Ann. Inst. Statist. Math.
Vol. 47, No. 4, 645–663 (1995)

DECONVOLVING A DENSITY FROM CONTAMINATED DEPENDENT OBSERVATIONS*

CHRISTIAN H. HESSE

Mathematisches Institut A, Universität Stuttgart, 70569 Stuttgart, Germany

(Received August 22, 1994; revised March 20, 1995)

Abstract. The paper studies the performance of deconvoluting kernel density estimators for estimating the marginal density of a linear process. The data stem from the linear process and are partially, respectively fully contaminated by iid errors with a known distribution. If $1 - p$ denotes the proportion of contaminated observations (and it is , of course, unknown which observations are contaminated and which are not) then for $1 - p \in (0, 1)$ and under mild conditions almost sure deconvolution rates of order $O(n^{-2/5}(\log n)^{9/10})$ can be achieved for convergence in $\mathcal{L}_\infty$. This rate compares well with the existing rates for *iid uncontaminated* observations. For $p = 0$ and exponentially decreasing error characteristic function the corresponding rates are of merely logarithmic order. As a by-product the paper also gives a rate of convergence result for the empirical characteristic function in the linear process context and utilizes this to demonstrate that deconvoluting kernel density estimators attain the optimal rate in the dependence case with exponentially decreasing error characteristic function.

Key words and phrases: Deconvolution, density estimation, contamination, identifiability, dependence.

1. Introduction

Suppose that (X_j, ε_j), $j = 1, 2, \dots, n$ are iid bivariate random vectors where X_1 has an unknown density g, ε_1 has a known density and is independent of X_1. It is desired to estimate the density g based on observations that are corrupted by additive noise, i.e. based on observations

$$Y_j = X_j + \varepsilon_j. \tag{1.1}$$

Problems in which these mixture models are relevant do occur in many branches of statistics. One area is the empirical Bayes approach to compound decision problems, as discussed in Robbins (1964). To connect this with our notation assume

* This work was partially supported by a grant from the Deutsche Forschungsgemeinschaft.

that $\{X_1, \ldots, X_n\}$ is a set of parameters, $G(x)$ is an unknown prior distribution and conditionally on $X_j = x$ the observed Y_j are realizations of independent random variables with known parametric density $f(y/x)$. Then unconditionally the Y_j are realizations from the density $\int f(y/x)dG(x)$. Robbins (1964) recommends the use of $Y_1, \ldots, Y_{n-1}$ to estimate $G(x)$ and then to use this estimate to compute the posterior distribution of X_n given Y_n.

Deconvolution problems also appear in errors-in-variables models for nonlinear regression, see e.g. Carroll *et al.* (1984). Further applications are mentioned in Carroll and Hall (1988) and Zhang (1990), Crump and Seinfeld (1982), Mendelsohn and Rice (1982), Snyder *et al.* (1988) and some of the references therein.

In the convolution context the question of identifiability arises. We shall call a density q identifiable in a general convolution with a density p if the following implication holds a.e.

$$\int p(x-y)q(y)dy = \int p(x-y)\tilde{q}(y)dy \Rightarrow q(y) = \tilde{q}(y).$$

This may be characterized in terms of the characteristic functions Ψ_p, Ψ_q, $\Psi_{\tilde{q}}$ as

$$\Psi_p(t) \cdot \Psi_q(t) = \Psi_p(t) \cdot \Psi_{\tilde{q}}(t) \Rightarrow \Psi_q(t) = \Psi_{\tilde{q}}(t) \forall t.$$

The last implication holds if $\Psi_p(t) \neq 0$ for all t and this is what we will require of our error characteristic function although there are weaker conditions that ensure identifiability in the present set-up.

A variety of approaches for estimation of g have been considered in the literature: a maximum likelihood method is given in Snyder *et al.* (1988), B-Splines are used by Mendelsohn and Rice (1982), and in Masry and Rice (1992) Gaussian deconvolution is based on estimates of derivatives of g. Liu and Taylor (1989) seems to be the first published work that investigates the performance of kernel-type estimators in this context. Other work on deconvolving kernel estimators includes Zhang (1990), Stefanski (1990), Fan *et al.* (1990), Stefanski and Carroll (1990), Fan (1991*a*, 1991*b*, 1992), Fan and Truong (1993), Fan and Masry (1992), and Masry (1991*a*, 1991*b*, 1993*a*, 1993*b*). We give a brief discussion of this literature devoted to regression with errors-in-variables and density estimation from contaminated observations as it relates to the present paper.

In the iid case the nonparametric regression estimation problem with errors-in-variables was studied by Fan *et al.* (1990) and Fan and Truong (1993) where optimal rates of convergence and asymptotic normality are established for the estimator

$$\hat{m}(x) = \frac{\sum_{j=1}^{n} Y_j W_\lambda((x-X_j)/\lambda(n))}{\sum_{j=1}^{n} W_\lambda((x-X_j)/\lambda(n))} \tag{1.2}$$

of the regression function $m(x) = E(Y \mid X = x)$. The estimator is based on observations $(X_j + \varepsilon_j, Y_j)$, i.e. covariates perturbed by error. In (1.2) W_λ denotes a deconvolution kernel and $\lambda(n)$ is the bandwidth parameter.

Fan and Masry (1992) obtain asymptotic normality for $\hat{m}(x)$ in the more general setting where $\{X_j\}$ and $\{Y_j\}$ are individually and jointly dependent. Finally, for this general setting Masry (1991*b*) gives sharp almost sure rates.

The other papers cited above are concerned with versions of the deconvolution problem when observations Y_j are of type (1.1) and estimation of the density of X_j is desired. The main focus of Liu and Taylor (1989), Stefanski (1990), Stefanski and Carroll (1990), Zhang (1990), and Fan (1991*a*, 1992) is on the iid case and on bounds for the rate of quadratic-mean convergence for deconvoluting kernel density estimators: Liu and Taylor (1989) study the mean square error at a fixed point, Stefanski and Carroll (1990) provide bounds for the integrated mean square error, Zhang (1990) contains both upper and lower bounds for $\mathcal{L}_2$-loss and Fan (1991*a*, 1992) obtains optimal local rates for the mean-square error at a point and optimal global rates under $\mathcal{L}_p$ and weighted $\mathcal{L}_p$-loss uniformly over a class of densities.

The more general context of processes $\{X_j\}$ satisfying a variety of mixing conditions, with dependence among $\{X_j\}$ and $\{Y_j\}$ and the estimation problem being extended to the joint density function $f(x_1, \dots, x_p)$ of the random variables $X_1, \dots, X_p$ $(p \geq 1)$ has been considered by Masry (1991*a*). There bounds as well as precise asymptotic expressions for the mean square estimation error at a point are provided.

In Masry (1993*b*) for essentially the same framework almost sure uniform convergence rates over compact subsets of $\mathbb{R}^p$ are given for estimators of $f(x_1, \dots, x_p)$. Both ordinary smooth and super smooth noise distributions are considered and are found to significantly influence the convergence.

Finally, we also mention that Fan (1991*b*) has proved the asymptotic normality of deconvoluting kernel estimates for the iid case and that Masry (1993*a*) recently extended this to the multivariate set-up.

In general, deconvolution rates of kernel estimators are intimately connected to the decay of the characteristic function of the error density. For example, in the practically important case of normally distributed errors and iid X_j Carroll and Hall (1988) show that the rate of convergence of any estimator cannot be faster than $O\left((\log n)^{-s/2}\right)$ over densities in $C_s(M) = \{g : \sup_x g(x) \leq M$ and $\sup_x |g^{(s)}(x)| \leq M\}$.

The contents of the present paper generalizes existing results in several directions. As in Masry (1993*b*) we also allow for dependence among the $\{X_j\}$ but focus on a different dependence structure, namely linear process dependence. Secondly, we aim for sharp uniform rates over the entire domain and not merely over compact subsets thereof, and thirdly and most importantly, we allow for partially contaminated observations. In this context the somewhat surprising fact is found that in terms of uniform rates of the deconvolution kernel estimators any fixed non-zero proportion of uncontaminated observations within a set of contaminated observations is essentially as good as a full set of uncontaminated observations. In addition, the methods employed which are related to the notion of metric entropy seem to be new in the deconvolution context. Our basic set-up is

$$Y(j) = \sum_{k=0}^{\infty} \rho_k \varepsilon(j-k) + T(j)e(j) \qquad j = 1, \dots, n, \tag{1.3}$$

where $\varepsilon(j)$, $e(j)$ are (mutually) iid random variables with unknown (in the case of $\varepsilon(j)$), respectively known (in the case of $e(j)$) density; $T(j)$ are independent (of each other and of the $\varepsilon(j)$ and the $e(j)$) Bernoulli random variables with parameter $1-p \in (0,1)$. For practical examples of partially contaminated distributions in the iid context see e.g. Huber (1981), in particular the discussion of the gross error model, and also Tukey (1960).

As an estimator of the density g we propose

$$\hat{g}_n(x) = \frac{1}{2\pi\lambda(d_n)} \int_{-1}^{+1} e^{-itx\lambda^{-1}(d_n)} \hat{\Psi}_Y(t\lambda^{-1}(d_n)) \frac{\Psi_W(t)}{\Psi_{Te}(t\lambda^{-1}(d_n))} dt \tag{1.4}$$

where $\hat{\Psi}_Y(t)$ denotes the empirical characteristic function of the observations $Y(1), \dots, Y(n)$; Ψ_{Te}, Ψ_W are the characteristic functions of $T(1)e(1)$ and the random variable W, respectively, whose density is used as a kernel and will be introduced below. Furthermore, $\lambda(n) = cn^{-\delta}$ with $\delta \in (0,1)$ to be determined later is a bandwidth function and $d_n = cn/\log n$. Here and below c always denotes a generic finite constant which may change from one occurrence to another.

Apart form a sharp uniform rate of convergence result for $\hat{g}_n(x)$ the paper also gives a rate of convergence for the empirical characteristic function in the linear process context and utilizes this to demonstrate that deconvoluting kernel density estimators attain the optimal rate in the dependence case with exponentially decreasing error characteristic function.

2. Convergence

In this section we will show that with an optimal choice of the bandwidth $\lambda(n)$ the estimator $\hat{g}_n(x)$ converges to $g(x)$ at a rate of $O(n^{-2/5}(\log n)^{9/10})$ a.s. This rate compares well with the one for uncontaminated iid observations, see Karunamuni and Mehra (1990), and is significantly faster than the deconvolution rates (even in the iid case) for normally distributed contaminating errors, see Carroll and Hall (1988). Hence the result is insofar surprising as the "best" available observations essentially determine the almost sure rate of the deconvoluting kernel estimator rather than those observations of lesser quality and this is so irrespective of what their relative proportions are.

The conditions we require are fairly mild and will now be introduced.

A: The random variables $\varepsilon(j)$ are iid with absolutely continuous distribution and a finite absolute moment of order $\alpha > 0$. $T(j)$ are iid Bernoulli random variables with parameter $(1-p) \in (0,1)$; $|\rho_k| \le c\rho^k$ for some $\rho \in (0,1)$.

B: The density $g(x)$ of $X(j)$ is bounded and has uniformly absolutely bounded and continuous derivatives up to order two, i.e. $\sup_{x\in\mathbb{R}} g(x) \le M$, $\sup_{x\in\mathbb{R}} |g'(x)| \le M$, $\sup_{x\in\mathbb{R}} |g''(x)| \le M$.

C: $\inf_t |\Psi_{Te}(t)| \geq p$ and $x^2 \int_{-1}^{+1} e^{-itx}\Psi_W(t)/\Psi_{Te}(t/\lambda(n))dt \longrightarrow^{x\to\infty} 0$ for $\lambda(n) = O(n^{-\delta})$ and any $\delta \in (0,1)$.

D: W is a zero mean and finite variance random variable with a symmetric density $W(z)$ and a characteristic function $\Psi_W(t)$ that vanishes off $[-1,+1]$.

E: $P(|Y(i)| > a) \leq a^{-\tau}$ for all $a > 0$ and some $\tau > 2(1+\delta)/(1-3\delta)$.

It is clear by the Riemann-Lebesgue lemma that the second part of Condition C is satisfied for example if $\Psi_W(t)/\Psi_{Te}\left(t\lambda^{-1}(n)\right)$ has two continuous integrable derivatives. An example of a density satisfying Condition D is

$$W(z) = \frac{3}{8\pi}\left(\frac{\sin(z/4)}{z/4}\right)^4$$

with characteristic function

$$\Psi_W(t) = \begin{cases} 6t^3 - 6t^2 + 1 & \text{for } 0 \leq t \leq 1/2 \\ -2t^3 + 6t^2 - 6t + 2 & \text{for } 1/2 < t \leq 1, \end{cases}$$

and for $t \in [-1, 0)$ given by symmetry. $\Psi_W(t)$ also possesses two continuous integrable derivatives.

We are now ready to state the main

THEOREM 2.1. *Under Conditions* A-E *the following rate is obtained for convergence of* $\hat{g}_n(x)$ *to* $g(x)$:

$$P\left(\limsup_{n\to\infty} n^{2/5}(\log n)^{-9/10} \sup_{x\in\mathbb{R}} |\hat{g}_n(x) - g(x)| < \infty\right) = 1.$$

This rate is achieved for the bandwidth choice $\lambda(n) = cn^{-1/5}$.

PROOF. Define the process

$$\begin{aligned} \tilde{Y}(j) &= \sum_{k=0}^{m(n)-1} \rho_k \varepsilon(j-k) + T(j)e(j) \\ &=: \tilde{X}(j) + T(j)e(j), \qquad j = 1, \ldots, n \end{aligned}$$

where $m(n) = [\tilde{c}\log n]$ with some appropriate constant $\tilde{c}$ to be chosen later. $[x]$ denotes the largest integer smaller than or equal to x. The density of $\tilde{X}(j)$ is the convolution of $\{(\rho_k)^{-1}f((\rho_k)^{-1}x) : k = 0, 1, \ldots, m(n)-1\}$ where f denotes the density of $\varepsilon(j)$.

Consider first

$$(2.1) \qquad \hat{g}_{n,j}(x) = \frac{1}{2\pi\lambda(d_n)} \int_{-1}^{+1} e^{-itx\lambda^{-1}(d_n)} \hat{\Psi}^j_{\tilde{Y}}(t\lambda^{-1}(d_n)) \frac{\Psi_W(t)}{\Psi_{Te}(t\lambda^{-1}(d_n))} dt$$

where $\hat{\Psi}^j_{\tilde{Y}}(t) = \frac{1}{n_j}\sum_{k=0}^{n_j-1} \exp(it\tilde{Y}(j+km(n))$ is the empirical characteristic function of $\tilde{Y}$ based on $\tilde{Y}(j), \tilde{Y}(j+m(n)), \ldots, \tilde{Y}(j+(n_j-1)m(n))$ and (n_j-1) is the largest integer so that $(n_j-1)[\tilde{c}\log n] \le n$. Observe that n_j is of order $O(n/\log n)$. Clearly, $\hat{g}_{n,j}(x)$ are not statistics since they are not functions of the observations $Y_1, \dots, Y_n$. It is crucial for the proof to link them to $\hat{g}_n(x)$. This will be done later, first we will evaluate the properties of $\hat{g}_{n,j}(x)$ as an "estimator" of the density of $\tilde{X}(1)$ which we denote by $\tilde{g}$.

In particular we will show that

$$P\left(\limsup_{n\to\infty} n^{(1-\delta)/2}(\log n)^{\delta/2-1} \sup_{x\in\mathbb{R}} |\hat{g}_{n,j}(x) - E\hat{g}_{n,j}(x)| < B\right) = 1, \tag{2.2}$$

and for the bias

$$P\left(\limsup_{n\to\infty} n^{2\delta}(\log n)^{-2\delta} \sup_{x\in\mathbb{R}} |E\hat{g}_{n,j}(x) - \tilde{g}(x)| < B\right) = 1 \tag{2.3}$$

for all $j = 1, 2, \dots, m(n)$, B sufficiently large and the bandwidth function $\lambda(n) = cn^{-\delta}$, $\delta \in (0,1)$. To demonstrate (2.2) it suffices to show that

$$\begin{aligned} P\Bigg(\limsup_{n\to\infty} \sup_{W_n^x\in\tilde{W}_n} \Bigg| \frac{1}{n_j}\sum_{i=0}^{n_j-1} W_n^x(\tilde{Y}(j+im(n))) \\ - E(W_n^x(\tilde{Y}(j+im(n))))\Bigg| < B\Bigg) = 1 \end{aligned} \tag{2.4}$$

where

$$\begin{aligned} W_n^x(y) &= n^{(1+\delta)/2}(\log n)^{-\delta/2-1} W_n((y-x)/\lambda(d_n)), \\ W_n(z) &= \frac{cn^{-\delta}(\log n)^\delta}{2\pi} \int_{-cn^\delta(\log n)^{-\delta}}^{+cn^\delta(\log n)^{-\delta}} e^{itz\lambda(d_n)} \frac{\Psi_W(t\lambda(d_n))}{\Psi_{Te}(t)}\,dt, \end{aligned}$$

and

$$\tilde{W}_n = \{W_n^x(y) : x \in \mathbb{R}\}.$$

$\tilde{W}_n$ is a sequence of sets of functions whose elements depend on n. We prove (2.4) by introducing

$$\begin{aligned} {}^*W_n^{x_j}(y) &= (2\pi)^{-1} n^{(1-\delta)/2}(\log n)^{\delta/2-1} \\ &\cdot \int_{-cn^\delta(\log n)^{-\delta}}^{+cn^\delta(\log n)^{-\delta}} e^{-it(x_j-y)} \frac{\Psi_W(ctn^{-\delta}(\log n)^\delta)}{\Psi_{Te}(t)}\,dt \end{aligned} \tag{2.5}$$

for $x_j = j\Delta$ and $j \in \{0, \pm 1, \dots, \pm[2KD^{-1}]\} := I$ with

$$\begin{aligned} K &= c\varepsilon^{-1/2} n^{(1-3\delta)/4}(\log n)^{-1/2+3\delta/4}, \\ D &= c\varepsilon n^{-3\delta/2-1/2}(\log n)^{1+3\delta/2}. \end{aligned}$$

In addition, for $j = [2KD^{-1}] + 1$ set

$$(2.6) \qquad {}^{*}W_n^{x_j}(y) = \begin{cases} (2\pi p)^{-1} n^{(1+\delta)/2} (\log n)^{-\delta/2-1} & \text{for } |y| > K \\ \varepsilon/2 & \text{for } |y| \le K. \end{cases}$$

Similarly,

$$(2.7) \qquad {}_{*}W_n^{x_j}(y) = (2\pi)^{-1} n^{(1-\delta)/2} (\log n)^{\delta/2-1} \cdot \int_{-cn^{\delta}(\log n)^{-\delta}}^{+cn^{\delta}(\log n)^{-\delta}} e^{-it(x_j-y)} \frac{\Psi_W(ctn^{-\delta}(\log n)^{\delta})}{\Psi_{Te}(t)} dt - \varepsilon$$

for $x_j = j\Delta$, $j \in I$, and finally

$$(2.8) \qquad {}_{*}W_n^{x_j}(y) \equiv 0$$

for $j = [2KD^{-1}] + 1$. There are $L(n) = O(n^{3/4+3\delta/4}(\log n)^{-5/4-3\delta/4})$ pairs of functions ${}^{*}W_n^{x_j}(y)$, ${}_{*}W_n^{x_j}(y)$. They exhibit the properties stated in Lemma 2.3 below.

Now, with $J = I \cup \{[2K/D] + 1\}$

$$\begin{aligned}
(2.9) \qquad & P\left(\sup_{W_n^x \in \tilde{W}_n} \left| c \log n/n \sum_{i=0}^{cn/\log n - 1} W_n^x(\tilde{Y}(j + im(n))) - E(W_n^x(\tilde{Y}(j+im(n)))) \right| > c_1 + 2c_2 + 4\varepsilon \right) \\
& \le P\left(\max_{k \in J} \left| c \log n/n \sum_{i=0}^{cn/\log n - 1} \tilde{W}_n^{x_k}(\tilde{Y}(j + im(n))) - E(\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n)))) \right| > c_1 \right) \\
& + P\left(\left| \max_{k \in J} \left| c \log n/n \sum_{i=0}^{cn/\log n - 1} \tilde{W}_n^{x_k}(\tilde{Y}(j + im(n))) - E(\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n)))) \right| \right.\right. \\
& \left.\left. - \sup_{W_n^x \in \tilde{W}_n} \left| c \log n/n \sum_{i=0}^{cn/\log n - 1} W_n^x(\tilde{Y}(j + im(n))) - E(W_n^x(\tilde{Y}(j+im(n)))) \right| \right| > 2c_2 + 4\varepsilon \right)
\end{aligned}$$

where $\tilde{W}_n^{x_k}$ is a fixed but arbitrary element of the set W_n^k which we define as

$$W_n^k = \{W_n^x(y) : |W_n^x(y) - {}_{*}W_n^{x_k}(y)| \le |{}^{*}W_n^{x_k}(y) - {}_{*}W_n^{x_k}(y)| \forall y \in \mathbb{R}\}$$

and clearly $\tilde{W}_n = \bigcup\{W_n^k : k \in J\}$. The introduction of $\tilde{W}_n$ and the construction (2.5)–(2.9) is related to the metric entropy techniques introduced in Dudley (1978).

We now bound the variance of $\tilde{W}_n^{x_k}(\tilde{Y}(j + im(n)))$. We have

$$\begin{aligned}
&\mathrm{var}(\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n))))\\
&\quad\le \max_{1\le j\le m(n)} \max_{k\in J} \sup_{\tilde{W}_n^{x_k}\in W_n^k} E((\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n))))^2)\\
&\quad\le \max_j \max_k \max_{\tilde{W}_n^{x_k}} \left(\sup_{y\in\mathbb{R}} |\tilde{W}_n^{x_k}(y)|\right) E|\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n)))|\\
&\quad\le cn(\log n)^{-2}\\
&\qquad\cdot \max_j \max_k \max_{\tilde{W}_n^{x_k}} \left(\sup_y \left|\int_{-1}^{+1} e^{ict(y-x_k)n^\delta(\log n)^{-\delta}} \frac{\Psi_W(t)}{\Psi_{Te}(ctn^\delta(\log n)^{-\delta})} dt\right|\right)\\
&\qquad\cdot E\left(\left|\int_{-cn^\delta(\log n)^{-\delta}}^{+cn^\delta(\log n)^{-\delta}} e^{-itx_k} e^{it\tilde{Y}(1)} \frac{\Psi_W(ctn^{-\delta}(\log n)^\delta)}{\Psi_{Te(t)}} dt\right|\right)
\end{aligned}$$

and since as a consequence of Condition D, $\Psi_W(t)$ is integrable and we can obtain the bound

$$\begin{aligned}
&\mathrm{var}(\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n))))\\
&\quad\le cn(\log n)^{-2} \max_j E(|\hat{g}_{n,j}(x)|)\\
&\quad\le cn(\log n)^{-2} \max_j \left(\sup_{x\in\mathbb{R}} \tilde{g}(x) + \sup_{x\in\mathbb{R}} |\tilde{g}(x) - E(|\hat{g}_{n,j}(x)|)|\right)\\
&\quad\le cn(\log n)^{-2} \max_j \Bigg(\sup_{x\in\mathbb{R}} g(x) + \sup_{x\in\mathbb{R}} |g(x) - \tilde{g}(x)|\\
&\qquad\qquad + \sup_{x\in\mathbb{R}} |\tilde{g}(x) - E(|\hat{g}_{n,j}(x)|)|\Bigg)\\
&\quad\le cn(\log n)^{-2}
\end{aligned}$$

by Lemma 2.2 below and by an argument similar to the one that establishes (2.3) which is proved below. In addition,

$$\begin{aligned}
&\max_{1\le j\le m(n)} \max_{k\in J} \sup_{\tilde{W}_n^{x_k}} |\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n))) - E(\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n))))|\\
&\quad\le 2\max_j \max_k \sup_{\tilde{W}_n^{x_k}} \sup_{y\in\mathbb{R}} |\tilde{W}_n^{x_k}(y)|\\
&\quad\le cn^{(1+\delta)/2}(\log n)^{-1-\delta/2}
\end{aligned}$$

so that

$$P\left(\max_{k\in J}\left| c\log n/n \sum_{i=0}^{cn/\log n - 1} \tilde{W}_n^{x_k}(\tilde{Y}(j+im(n)))\right.\right.$$

$$
\begin{aligned}
&- E(\tilde W_n^{x_k}(\tilde Y(j+im(n))))\Big| > c_1\Big) \\
&\le L(n)P\left(\left|c\log n/n \sum_{i=0}^{cn/\log n-1} \tilde W_n^{x_k}(\tilde Y(j+im(n)))\right.\right. \\
&\qquad\left.\left. - E(\tilde W_n^{x_k}(\tilde Y(j+im(n))))\right| > c_1\right) \\
&\le 2L(n)\exp\Big(-cn(\log n)^{-1}c_1^2/\Big(2\operatorname{var}(\tilde W_n^{x_k}(\tilde Y(j+im(n)))) \\
&\qquad + (4/3)c_1\sup_y|\tilde W_n^{x_k}(y)|\Big)\Big) \\
&\le cL(n)n^{-cc_1^2}
\end{aligned}
$$

by Bernstein's inequality. We therefore have established a bound for the first summand on the right-hand side of (2.9).

For the remaining term in (2.9) note that for sufficiently large n

$$
\begin{aligned}
&\Bigg|\max_{k\in J}\Bigg|c\log n/n \sum_{i=0}^{cn/\log n-1} \tilde W_n^{x_k}(\tilde Y(j+im(n))) - E(\tilde W_n^{x_k}(\tilde Y(j+im(n))))\Bigg| \\
&\qquad - \sup_{W_n^x\in\tilde W_n}\Bigg|c\log n/n \sum_{i=0}^{cn/\log n-1} W_n^x(\tilde Y(j+im(n))) \\
&\qquad\qquad - E(W_n^x(\tilde Y(j+im(n))))\Bigg|\Bigg| \\
&\le \max_{k\in J}\sup_{W_n^x\in W_n^k}\Bigg|c\log n/n \sum_{i=0}^{cn/\log n-1} (\tilde W_n^{x_k}(\tilde Y(j+im(n))) \\
&\qquad - W_n^x(\tilde Y(j+im(n)))) \\
&\qquad - E(\tilde W_n^{x_k}(\tilde Y(j+im(n))) - W_n^x(\tilde Y(j+im(n))))\Bigg| \\
&\le \max_{k\in J}\sup_{W_n^x\in W_n^k}\Bigg|c\log n/n \sum_{i=0}^{cn/\log n-1} (\tilde W_n^{x_k}(\tilde Y(j+im(n))) \\
&\qquad - {}_*W_n^{x_k}(\tilde Y(j+im(n)))) \\
&\qquad - E(\tilde W_n^{x_k}(\tilde Y(j+im(n))) - {}_*W_n^{x_k}(\tilde Y(j+im(n))))\Bigg| \\
&+ \Bigg|c\log n/n \sum_{i=0}^{cn/\log n-1} ({}_*W_n^{x_k}(\tilde Y(j+im(n))) - W_n^x(\tilde Y(j+im(n))))
\end{aligned}
$$

$$- E({}_*W_n^{x_k}(\tilde{Y}(j+im(n))) - W_n^x(\tilde{Y}(j+im(n))))\Bigg|$$

$$\leq \max_k \sup_{W_n^x} \Bigg(\Bigg| c\log n/n \sum_{i=0}^{cn/\log n - 1} (\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n))) - {}_*W_n^{x_k}(\tilde{Y}(j+im(n)))) - E(\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n))) - {}_*W_n^{x_k}(\tilde{Y}(j+im(n))))\Bigg|$$

$$+ \Bigg| c\log n/n \sum_{i=0}^{cn/\log n - 1} |{}^*W_n^{x_k}(\tilde{Y}(j+im(n))) - {}_*W_n^{x_k}(\tilde{Y}(j+im(n)))| - E(|{}^*W_n^{x_k}(\tilde{Y}(j+im(n))) - {}_*W_n^{x_k}(\tilde{Y}(j+im(n)))|)\Bigg| + 4\varepsilon\Bigg)$$

since $E(|{}^*W_n^{x_k}(Y(j+im(n))) - {}_*W_n^{x_k}(Y(j+im(n)))|) \leq \varepsilon$ by construction and $\tilde{Y}(k)$ converges to $Y(k)$ a.s. at an arbitrary rate if $\tilde{c}$ is increased so that for n large enough

$$E(|{}^*W_n^{x_k}(\tilde{Y}(j+im(n))) - {}_*W_n^{x_k}(\tilde{Y}(j+im(n)))|) \leq 2\varepsilon.$$

Therefore the second term on the right side of (2.9) is upper-bounded by

$$(2.10)\quad P\Bigg(\max_{k\in J}\Bigg| c\log n/n \sum_{i=0}^{cn/\log n - 1} (\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n))) - {}_*W_n^{x_k}(\tilde{Y}(j+im(n)))) - E(\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n))) - {}_*W_n^{x_k}(\tilde{Y}(j+im(n))))\Bigg| > c_2\Bigg)$$

$$+ P\Bigg(\max_{k\in J}\Bigg| c\log n/n \sum_{i=0}^{cn/\log n - 1} |{}^*W_n^{x_k}(\tilde{Y}(j+im(n))) - {}_*W_n^{x_k}(\tilde{Y}(j+im(n)))| - E(|{}^*W_n^{x_k}(\tilde{Y}(j+im(n))) - {}_*W_n^{x_k}(\tilde{Y}(j+im(n)))|)\Bigg| > c_2\Bigg)$$

and again both terms can be bounded by Bernstein's inequality:

$$E(|\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n))) - {}_*W_n^{x_k}(\tilde{Y}(j+im(n)))|^2)$$
$$\leq \left(\sup_{y\in\mathbb{R}} |\tilde{W}_n^{x_k}(y) - {}_*W_n^{x_k}(y)|\right) \cdot E(|\tilde{W}_n^{x_k}(\tilde{Y}(j+im(n))) - {}_*W_n^{x_k}(\tilde{Y}(j+im(n)))|)$$
$$\leq c\varepsilon n^{(1+\delta)/2}(\log n)^{-1-\delta/2}$$

by construction of ${}_*W_n^{x_k}$, ${}^*W_n^{x_k}$ and $\tilde{W}_n^{x_k}$. Similarly,

$$E(|{}^*W_n^{x_k}(\tilde{Y}(j+im(n))) - {}_*W_n^{x_k}(\tilde{Y}(j+im(n)))|^2) \le c\varepsilon n^{(1+\delta)/2}(\log n)^{-1-\delta/2}$$

so that both probabilities in (2.10) may be handled in the same fashion. By Bernstein's inequality (2.10) is bounded by

$$4L(n)\exp(-cc_2^2 n(\log n)^{-1}/c\varepsilon n^{(1+\delta)/2}(\log n)^{-1-\delta/2}).$$

Combining the bounds for the terms in (2.9) we finally arrive at

$$\begin{aligned}(2.11)\qquad & P\Bigg(\sup_{W_n^x\in\tilde{W}_n}\Bigg|c\log n/n\sum_{i=0}^{cn/\log n-1} W_n^x(\tilde{Y}(j+im(n))) \\ &\qquad - E(W_n^x(\tilde{Y}(j+im(n))))\Bigg| > c_1+2c_2+4\varepsilon\Bigg) \\ &\le cL(n)n^{-cc_1^2} \\ &\quad + 4L(n)\exp(-cc_2^2 n(\log n)^{-1}/c\varepsilon n^{(1+\delta)/2}(\log n)^{-1-\delta/2}).\end{aligned}$$

The right side of (2.11) is obviously summable and by the Borel-Cantelli lemma (2.2) is found to hold. For (2.3) we first evaluate the expectation

$$\begin{aligned}(2.12)\quad & E\left(\frac{1}{2\pi}\int_{-\lambda^{-1}(d_n)}^{+\lambda^{-1}(d_n)} e^{-itx}\hat{\Psi}_{\tilde{Y}}^j(t)\frac{\Psi_W(t\lambda(d_n))}{\Psi_{Te}(t)}dt\right) \\ &= \frac{1}{2\pi}\int_{-\lambda^{-1}(d_n)}^{+\lambda^{-1}(d_n)} e^{-itx}\frac{\Psi_W(t\lambda(d_n))}{\Psi_{Te}(t)} \\ &\quad\cdot(pE(e^{it\tilde{X}(k)}) + (1-p)E(e^{it(\tilde{X}(k)+e(k))}))dt \\ &= \frac{1}{2\pi}\int_{-\lambda^{-1}(d_n)}^{+\lambda^{-1}(d_n)} e^{-itx}\frac{\Psi_W(t\lambda(d_n))}{\Psi_{Te}(t)} \\ &\quad\cdot\left(p\int_{\mathbb{R}} e^{itz}\tilde{g}(z)dz + (1-p)\int_{\mathbb{R}} e^{ity}\left(\int_{\mathbb{R}}\tilde{g}(z)k(y-z)dz\right)dy\right)dt\end{aligned}$$

where $k(y)$ is the density of $e(1)$. Further,

$$\begin{aligned}&(1-p)\frac{1}{2\pi}\int_{-\lambda^{-1}(d_n)}^{+\lambda^{-1}(d_n)} e^{-itx}\frac{\Psi_W(t\lambda(d_n))}{\Psi_{Te}(t)}\left(\int_{\mathbb{R}} e^{ity}\left(\int_{\mathbb{R}}\tilde{g}(z)k(y-z)dz\right)dy\right)dt \\ &= (1-p)\int_{\mathbb{R}}\frac{1}{2\pi}\int_{-\lambda^{-1}(d_n)}^{+\lambda^{-1}(d_n)} e^{it(z-x)}\frac{\Psi_W(t\lambda(d_n))\Psi_e(t)}{\Psi_{Te}(t)}\tilde{g}(z)dzdt\end{aligned}$$

and so, in summary,

$$\begin{aligned}E\hat{g}_{n,j}(x) &= \frac{1}{2\pi}\int_{\mathbb{R}}\int_{-\lambda^{-1}(d_n)}^{+\lambda^{-1}(d_n)} e^{it(z-x)}\Psi_W(t\lambda(d_n))\tilde{g}_n(z)dtdz \\ &= \int_{\mathbb{R}}\lambda^{-1}(d_n)W((x-z)\lambda^{-1}(d_n))g(z)dz\end{aligned}$$

since the characteristic function of $T(j)e(j)$ is $p + (1-p)\Psi_e(t)$ with $\Psi_e(t)$ being the characteristic function of $e(j)$. Then

$$E\hat{g}_{n,j}(x) - g(x) = \int_{\mathbb{R}} W(y)[g(x - \lambda(d_n)y) - g(x)]dy.$$

Due to Condition B the density g is twice differentiable with uniformly absolutely bounded derivatives so that (remembering that c is a positive constant, not always the same one)

$$\begin{aligned}\sup_{x\in\mathbb{R}} |E\hat{g}_{n,j}(x) - g(x)| &\le c\lambda(d_n) \sup_{x\in\mathbb{R}} |g'(x)| \int_{\mathbb{R}} yW(y)dy \\ &\quad + c\lambda^2(d_n) \sup_{x\in\mathbb{R}} |g''(x)| \int_{\mathbb{R}} y^2 W(y)dy \\ &\le cn^{-2\delta}(\log n)^{+2\delta}\end{aligned}$$

and (2.3) follows with the help of part (b) of Lemma 2.2. Now, to link $\hat{g}_{n,j}(x)$ defined in (2.1) to the sample statistic $\hat{g}_n(x)$ defined in (1.4) we first observe that

$$\min_{1\le j\le m(n)} \hat{g}_{n,j}(x) \le g_n^*(x) \le \max_{1\le j\le m(n)} \hat{g}_{n,j}(x)$$

where

$$g_n^*(x) = \frac{1}{2\pi\lambda(d_n)} \int_{-1}^{+1} e^{-itx\lambda^{-1}(d_n)} \hat{\Psi}_{\tilde{Y}}(t\lambda^{-1}(d_n)) \frac{\Psi_W(t)}{\Psi_{Te}(t\lambda^{-1}(d_n))} dt$$

with

$$\hat{\Psi}_{\tilde{Y}}(t) = \frac{1}{n}\sum_{k=0}^{n} \exp(it\tilde{Y}(k)).$$

By Lemma 2.2

$$\sup_{x\in\mathbb{R}} |\hat{g}_n(x) - g_n^*(x)| = O(n^{-1/2}) \qquad \text{a.s.}$$

for $\tilde{c}$ sufficiently large. Therefore

$$\sup_{x\in\mathbb{R}} |\hat{g}_n(x) - E\hat{g}_n(x)| \le \max_{1\le j\le m(n)} \sup_{x\in\mathbb{R}} |\hat{g}_{n,j}(x) - E\hat{g}_{n,j}(x)| + O(n^{-1/2})$$

and

$$\begin{aligned}&P\left(n^{(1-\delta)/2}(\log n)^{\delta/2-1} \sup_{x\in\mathbb{R}} |\hat{g}_n(x) - E\hat{g}_n(x)| > c + O(n^{-1/2})\right) \\ &\quad \le P\left(\max_{1\le j\le m(n)} \sup_{x\in\mathbb{R}} n^{(1-\delta)/2}(\log n)^{\delta/2-1} |\hat{g}_{n,j}(x) - E\hat{g}_{n,j}(x)| > c\right).\end{aligned}$$

Summation over n gives

$$\sum_{n=n_0}^{\infty} P\left(\max_{1\le j\le m(n)} \sup_{x\in\mathbb{R}} n^{(1-\delta)/2}(\log n)^{\delta/2-1}|\hat{g}_{n,j}(x) - E\hat{g}_{n,j}(x)| > c_1 + 2c_2 + 4\varepsilon\right)$$

$$\le \sum_{n=n_0}^{\infty} \sum_{j=1}^{m(n)} 4L(n)\exp(-cc_2^2 n(\log n)^{-1}/c\varepsilon n^{(1+\delta)/2}(\log n)^{-\delta/2-1})$$

$$+ cL(n)n^{-cc_1^2} < \infty$$

for any $\delta \in (0,1)$ and $\tilde{c}$, c_1, c_2 sufficiently large. Comparing the rates of bias and variance, namely $O(n^{2\delta}(\log n)^{-2\delta})$ and $O(n^{(1-\delta)/2}(\log n)^{\delta/2-1})$, the rate stated in the theorem results from $\delta = 1/5$.

The proof of Theorem 2.1 refers to Lemmas 2.1, 2.2, 2.3 which we establish now.

LEMMA 2.1. *Let* $\tilde{X}(j) = \sum_{k=0}^{m(n)-1} \rho_k \varepsilon(j-k)$ *with* $m(n) = [\tilde{c}\log n]$ *and* $\tilde{Y}(j) = \tilde{X}(j) + T(j)e(j)$. *Under Conditions* A–D *we have*

$$\max_{1\le j\le n} |Y(j) - \tilde{Y}(j)| \le cZn^{-\tau_1}$$

where τ_1 *may be made arbitrarily large by increasing* $\tilde{c}$ *and* Z *is a random variable that is a.s. finite with* $P(Z \ge z) \le cz^{-\alpha}$ *for all* $z \in \mathbb{R}_+$, α *being defined in Condition* A. *In addition*

$$\sup_{x\in\mathbb{R}} |F(x) - \tilde{F}(x)| = O(n^{-\alpha\tau_1/(\alpha+1)}) \qquad a.s.$$

where $F(x)$, $\tilde{F}(x)$ *denote the distribution functions of* $X(j)$ *and* $\tilde{X}(j)$, *respectively.*

PROOF. The proof is similar to the one in Hesse (1987), Lemma 2 and is therefore omitted here.

LEMMA 2.2. *Let* $\tilde{X}(j) = \sum_{k=0}^{m(n)-1} \rho_k \varepsilon(j-k)$ *with* $m(n) = [\tilde{c}\log n]$ *and* $\tilde{Y}(j) = \tilde{X}(j) + T(j)e(j)$. *Also, define*

$$g_n^*(x) = \frac{1}{2\pi\lambda(d_n)} \int_{-1}^{+1} e^{-itx\lambda^{-1}(d_n)} \hat{\Psi}_{\tilde{Y}}(t\lambda^{-1}(d_n)) \frac{\Psi_W(t)}{\Psi_{Te}(t\lambda^{-1}(d_n))} dt$$

as an "estimator" of the density $\tilde{g}$ *of* $\tilde{X}(j)$ *where* $\hat{\Psi}_{\tilde{Y}}$ *is the empirical characteristic function of* $\tilde{Y}(1), \ldots, \tilde{Y}(n)$.

Then under Conditions A–E

$$\text{(a)} \quad \sup_{x\in\mathbb{R}} |g_n^*(x) - \hat{g}_n(x)| = O(n^{-\tau_2})$$

$$\text{(b)} \quad \sup_{x\in\mathbb{R}} |\tilde{g}(x) - g(x)| = O(n^{-\tau_3})$$

where $\tau_i(>0)$ may be made arbitrarily large by increasing $\tilde{c}$ (see definition of $\tilde{Y}(j)$) and $\hat{g}_n(x)$ is as in (1.4).

PROOF. (a) By Lemma 2.1 for $\tilde{c}$ sufficiently large

$$\begin{aligned}\max_{1\leq j\leq n} |\exp(it((Y(j)-x)\lambda^{-1}(d_n))) - \exp(it((\tilde{Y}(j)-x)\lambda^{-1}(d_n)))| \\ \leq \max_{1\leq j\leq n}(ct|Y(j)-\tilde{Y}(j)|\lambda^{-1}(d_n)) \\ \leq ctZn^{-c}\end{aligned}$$

and hence

$$\begin{aligned}\sup_{x\in\mathbb{R}}\Bigg(\frac{1}{2\pi\lambda(d_n)}\int_{-1}^{+1} |\exp(it((Y(j)-x)\lambda^{-1}(d_n))) \\ - \exp(it((\tilde{Y}(j)-x)\lambda^{-1}(d_n)))|\frac{|\Psi_W(t)|}{|\Psi_{Te}(t\lambda^{-1}(d_n))|}dt\Bigg) \\ \leq c\lambda^{-1}(d_n)Zn^{-c}\int_{-1}^{+1}|t\Psi_W(t)|dt = O(n^{-\tau_2}) \qquad \text{a.s.}\end{aligned}$$

(b) We know that

$$\left|\frac{F(x+h)-F(x)}{h} - \frac{\tilde{F}(x+h)-\tilde{F}(x)}{h}\right| \xrightarrow{h\to 0} |g(x)-\tilde{g}(x)|$$

and also

$$\begin{aligned}\sup_{x\in\mathbb{R}}\left|\frac{F(x+h)-F(x)}{h} - \frac{\tilde{F}(x+h)-\tilde{F}(x)}{h}\right| \\ \leq \sup_{x\in\mathbb{R}}\left(\frac{|F(x+h)-\tilde{F}(x+h)|}{h} + \frac{|\tilde{F}(x)-F(x)|}{h}\right).\end{aligned}$$

With $h = n^{-\gamma}$ for some $\gamma > 0$ this is, by Lemma 2.1, of the order

$$n^{\gamma}O(n^{-\alpha\tau_1/(\alpha+1)}) = O(n^{-\alpha\tau_1/(\alpha+1)+\gamma}) =: O(n^{-\tau_2}).$$

LEMMA 2.3. *The functions* ${}^*W_n^{x_j}$ *and* ${}_*W_n^{x_j}$ *defined in* (2.5)–(2.8) *have the following properties*:

(a) *For all* $W_n^x(y) = n^{(1+\delta)/2}(\log n)^{-\delta/2-1}W_n((y-x)\lambda^{-1}(cn/\log n)) \in \tilde{W}_n$ *there exists an index* $j\in J$ *such that for all* $y\in\mathbb{R}$

$$|W_n^x(y) - {}^*W_n^{x_j}(y)| \leq |{}^*W_n^{x_j}(y) - {}_*W_n^{x_j}(y)|.$$

(b) *For all* $j\in J$ *and* $\varepsilon \geq 1$,

$$E(|{}^*W_n^{x_j}(Y(1)) - {}_*W_n^{x_j}(Y(1))|) \leq \varepsilon.$$

PROOF. The proof of these results is straightforward and is therefore omitted.

Remark 1. The result of Theorem 2.1 can be extended beyond the context of partial contamination. Subject to E it holds more generally for any error distribution whose characteristic function is twice continuously differentiable and uniformly bounded away from zero in absolute value. The class thus specified is rich and includes elementary examples such as the distributions with characteristic functions $p+(1-p)e^{-t^2/2}$, $p+(1-p)(1+t^2)^{-1}$ (i.e. partially contaminated normal, Laplace, respectively, for any amount of contamination $(1-p)\in(0,1)$), and in general any partially contaminated distribution if that distribution's characteristic function is real and non-negative. Also included are the Bernoulli$(1-p)$ distribution with $p\neq 1/2$ and the Poisson(λ) distribution. For the latter this is so since clearly

$$|\exp(\lambda(\exp(it)-1))| = \exp(\mathrm{Re}[\lambda(\exp(it)-1)]) = \exp(\lambda(\cos t-1)) \geq e^{-2\lambda}$$

where Re$[x]$ denotes the real part of x. Additional and less elementary examples are provided by the distribution of a compound Poisson random variable $Z=\sum_{i=1}^{m} Z_i$ for appropriate choices of the distribution of the iid Z_i random variables (m is a Poisson(λ) random variable) as well as by the distributions corresponding to the characteristic functions

$$\Psi_1(t) = \exp[\exp(-\alpha t^2)-1]$$
$$\Psi_2(t) = \exp[\exp[\lambda\exp(it)]-\exp\lambda].$$

Remark 2. The result of Theorem 2.1 is valid if the amount of contamination is known. If $p\in[0,1]$ is also unknown then the question of identifiability re-emerges with respect to p. Since the density g to be estimated is unknown the parameter p is identifiable for a given error density f if for the characteristic functions Ψ_{g_1}, Ψ_{g_2} and Ψ_f of densities g_1, g_2 and f, respectively,

$$\Psi_{g_1}(t)[p_1+(1-p_1)\Psi_f(t)] = \Psi_{g_2}(t)[p_2+(1-p_2)\Psi_f(t)]\forall t$$

implies $p_1=p_2$ for all smooth g_1 and g_2 satisfying the conditions of Theorem 2.1. However, with

$$\Psi_f(t) = \exp(-t^2/2), \qquad p_1\in(0,1),\ \ p_2=1$$
$$\Psi_{g_1}(t) = \exp(-t^2/2)$$
$$g_2(x) = p_1\varphi(x)+(1-p_1)2^{-1/2}\varphi(2^{-1/2}x)$$

the above implication does not hold, where $\varphi(x)$ is the standard normal density.

Remark 3. The rate given in Theorem 2.1 for the contamination and dependence case compares well with existing results for ordinary density estimation from iid uncontaminated observations. For a second order kernel with bounded support the relevant rates are $o((n^{-1}\log n)^{2/5} M_n)$ a.s. where $M_n \le \log\log n$ and $M_n \to \infty$ (Karunamuni and Mehra (1990)).

It is interesting to compare deconvolution of partially contaminated linear processes with deconvolution of fully contaminated linear processes, i.e. with the situation arising for $p = 0$ in the above model. Towards this end we first present a theorem for the empirical characteristic function of

$$Y(j) = \sum_{k=0}^{\infty} \rho_k \varepsilon(j-k) + e(j), \qquad j = 1, \ldots, n.$$

THEOREM 2.2. *Under Conditions* A, B *and* E *but with* $p = 0$ *in Condition* A *we have*

$$P\left(\limsup_{n\to\infty} n^{1/2}(\log n)^{-1} \sup_{|t|\le n^{\theta}} |\hat{\Psi}_Y(t) - \Psi_Y(t)| < \infty\right) = 1$$

where θ *is an arbitrary positive constant.*

PROOF. Because of Hesse (1990) Propositions 2 and 3 with $g(i) = c\rho^i$, $\rho \in (0,1)$ and $h(n) = [\tilde{c}\log n] = m(n)$ it suffices to obtain the stated rate $O(n^{-1/2}\log n)$ for convergence of the characteristic function $\Psi_{\tilde{Y}}$ of

$$\tilde{Y}(j) = \sum_{k=0}^{[\tilde{c}\log n]} \rho_k \varepsilon(j-k) + e(j), \qquad j = 1, \ldots, n.$$

With

$$\hat{\Psi}^j_{\tilde{Y}}(t) = \frac{1}{n_j}\sum_{k=0}^{n_j - 1} \exp(it(\tilde{Y}(j + km(n))))$$

we have

$$\sup_{|t|\le (m(n))^{\theta}} |\hat{\Psi}_{\tilde{Y}}(t) - \Psi_{\tilde{Y}}(t)| \le \max_{1\le j\le m(n)} \sup_{|t|\le (m(n))^{\theta}} |\hat{\Psi}^j_{\tilde{Y}}(t) - \Psi_{\tilde{Y}}(t)|$$

and by the methods of Csörgö (1985)

$$P\left(n^{1/2}(\log n)^{-1} \sup_{|t|\le (m(n))^{\theta}} |\hat{\Psi}^j_{\tilde{Y}}(t) - \Psi_{\tilde{Y}}(t)| > c\right)$$
$$\le c\exp(-c_1 \log n) + K_n n^{\theta+1/2}(\log n)^{-1-\theta}\exp(-c_2\log n)$$

where c_1, c_2 may be made arbitrarily large, $\theta > 0$ is arbitrary, and

$$K_n = \inf\{x > 0 : P(|\tilde{Y}(1)| > x) \le n^{-1/2}\log n\}.$$

Therefore by appropriate choices of c_1 and c_2

$$\sum_{n=n_0}^{\infty} P\left(\max_{1\le j\le m(n)} \sup_{|t|\le(m(n))^\theta} |\hat{\Psi}^j_{\tilde{Y}}(t) - \Psi_{\tilde{Y}}(t)| > c\right)$$
$$\le c\sum \log n(n^{-c_1} + K_n n^{\theta+1/2-c_2}(\log n)^{-1-\theta}) < \infty$$

since K_n is of polynomial order. With this theorem we may now obtain deconvolution rates for the model (1.3) with $p = 0$.

THEOREM 2.3. *If Conditions* A (*with* $p = 0$), B (*with "two" replaced by* s), E *hold*, $\Psi_W(t)$ *is even, real-valued and nonincreasing on* $[0,\infty)$ *with* $\max(r,s)+1$ *bounded derivatives such that* $\Psi_W(1) = \cdots = \Psi_W^{(r-1)}(1) = 0$, $\Psi_W^{(r)}(1) \neq 0$ *and* $\Psi_W^{(1)}(0) = 0 = \Psi_W^{(2)}(0) = \cdots = \Psi_W^{(s-1)}(0)$, $\Psi_W^{(s)}(0) \neq 0$ *and in addition* $\Psi_e(t)$ *is real-valued, non-vanishing with* $(\Psi_e(t))^{-1} \sim \alpha t^\beta \exp(\gamma t^\xi)$ *as* $t \to \infty$ *for constants* $\alpha, \xi, \gamma > 0$, $|\beta| < \infty$ *then*

$$P\left(\limsup_{n\to\infty}(\log n)^{s/\xi} \sup_{x\in\mathbb{R}} |\hat{g}_n(x) - g(x)| < \infty\right) = 1$$

where $\hat{g}_n(x)$ *is as in* (1.4) *with* $\lambda = \lambda(d_n) = (2\gamma/\log n)^{1/\xi}$.

PROOF. We have

$$\sup_{x\in\mathbb{R}} |\hat{g}_n(x) - g(x)| \le \sup_{x\in\mathbb{R}} |\hat{g}_n(x) - E\hat{g}_n(x)| + \sup_{x\in\mathbb{R}} |E\hat{g}_n(x) - g(x)|$$
$$\le \sup_{x\in\mathbb{R}} \left|\frac{1}{2\pi}\int_{-\lambda^{-1}}^{+\lambda^{-1}} e^{-itx}(\hat{\Psi}_Y(t) - \Psi_Y(t))\frac{\Psi_W(t\lambda)}{\Psi_e(t)}\right| dt + c\lambda^s \qquad \text{a.s.}$$
$$\le c \sup_{|t|\le\lambda^{-1}} |\hat{\Psi}_Y(t) - \Psi_Y(t)| \int_0^{\lambda^{-1}} \frac{\Psi_W(t\lambda)}{\Psi_e(t)} dt + c\lambda^s \qquad \text{a.s.}$$
$$\le cn^{-1/2}(\log n)\lambda^{(r+1)\xi-\beta-1}\exp(\gamma\lambda^{-\xi}) + c\lambda^s \qquad \text{a.s.}$$

by Lemmas 3.1–3.3 of Stefanski (1990) and by our Theorem 2.2. Now the optimal choice $\lambda = (2\gamma/\log n)^{1/\xi}$ produces the result.

In the normal case with $\xi = 2$ this theorem reproduces the optimal rate of order $(\log n)^{-s/2}$ obtained by Carroll and Hall (1988) for iid observations with normal contamination errors. We have shown that this rate of order $O((\log n)^{-s/\xi})$ is still valid if observations have the dependence structure of a linear process. If observations are merely partially contaminated then the results of the paper show that deconvolution rates improve in a significant fashion from logarithmic to polynomial order and get close to the best rates available for iid uncontaminated observations.

Acknowledgements

Part of the research was conducted while visiting the Center for Mathematics and its Applications at the Australian National University. I thank Peter Hall for the invitation and the Center for its hospitality. I would also like to thank two anonymous referees for their thoughtful comments.

References

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

Carroll, R. J., Spiegelman, C. H., Lan, K. K., Bailey, K. R. and Abbott, R. D. (1984). On errors-in-variables for binary regression models, *Biometrika*, **71**, 19–25.

Crump, J. G. and Seinfeld, J. H. (1982). A new algorithm for inversion of aerosol size distribution data, *Aerosol Science and Technology*, **1**, 15–34.

Csörgö, S. (1985). Rates of uniform convergence for the empirical characteristic function, *Acta Sci. Math.*, **48**, 97–102.

Dudley, R. M. (1978). Central limit theorems for empirical measures, *Ann. Probab.*, **6**, 899–929 (Correction: ibid. (1978), **7**, 909–911).

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

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

Fan, J. (1992). Deconvolution with supersmooth distributions, *Canad. J. Statist.*, **20**, 155–169.

Fan, J. and Masry, E. (1992). Multivariate regression estimation with errors-in variables: Asymptotic normality for mixing processes, *J. Multivariate Anal.*, **43**, 237–271.

Fan, J. and Troung, Y. (1993). Nonparametric regression with errors-in-variables, *Ann. Statist.*, **21**, 1900–1925.

Fan, J., Troung, Y. and Wang, Y. (1990). Measurement errors regression: A nonparametric approach, Institute of Statistics Mimeo Series, No. 2042, University of North Carolina, Chapel Hill.

Hesse, C. H. (1987). Limit theorems for linear processes and applications, Ph.D. Dissertation, Dept. of Statistics, Harvard University.

Hesse, C. H. (1990). Rates of convergence for the empirical distribution function and the empirical characteristic function of a broad class of linear processes, *J. Multivariate Anal.*, **35**, 186–202.

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Karunamuni, R. J. and Mehra, K. L. (1990). Improvements on strong uniform consistency of some known kernel estimates of a density and its derivatives, *Statist. Probab. Lett.*, **9**, 133–140.

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

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

Masry, E. (1991*b*). Regression estimation with errors-in-variables for stationary processes, Tech. Report, **173**, Dept. of Electrical and Computer Engineering, University of California, San Diego.

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

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

Masry, E. and Rice, J. A. (1992). Gaussian deconvolution via differentiation, *Cand. J. Statist.*, **20**, 9–21.

Mendelsohn, J. and Rice, J. (1982). Deconvolution of microfluorometric histograms with *B*-splines, *J. Amer. Statist. Assoc.*, **77**, 748–753.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems, *Ann. Math. Statist.*, **35**, 1–20.

Snyder, D. L., Miller, M. I. and Schultz, T. J. (1988). Constrained probability density estimation from noisy data, *Proceedings of the 1988 Conference on Information Sciences and Systems*, 170–172, Johns Hopkins University, Baltimore, Maryland.

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

Stefanski, L. A. and Carroll, R. J. (1990). Deconvolving kernel density estimators, *Statistics*, **21**, 169–184.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions, *Contributions to Probability and Statistics* (ed. I. Olkin), Stanford University Press, Stanford.

Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions, *Ann. Statist.*, **18**(2), 806–831.