_2 defined in (2.4). It is easy to notice that $b_{j,k}^{(l)}$ are the Fourier coefficients of the functions $(2\pi)^{-1}2^{j/2}\tilde{g}(2^j\omega)\tilde{\psi}(\omega)I(\omega \in \Omega_l)$, $l = 1, 2$, so that $\sum_{k\in\mathbf{Z}} |b_{j,k}^{(l)}|^2 = (2\pi)^{-1}2^j \int_{\Omega_l} |\tilde{g}(2^j\omega)|^2 |\tilde{\psi}(\omega)|^2 d\omega$. Since $\sum_{k\in\mathbf{Z}} |b_{j,k}|^2 \le 2\sum_{k\in\mathbf{Z}} |b_{j,k}^{(1)}|^2 + 2\sum_{k\in\mathbf{Z}} |b_{j,k}^{(2)}|^2$, we conclude that

$$\sum_{k\in\mathbf{Z}} |b_{j,k}|^2 \le \pi^{-1}\, \|\tilde{\psi}\|_c^2\, (2\pi/3)^{-2\alpha}\; 2^{-2j\alpha} \exp\{-2\varrho(2\pi/3)^{\nu} 2^{j\nu}\}\, \varepsilon_j.$$

Here $\varepsilon_j = \int_{W_j} |\tilde{g}(\omega)|^2(\omega^2+1)^{\alpha} \exp\{2\varrho|\omega|^{\nu}\}\, d\omega$ with $W_j = [-2^j 8\pi/3, -2^j 2\pi/3] \cup [2^j 2\pi/3, 2^j 8\pi/3]$ which implies $\sum_j \varepsilon_j \le 2A_\alpha^2 < \infty$.

LEMMA 3. *For $\Delta_1(m)$ and $\Delta_2(j)$ defined in* (3.3) *the following inequalities are valid:*

$$\sup_x \sum_{k\in\mathbf{Z}} \left|V_j(2^j x - k)\right|^2 \le 2\Delta_2(j), \tag{A.2}$$

$$\sup_{g\in\mathcal{S}_\alpha} \left[\sum_{k\in\mathbf{Z}} \operatorname{Var}(\hat{b}_{j,k})\right] \le n^{-1} 2^{j+1} \Delta_2(j), \tag{A.3}$$

$$\sup_{g \in \mathscr{S}_\alpha} \left[\sum_{k \in \mathbf{Z}} \operatorname{Var}(\hat{a}_{m,k}) \right] \le n^{-1} 2^{m+1} \Delta_1(m). \tag{A.4}$$

PROOF. Observe that $V_j(2^j x - k) = \beta_k^{(1)}(2^j x) + \beta_k^{(2)}(2^j x)$ where the functions $\beta_k^{(l)}(2^j x) = (2\pi)^{-1} \int_{\Omega_l} \exp(-ik\omega) \exp(i 2^j x \omega) \tilde{V}_j(\omega)\, d\omega$, $l = 1, 2$, are the Fourier coefficients of the functions $\exp(i2^j x\omega)\tilde{V}_j(\omega) I(\omega \in \Omega_l) = \exp(i 2^j x \omega) \cdot \tilde{\psi}(\omega)[\tilde{q}(-2^j\omega)]^{-1} I(\omega \in \Omega_l)$, $l = 1, 2$. Therefore,

$$\sum_{k \in \mathbf{Z}} \left| V_j(2^j x - k) \right|^2 \le 2 \sum_{l=1}^{2} \sum_{k \in \mathbf{Z}} \left| \beta_k^{(l)}(2^j x) \right|^2 \le 2\Delta_2(j),$$

which implies (A.2). To prove (A.3) notice that $\sum_{k \in \mathbf{Z}} \operatorname{Var}(\hat{b}_{j,k}) \le n^{-1} \int_{-\infty}^{\infty} 2^j \cdot \sum_{k \in \mathbf{Z}} |V_j(2^j x - k)|^2 p(x)\, dx \le n^{-1} 2^j \sup_x \{\sum_{k \in \mathbf{Z}} |V_j(2^j x - k)|^2\}$. Inequality (A.4) can be derived in a manner similar to (A.3).

LEMMA 4. *Let* $\varrho_{k,l}(j) = \int_{-\infty}^{\infty} 2^j V_j(2^j x - k) V_j(2^j x - l) p(x)\, dx$. *Then*

$$\sup_{g \in \mathscr{S}_\alpha} \left[\sum_{k, l \in \mathbf{Z}} \left| \varrho_{k,l}(j) \right|^2 \right] \le C_\varrho 2^j \Delta_4(j),$$

where $\mathscr{S}_\alpha(A_\alpha)$ *and* $\Delta_4(j)$ *are defined in* (3.1) *and* (3.3), *respectively, and* $C_\varrho = 2\pi A_\alpha^2 \|q\|_{L^2}^2$.

PROOF. To simplify notation, we drop the argument j of $\varrho_{k,l}(j)$ in the proof and refer to $\varrho_{k,l}(j)$ as $\varrho_{k,l}$. Using Parseval's identity and properties of Fourier transform, we write $\varrho_{k,l}$ as

$$\varrho_{k,l} = \frac{2^j}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-izk - iyl) \frac{\tilde{\psi}(z)\tilde{\psi}(y)}{\tilde{q}(-2^j z)\tilde{q}(-2^j y)} \times \tilde{p}(-2^j(y+z))\, dy\, dz. \tag{A.5}$$

Therefore $\varrho_{k,l} = \sum_{s=1}^{2} \sum_{r=1}^{2} \varrho_{k,l}^{(s,r)}$, where the only difference between $\varrho_{k,l}^{(s,r)}$ and $\varrho_{k,l}$ is that the integral for $\varrho_{k,l}^{(s,r)}$ is calculated over $\Omega_s \cup \Omega_r$, $s, r = 1, 2$ [see (2.4)]. It is easy to see that $\varrho_{k,l}^{(s,r)}$ are the Fourier coefficients of the functions

$$[\tilde{q}(-2^j z)\tilde{q}(-2^j y)]^{-1} \tilde{\psi}(z)\tilde{\psi}(y)\tilde{p}(-2^j(y+z)) I(y \in \Omega_s) I(z \in \Omega_r), \qquad r, s = 1, 2,$$

respectively, which implies that $\sum_{k, l \in \mathbf{Z}} |\varrho_{k,l}^{(s,r)}|^2$ is equal to the integral of the square of the magnitude of the function generating the coefficients $\varrho_{k,l}^{(s,r)}$. To complete the proof, note that the function $|\tilde{q}(\omega)|$ is even and that $\|\tilde{p}\|_{L^2} \le A_\alpha \|\tilde{q}\|_{L^2}$.

LEMMA 5. *Denote* $\lambda_{j,n} = \lambda_n \sqrt{2^j \Delta_2(j)}$. *Then for any* $\lambda_n > 0$,

$$\begin{aligned}\sup_{g\in\mathscr{S}_\alpha} P\left\{\sum_{k\in\mathbf{Z}} (\hat{b}_{j,k} - b_{j,k})^2 \ge \lambda_{j,n}^2 \left[1 + \frac{\sqrt{2}K_2}{\sqrt{n}\lambda_n}\right]^2\right\} \\ \le \exp\left\{-nK_1\left(\frac{2^{j/2}\Delta_2(j)\lambda_n^2}{\sqrt{C_\varrho\,\Delta_4(j)}} \wedge \frac{\lambda_n}{\sqrt{2}}\right)\right\}.\end{aligned} \tag{A.6}$$

Here K_1 and K_2 are the absolute constants in Talagrand's inequality, $a \wedge b = \min(a, b)$ and C_ϱ is defined in Lemma 4.

PROOF OF LEMMA 5. The proof is based on Talagrand's inequality (Lemma 1). Consider an infinite sequence $\mathscr{E} = \{(e_k)_{k\in\mathbf{Z}}: \sum_{k\in\mathbf{Z}} e_k^2 \le 1\}$ and a class of functions $\mathscr{F} = \{f: f(x) = 2^{j/2}\sum_{k\in\mathbf{Z}} e_k V_j(2^j x - k)\}$. It is easy to notice that

$$\sum_{k\in\mathbf{Z}} (\hat{b}_{j,k} - b_{j,k})^2 = \left[\sup_{e\in\mathscr{E}} \sum_{k\in\mathbf{Z}} (\hat{b}_{j,k} - b_{j,k}) e_k\right]^2 = \sup_{f\in\mathscr{F}} [\nu_n(f)]^2.$$

Now to complete the proof we need to find v, H and T in Talagrand's inequality. From Lemma 3 it follows that

$$|f(x)| \le \left(\sum_{k\in\mathbf{Z}} e_k^2\right)^{1/2} \left(\sum_{k\in\mathbf{Z}} 2^j |V_j(2^j x - k)|^2\right)^{1/2} \le 2^{(j+1)/2}\sqrt{\Delta_2(j)},$$

that is, $T = 2^{(j+1)/2}\sqrt{\Delta_2(j)}$. Also, if $\varepsilon_1, \ldots, \varepsilon_n$ are independent Rademacher variables, then

$$\begin{aligned}\mathbf{E}\left[\sup_{e\in\mathscr{E}} \sum_{l=1}^n \left\{\sum_{k\in\mathbf{Z}} e_k 2^{j/2} V_j(2^j X_l - k)\varepsilon_l\right\}\right] &\le \left[\sum_{k\in\mathbf{Z}} \mathbf{E}\left\{\sum_{l=1}^n 2^{j/2} V_j(2^j X_l - k)\varepsilon_l\right\}^2\right]^{1/2} \\ &\le \left[\sum_{k\in\mathbf{Z}} \mathbf{E}\left\{\sum_{l=1}^n 2^j V_j^2(2^j X_l - k)\right\}\right]^{1/2} \\ &\le 2^{(j+1)/2}\sqrt{n\,\Delta_2(j)},\end{aligned}$$

which implies that $H = n^{-1/2}\,2^{(j+1)/2}\sqrt{\Delta_2(j)}$. To obtain an upper bound for v we apply Lemma 4: $\sup_{f\in\mathscr{F}} \operatorname{Var} f(X) \le \sup_{e\in\mathscr{E}} \sum_{k,l} e_k e_l \varrho_{k,l} \le \left[\sum_{k,l} |\varrho_{k,l}|^2\right]^{1/2} \cdot [\sum_{k\in\mathbf{Z}} e_k^2]$, that is, $v \le 2^{j/2}\sqrt{C_\varrho\,\Delta_4(j)}$. To complete the proof, rewrite (A.1) with $t = \lambda_{j,n}$ and take the supremum of both sides over $\mathscr{S}_\alpha(A_\alpha)$. □

PROOF OF THEOREM 1. The validity of Theorem 1 follows directly from Lemma 2 with $\varrho = 0$, Lemma 3 and the inequality

$$\text{MISE}\ (\hat{g}_n^{(L)}) \le \sum_{j=m}^{\infty} \sum_{k\in\mathbf{Z}} b_{j,k}^2 + \sum_{k\in\mathbf{Z}} \operatorname{Var}\ \hat{a}_{m,k}. \qquad \square$$

PROOF OF COROLLARY 1. Formulas (3.5) and (3.6) follow directly from Theorem 1 and the asymptotic equality $\Delta_1(m) \sim 2^{2\gamma m} \exp\{2B(4\pi/3)^\beta 2^{m\beta}\}$. □

PROOF OF THEOREM 2. Note that MISE $(\hat{g}_n^{(N)})$ can be written as the sum of four components,

$$\text{(A.7)}\qquad \begin{aligned} \text{MISE}\left(\hat{g}_n^{(N)}\right) &= R_1 + R_2 + R_3 + R_4 \\ &\equiv \sum_{j=m}^{m+r} \mathbf{E}\left\{\sum_{k\in\mathbf{Z}} (\hat{b}_{j,k} - b_{j,k})^2\, I\left(\sum_{k\in\mathbf{Z}} \hat{b}_{j,k}^2 \geq \delta_{j,n}^2\right)\right\} \\ &\quad + \sum_{j=m}^{m+r} \left(\sum_{k\in\mathbf{Z}} b_{j,k}^2\right) P\left(\sum_{k\in\mathbf{Z}} \hat{b}_{j,k}^2 \leq \delta_{j,n}^2\right) + \sum_{j=m+r+1}^{\infty} \sum_{k\in\mathbf{Z}} b_{j,k}^2 \\ &\quad + \sum_{k\in\mathbf{Z}} \mathbf{E}(\hat{a}_{m,k} - a_{m,k})^2, \end{aligned}$$

so that

$$\text{(A.8)}\qquad \sup_{g\in\mathscr{S}_\alpha} \text{MISE}\left(\hat{g}_n^{(N)}\right) \leq \sup_{g\in\mathscr{S}_\alpha} R_1 + \sup_{g\in\mathscr{S}_\alpha} R_2 + \sup_{g\in\mathscr{S}_\alpha} R_3 + \sup_{g\in\mathscr{S}_\alpha} R_4.$$

Let us analyze each term R_i, $i = 1, 2, 3, 4$ in turn.

The upper bound for $\sup_{g\in\mathscr{S}_\alpha} R_1$. Observe that

$$I\left(\sum_{k\in\mathbf{Z}} \hat{b}_{j,k}^2 > \delta_{j,n}^2\right) \leq I\left(\sum_{k\in\mathbf{Z}} (\hat{b}_{j,k} - b_{j,k})^2 > 0.25\,\delta_{j,n}^2\right) + I\left(\sum_{k\in\mathbf{Z}} b_{j,k}^2 > 0.25\,\delta_{j,n}^2\right),$$

so that $R_1 \leq R_{1,1} + R_{1,2}$ where

$$\text{(A.9)}\qquad R_{1,1} = \sum_{j=m}^{m+r} \mathbf{E}\left\{\sum_{k\in\mathbf{Z}} (\hat{b}_{j,k} - b_{j,k})^2\, I\left(\sum_{k\in\mathbf{Z}} (\hat{b}_{j,k} - b_{j,k})^2 > 0.25\,\delta_{j,n}^2\right)\right\},$$

$$\text{(A.10)}\qquad R_{1,2} = \sum_{j=m}^{m+r} \mathbf{E}\left\{\sum_{k\in\mathbf{Z}} (\hat{b}_{j,k} - b_{j,k})^2\right\} I\left(\sum_{k\in\mathbf{Z}} b_{j,k}^2 > 0.25\,\delta_{j,n}^2\right).$$

The first term, $R_{1,1}$, is dominated by

$$\text{(A.11)}\qquad \begin{aligned} R_{1,1} \leq \textstyle\sum_{j=m}^{m+r} \Bigg[&\mathbf{E}\left\{\textstyle\sum_{k\in\mathbf{Z}} (\hat{b}_{j,k} - b_{j,k})^2\right\}^2 \\ &\times P\left(\textstyle\sum_{k\in\mathbf{Z}} (\hat{b}_{j,k} - b_{j,k})^2 > 0.25\,\delta_{j,n}^2\right)\Bigg]^{1/2}, \end{aligned}$$

the sum of the products of two factors. The first factors in (A.11) are majorized by

$$\begin{aligned}
&\mathbf{E}\left\{\sum_{k\in\mathbf{Z}}(\hat{b}_{j,k}-b_{j,k})^2\right\}^2\\
&\quad\le \mathbf{E}\left\{\sum_{k\in\mathbf{Z}}(\hat{b}_{j,k}-b_{j,k})^2\sum_{k\in\mathbf{Z}}\left[n^{-1}\sum_{l=1}^{n}2^{j/2}V_j(2^jX_l-k)-b_{j,k}\right]^2\right\} \qquad \text{(A.12)}\\
&\quad\le \left\{2^{j+1}\sup_x\left[\sum_{k\in\mathbf{Z}}\left|V_je(2^jx-k)\right|^2\right]+2\sum_{k\in\mathbf{Z}}b_{j,k}^2\right\}\sum_{k\in\mathbf{Z}}\operatorname{Var}\hat{b}_{j,k}\\
&\quad= O\big(n^{-1}\,2^{2j+4}[\Delta_2(j)]^2\big),
\end{aligned}$$

according to Lemmas 2 and 3. Construction of the upper bounds for the second factors in (A.11) is based on Lemma 5. Choose $\lambda_n = \lambda_0\,\delta_n$ with $\lambda_0 \le 0.5\delta_0 - \sqrt{2}K_2$ and note that in this case $0.25\,\delta_0^2\,\lambda_0^{-2} \ge (1+\sqrt{2}\lambda_0^{-1}K_2)^2$. Therefore, $0.25\,\delta_{j,n}^2 = 0.25\,\delta_0^2\,n^{-1}\,2^j\,\Delta_2(j) \ge \lambda_{j,n}^2(1+\lambda_n^{-1}n^{-1/2}\sqrt{2}K_2)^2$. Hence, from Lemma 5 it follows that

$$P\left(\sum_{k\in\mathbf{Z}}(\hat{b}_{j,k}-b_{j,k})^2 > 0.25\,\delta_{j,n}^2\right) \le \exp\left[-(C_1 2^{j/2}\wedge C_2\sqrt{n})\right] \qquad \text{(A.13)}$$

with $C_1 = K_1\,\lambda_0^2\,C_0^{-1}\,C_\varrho^{-1/2}$ and $C_2 = \lambda_0 K_1/\sqrt{2}$, since $\Delta_2^2(j)\Delta_4^{-1}(j) \ge C_0^{-1}$. Combining (A.11)–(A.13), taking into account that $2^{j/2} > (\ln n)^{1+0.5\,\varepsilon}$ and taking supremum over $\mathscr{S}_\alpha(A_\alpha)$, we obtain

$$\sup_{g\in\mathscr{S}_\alpha} R_{1,1} = o\big(n^{-1}\big). \qquad \text{(A.14)}$$

Now let us calculate $R_{1,2}$ [see (A.10)]. Since $\Delta_2(j) \le C_\Delta 2^{2\gamma j}$ with $C_\Delta = (A_0)^2[(8\pi/3)^2+1]$, according to Lemmas 2 and 3,

$$R_{1,2} \le n^{-1}\sum_{j=m}^{m+r} 2^{j+1}C_\Delta 2^{2\gamma j}\; I\Big(C_b\varepsilon_j 2^{-2\alpha j} > 0.25\,\delta_0^2\,C_\Delta\,n^{-1}\,2^{j(2\gamma+1)}\Big).$$

Rearranging the last formula, we have

$$R_{1,2} \le 2C_\Delta\delta_0^{-(2(2\gamma+1))/(2\alpha+2\gamma+1)}\left(\frac{4C_b\varepsilon_j}{C_\Delta}\right)^{(2\gamma+1)/(2\alpha+2\gamma+1)} n^{-2\alpha/(2\alpha+2\gamma+1)},$$

which implies that

$$\sup_{g\in\mathscr{S}_\alpha} R_{1,2} = O\Big(n^{-2\alpha/(2\alpha+2\gamma+1)}\Big), \qquad \text{(A.15)}$$

since $\sum_{j\in\mathbf{Z}}\varepsilon_j \le 2\,A_\alpha^2$. Combining (A.14) and (A.15), we obtain $\sup_{g\in\mathscr{S}_\alpha} R_1 = O(n^{-2\alpha/(2\alpha+2\gamma+1)})$, as $n\to\infty$.

The upper bound for $\sup_{g\in\mathscr{S}_\alpha} R_2$. To find the upper bound for R_2 we introduce $M=(2\alpha+2\gamma+1)^{-1}\log_2 n$ and note that

$$I\Bigl(\sum_{k\in\mathbf{Z}}\hat b_{j,k}^2\le\delta_{j,n}^2\Bigr)\le I\Bigl(\sum_{k\in\mathbf{Z}}b_{j,k}^2\le 2.5\,\delta_{j,n}^2\Bigr)+I\Bigl(\sum_{k\in\mathbf{Z}}(\hat b_{j,k}-b_{j,k})^2>0.25\,\delta_{j,n}^2\Bigr)$$

by virtue of the inequality $\hat b_{j,k}^2\ge 0.5\,b_{j,k}^2-(\hat b_{j,k}-b_{j,k})^2$. Thus,

$$R_2\le R_{2,1}+R_{2,2}+R_{2,3}, \tag{A.16}$$

where

$$R_{2,1}=\sum_{j=M+1}^{m+r}\Bigl(\sum_{k\in\mathbf{Z}}b_{j,k}^2\Bigr)P\Bigl(\sum_{k\in\mathbf{Z}}\hat b_{j,k}^2\le\delta_{j,n}^2\Bigr),$$

$$R_{2,2}=\sum_{j=m}^{M}\Bigl(\sum_{k\in\mathbf{Z}}b_{j,k}^2\Bigr)I\Bigl(\sum_{k\in\mathbf{Z}}b_{j,k}^2\le 2.5\,\delta_{j,n}^2\Bigr),$$

$$R_{2,3}=\sum_{j=m}^{M}\Bigl(\sum_{k\in\mathbf{Z}}b_{j,k}^2\Bigr)P\Bigl(\sum_{k\in\mathbf{Z}}(\hat b_{j,k}-b_{j,k})^2>0.25\,\delta_{j,n}^2\Bigr).$$

Now, from the choice of M and Lemma 2 with $\varrho=0$, it follows that

$$\sup_{g\in\mathscr{S}_\alpha}R_{2,1}\le 2C_bA_\alpha^2\sum_{j=M+1}^{m+r}2^{-2\alpha j}=O\Bigl(n^{-2\alpha/(2\alpha+2\gamma+1)}\Bigr).$$

Also, using Lemma 3 we derive that

$$R_{2,2}=\sum_{j=m}^{M}\Bigl(\sum_{k\in\mathbf{Z}}b_{j,k}^2\Bigr)I\Bigl(\sum_{k\in\mathbf{Z}}b_{j,k}^2\le 2.5\,C_\Delta\,\delta_0^2\,2^{(2\gamma+1)j}\,n^{-1}\Bigr),$$

which implies that

$$\sup_{g\in\mathscr{S}_\alpha}R_{2,2}\le 2.5\,C_\Delta\,\delta_0^2\,n^{-1}\,2^{M(2\gamma+1)}=O\Bigl(n^{-2\alpha/(2\alpha+2\gamma+1)}\Bigr).$$

The last term, $R_{2,3}$, can be majorized by applying Lemma 2 and formula (A.13),

$$\sup_{g\in\mathscr{S}_\alpha}R_{2,3}\le 2C_ba_\alpha^2\sum_{j=m}^{M}\Bigl[2^{-2\alpha j}\exp\Bigl\{-(C_12^{j/2}\wedge C_2\sqrt{n})\Bigr\}\Bigr],$$

and therefore $\sup_{g\in\mathscr{S}_\alpha}R_{2,3}=o(n^{-1})$. Combining all three components in (A.16), we conclude that $\sup_{g\in\mathscr{S}_\alpha}R_2=O(n^{-2\alpha/(2\alpha+2\gamma+1)})$, as $n\to\infty$.

The upper bounds for $\sup_{g\in\mathscr{S}_\alpha} R_3$ *and* $\sup_{g\in\mathscr{S}_\alpha} R_4$. The upper bounds for $\sup_{g\in\mathscr{S}_\alpha} R_3$ and $\sup_{g\in\mathscr{S}_\alpha} R_4$ follow directly from Lemma 2 with $\varrho = 0$ and Lemma 3, respectively,

$$\sup_{g\in\mathscr{S}_\alpha} R_3 \le \sum_{j=m+r}^{\infty} 2C_b A_\alpha^2 2^{-2\alpha j} = o(n^{-1}),$$

$$\sup_{g\in\mathscr{S}_\alpha} R_4 \le 2^{m+1}\Delta_1(m) n^{-1} = o\Big(n^{-2\alpha/(2\alpha+2\gamma+1)}\Big),$$

which completes the proof of the theorem. □

PROOF OF THEOREM 3. Assume that $g \in \mathscr{S}_\alpha^*(A_\alpha, A_g)$ where the class $\mathscr{S}_\alpha^*(A_\alpha, A_g)$ is defined in (3.8). Then the coefficients $a_{m,k}$ satisfy the chain of inequalities

$$\begin{aligned}
k|a_{m,k}| &\le 2^{m/2}\int_{-\infty}^{\infty}\big|(2^m\theta - k) - 2^m\theta\big|\,\big|\varphi(2^m\theta-k)\big| g(\theta)\,d\theta\\
&\le 2^{m/2}\sup_z[|z|\,|\varphi(z)|]\int_{-\infty}^{\infty} g(\theta)\,d\theta\\
&\quad + 2^{3m/2}\sup_\theta[|\theta| g(\theta)]\int_{-\infty}^{\infty}|\varphi(2^m\theta-k)|\,d\theta\\
&\le 2^{m/2}C_\varphi + 2^{m/2}\sup_\theta[|\theta| g(\theta)]\,\|\varphi\|_{L_1},
\end{aligned}$$

where $\|\varphi\|_{L_1} < \infty$ by virtue of (2.5). Let $C_\varphi^* = C_\varphi + \sup_\theta[|\theta| g(\theta)]\,\|\varphi\|_{L_1}$. Thus,

$$\sum_{|k|>K_n} a_{m,k}^2 \le [C_\varphi^*]^2 2^m \sum_{|k|>K_n} k^{-2} = O(2^m K_n^{-1}).$$

Repeating similar calculations for $b_{j,k}$ we obtain

$$\sum_{|k|>L_n} b_{j,k}^2 = O(2^j L_n^{-1}).$$

Then,

$$\begin{aligned}
\sup_{g\in\mathscr{S}_\alpha^*} \text{MISE}\left(\hat{g}_n^{(LF)}\right) &\le \sup_{g\in\mathscr{S}_\alpha^*} \text{MISE}\left(\hat{g}_n^{(L)}\right) + \sum_{|k|>K_n} a_{m,k}^2\\
&\le O\Big(n^{-2\alpha/(2\alpha+2\gamma+1)}\Big) + O\Big(n^{1/(2\alpha+2\gamma+1)} K_n^{-1}\Big)\\
&= O\Big(n^{-2\alpha/(2\alpha+2\gamma+1)}\Big)\\
&\quad + O\Big(n\, K_n^{-1}\; n^{-2\alpha/(2\alpha+2\gamma+1)}\; n^{-2\gamma/(2\alpha+2\gamma+1)}\Big),
\end{aligned}$$

which implies (3.9).

To obtain (3.11), note that MISE $(\hat{g}_n^{(NF)}) = R_1 + R_2 + R_3 + R_4 + R_5 + R_6$, where R_1, R_2, R_3 and R_4 have the same form as in (A.8), the only difference being that the infinite sums $\sum_{k\in\mathbf{Z}}$ are replaced by their finite analogs and

$$R_5 = \sum_{|k|>M_n} a_{m,k}^2, \qquad R_6 = \sum_{j=m}^{m+r} \sum_{|k|>L_n} b_{j,k}^2. \tag{A.17}$$

Repeating the proof of Theorem 2 with finite sums, we show that replacing infinite sums by finite sums does not increase $\sup_{g\in\mathscr{S}_\alpha^*}(R_1+R_2+R_3+R_4)$. For the last two components, R_5 and R_6, the following relations hold: $\sup_{g\in\mathscr{S}_\alpha^*} R_5 = O([\ln n]^{2+\varepsilon} M_n^{-1}) = o(n^{-2\alpha/(2\alpha+2\gamma+1)})$ and $\sup_{g\in\mathscr{S}_\alpha^*} R_6 = O(n^{1/(2\gamma+1)} L_n^{-1}) = o(n^{-2\alpha/(2\alpha+2\gamma+1)})$. This completes the proof. □

PROOF OF THEOREM 4. The proof follows directly from Theorem 1 and the fact that

$$\begin{aligned}\sup_{g\in\mathscr{S}_{\alpha,\nu,\varrho}^*} \sum_{|k|>K_n} a_{m,k}^2 &\le O\Big(n^{-2\alpha/(2\alpha+2\gamma+1)}\Big) + O\Big(n^{1/(2\alpha+2\gamma+1)} K_n^{-1}\Big)\\ &= O(n^{-1}).\end{aligned}$$

□

PROOF OF THEOREM 5. The proof is identical to the proof of Theorem 2; the only difference is that $\varrho > 0$ in Lemma 2. Therefore, $\sup_{g\in\mathscr{S}_{\alpha,\nu,\varrho}^*} R_{1,2} = O(n^{-1}(\ln n)^{(2\gamma+1)/\nu})$,

$$2^M = (2\varrho)^{-1/\nu}(3/2\pi)\big[\ln n - \nu^{-1}(2\alpha+2\gamma+1)\ln\ln n\big]^{1/\nu}$$

and $\sup_{g\in\mathscr{S}_{\alpha,\nu,\varrho}^*} R_{2,1} \sim \sup_{g\in\mathscr{S}_{\alpha,\nu,\varrho}^*} R_{2,2} = O(n^{-1}(\ln n)^{(2\gamma+1)/\nu})$. To complete the proof, note that $\sup_{g\in\mathscr{S}_{\alpha,\nu,\varrho}^*} R_4 = O(n^{-1}(\ln n)^{(2\gamma+1)(2+\varepsilon)})$. □

Acknowledgments. Both authors express their gratitude to the Editor and two anonymous referees for their encouragement and constructive criticism. The referees' suggestions improved the paper.

REFERENCES

ABRAMOVICH, F. and SILVERMAN, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika* **85** 115–129.

ANTONIADIS, A., GRÉGOIRE, G. and MCKEAGUE, I. W. (1994). Wavelet method for curve estimation. *J. Amer. Statist. Assoc.* **89** 1340–1353.

CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.

DESOUZA, C. M. (1991). An empirical Bayes formulation of cohort models in cancer epidemiology. *Statistics in Medicine* **10** 1241–1256.

DEVROYE, L. (1989). Consistent deconvolution in density estimation. *Canad. J. Statist.* **17** 235–239.

DIGGLE, P. J. and HALL, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Assoc. Ser. B* **55** 523–531.

DONOHO, D. and JOHNSTONE, I. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224.

DONOHO, D., JOHNSTONE, I., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539.

EFROMOVICH, S. (1997). Density estimation for the case of supersmooth measurement error. *J. Amer. Statist. Assoc.* **92** 526–535.

FAN, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problem. *Ann. Statist.* **19** 1257–1272.

FAN, J. (1991b). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A* **53** 97–110.

FAN, J. (1991c). Global behavior of deconvolution kernel estimates. *Statist. Sinica* **1** 541–551.

FAN, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.* **21** 600–610.

HALL, P., KERKYACHARIAN, G. and PICARD, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. *Ann. Statist.* **26** 922–942.

HALL, P. and PATIL, P. (1995). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.* **23** 905–928.

HALL, P., PENEV, S., KERKYACHARIAN, G. and PICARD, D. (1997). Numerical performance of block thresholded wavelet estimators. *Statistics Comput.* **7** 115–124.

HERNÁNDEZ, E. and WEISS, G. (1996). *A First Course on Wavelets*. CRC Press, Boca Raton, FL.

KERKYACHARIAN, G. and PICARD, D. (1992). Density estimation in Besov spaces. *Statist. Probab. Lett.* **13** 15–24.

LIU, M. C. and TAYLOR, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.* **17** 427–438.

LOUIS, T. A. (1991). Using empirical Bayes methods in biopharmaceutical research. *Statistics in Medicine* **10** 811–827.

MASRY, E. (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory* **37** 1105–1115.

MASRY, E. (1993a). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic Process. Appl.* **47** 53–74.

MASRY, E. (1993b). Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. *J. Multivariate Anal.* **44** 47–68.

MASRY, E. (1994). Probability density estimation from dependent observations using wavelet orthonormal bases. *Statist. Probab. Lett.* **21** 181–194.

PENSKAYA, M. (1985). Projection estimators of the density of an a priori distribution and of functionals of it. *Theory Probab. Math. Statist.* **31** 113–124.

PENSKY, M. (1999). Estimation of a smooth density function using Meyer-type wavelets. *Statist. Decisions* **17** 111–123

PITERBARG, V. and PENSKAYA, M. (1993). On asymptotic distribution of integrated squared error of an estimate of a component of a convolution. *Math. Methods Statist.* **2** 30–41.

STEFANSKY, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.* **9** 229–235.

STEFANSKI, L. and CARROL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184.

TAYLOR, R. L. and ZHANG, H. M. (1990). On a strongly consistent non-parametric density estimator for deconvolution problem. *Comm. Statist. Theory Methods* **19** 3325–3342.

TALAGRAND M. (1994). Sharper bounds for empirical processes. *Ann. Probab.* **22** 28–76.

WALTER, G. G. (1981). Orthogonal series estimators of the prior distribution. *Sankhyā Ser. A* **43** 228–245.

WALTER, G. G. (1994). *Wavelets and Other Orthogonal Systems with Applications*. CRC Press, Boca Raton, FL.

ZAYED, A. I. and WALTER, G. G. (1996). Characterization of analytic functions in terms of their wavelet coefficients. *Complex Variables* **29** 265–276.

ZHANG, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–831.

ZHANG, C. H. (1992). On deconvolution using time of flight information in positron emission tomography. *Statist. Sinica* **2** 553–575.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CENTRAL FLORIDA
ORLANDO, FLORIDA 32816
E-MAIL: mpensky@pegasus.cc.ucf.edu

INSTITUTE OF STATISTICS
AND DECISION SCIENCES
DUKE UNIVERSITY
DURHAM, NORTH CAROLINA 27708-0251
E-MAIL: brani@isds.duke.edu