Statistica Sinica **20** (2010), 1609-1636

DECONVOLUTION FROM NON-STANDARD ERROR DENSITIES UNDER REPLICATED MEASUREMENTS

Alexander Meister and Michael H. Neumann

Universität Rostock and Friedrich-Schiller-Universität Jena

Abstract: We propose a nonparametric density estimator based on data that are repeatedly observed with independent measurement errors. We particularly focus on cases where the Fourier transform of the error density has some zeros and shows oscillations. Our estimator attains the same rates of convergence as obtains under smooth error densities whose Fourier transform have the corresponding tails but no zeros. We prove minimax results for estimating the distribution function and for support estimation in the same model. A simulation study supports our findings.

Key words and phrases: Analysis of variance, characteristic function, components of variance, deconvolution, nonparametric density estimation, panel data, repeated measurements.

1. Model and Assumptions

Nonparametric density deconvolution is the problem of estimating a density function f_X, or the corresponding distribution, based on contaminated data under nonparametric conditions on f_X. In the basic model, i.i.d. observations $Y_1, \ldots, Y_n$, generated by $Y_j = X_j + \varepsilon_j$, are given. All $X_1, \varepsilon_1, \ldots, X_n, \varepsilon_n$ are independent; the ε_j represent the measurement errors and have the known density f_ε; the X_j have the density of interest f_X. Kernel methods for the deconvolution problem were introduced in Carroll and Hall (1988) and Stefanski and Carroll (1990).

A common condition on the error density f_ε is $f_\varepsilon^{ft}(t) \neq 0$ for all $t \in \mathbb{R}$, where g^{ft} denotes the Fourier transform of a generic function g. Also, one usually imposes upper and lower bounds on f_ε^{ft} whose ratio is independent of t. In the case of polynomially decaying f_ε^{ft}, the error density is called ordinary smooth; densities with exponentially decaying Fourier transform are called supersmooth; see Fan (1991, 1993). The Laplace density is an example of an ordinary smooth density whereas the normal and Cauchy densities are examples of supersmooth densities. We refer to both ordinary smooth and supersmooth f_ε as standard error densities. Here, we focus on error densities whose Fourier transforms have some zeros and show oscillatory or irregular behaviour. To better illustrate these ideas we provide four examples of non-standard error distributions.

(a) Uniform (continuous) distribution.

The density is

$$f_\varepsilon(x;a,b) = \begin{cases} \frac{1}{b-a}, & \text{if } a \le x \le b, \\ 0, & \text{otherwise,} \end{cases}$$

while the characteristic function, f_ε^{ft}, satisfies

$$f_\varepsilon^{ft}(t) = \frac{\exp(ibt) - \exp(iat)}{it(b-a)} = \frac{2\ \exp\{i(a+b)/2\}\ \sin\{(b-a)t/2\}}{t(b-a)}.$$

(b) Uniform (discrete) distribution.

This distribution has a probability function

$$p(k;j,l) = \begin{cases} \frac{1}{l}, & \text{if } k = j, j+1, \ldots, j+l-1, \\ 0, & \text{for other } k\,, \end{cases}$$

and the characteristic function

$$f_\varepsilon^{ft}(t) = \frac{\exp[it\{j + (l-1)/2\}]\ \sin(lt/2)}{l\ \sin(t/2)}.$$

(c) Ordinary smooth densities (e.g., Laplace density) convolved with a (continuous or discrete) uniform distribution.

(d) A density with characteristic function vanishing on some intervals.

The density

$$f_\varepsilon(x) = \frac{1-\cos(x)}{\pi x^2} \sum_{j=0}^{\infty} (1 + c|j|)^{-\alpha} \cos(\frac{5jx}{2})\,,$$

where $\alpha > 1$ and $c > 0$ is sufficiently large, has the Fourier transform

$$f_\varepsilon^{ft}(t) = \frac{1}{2}(1-|t|)_+ + \frac{1}{2} \sum_{j=-\infty}^{\infty} (1+c|j|)^{-\alpha}\ (1 - |t + \frac{5j}{2}|)_+$$

which vanishes on the set $\bigcup_{j=1}^{\infty} [5j/2 - 3/2, 5j/2 - 1] \cup [1 - 5j/2, 3/2 - 5j/2]$.

Uniform measurement error densities appear in astronomy. Sun et al. (2003) describe an experiment where contaminated measurements on the velocity of halo stars in the Milky Way are obtained. In some cases, the measurement error is uniform if the contamination is caused by such effects as mechanical stiffness of the spectrograph.

Nonparametric deconvolution estimators are usually obtained by a kernel method, where the deconvolution step is carried out in the spectral domain via a division by f_ε^{ft}. Some alternative approaches are based on splines, e.g., Mendelsohn and Rice (1982) and Carroll, Maca and Ruppert (1999), and on SIMEX methods, e.g., Carroll, Maca and Ruppert (1999), but here we choose to work with an inverse-Fourier-type estimator.

In Johnstone *et al.* (2004) and Kerkyacharian, Picard and Raimondo (2007), boxcar deconvolution problems, i.e. uniform deconvolution in a periodic setting, are considered in a white noise model. The authors derive nearly optimal rates under certain conditions on the ratio of the scaling parameter of the error density and the length of true signal support. Note that the periodic framework in these papers is significantly different from the usual deconvolution setting in density estimation. In the usual density estimation based on additively corrupted data, it has been shown by Devroye (1989) that consistent estimation of f_X is possible if the set $N_\varepsilon = \{t > 0 : f_\varepsilon^{ft}(t) = 0\}$ has Lebesgue measure zero. Under the even weaker condition that no open, non-void interval is included in N_ε, Meister (2005) establishes consistency with respect to a specific topology on the densities. However, if N_ε contains some open, non-empty intervals, as in example (d), then f_X is not identifiable (see Devroye (1989)), from which it follows that there is no consistent estimator of f_X in example (d). Hall and Meister (2007) consider error densities when f_ε^{ft} has some isolated and periodic zeros (example (a) and (b)). They show that the optimal convergence rates as derived in Fan (1991, 1993) for standard error densities remain valid for supersmooth f_ε; however, for ordinary smooth f_ε, these rates can be ensured only in rare cases.

The situation changes dramatically if repeated measurements, as in analysis of variance models, are available. In the current paper, we consider the replicated measurement model, where each X_j is observed at least twice with different errors. Therefore, the data $Y_{j,k}$, $j \in \{1, \ldots, n\}$, $k \in \{1, 2\}$, with

$$Y_{j,k} = X_j + \varepsilon_{j,k} \,, \tag{1.1}$$

are available, where the $\varepsilon_{j,k}$ have the density f_ε and all $X_j, \varepsilon_{j,k}$ are independent. The replicated measurement model has been studied by Horowitz and Markatou (1996), Li and Vuong (1998), Hall and Yao (2003), Schennach (2004a,b), Neumann (2007), Delaigle, Hall and Meister (2008), where it has been shown that f_X can be consistently estimated in some cases even if f_ε is not known in advance. A similar problem is studied in Neumann (1997) and Efromovich (1997), where the error density is estimable from additional direct observations; this is a major advantage in comparison to the basic deconvolution model where prior knowledge of the error distribution is essential.

We show here that optimal rates of convergence are unaffected by oscillations of f_ε^{ft} under model (1.1). For all f_ε that satisfy (1.2), we obtain the same convergence rates as under ordinary smooth error densities whose Fourier transforms have the corresponding tails, but no zeros. This phenomenon does not occur in the case without any replications, rather the convergence rates usually show deterioration under error densities with isolated zeros in the Fourier domain (example (a)); see Hall and Meister (2007). However, our estimator even attains these rates for error densities such as example (d); inconsistency occurs here in the basic deconvolution model without any replications. We further extend this theory to the problem of estimating the probability of an interval and the support of f_X in model (1.1).

To provide some intuition about the empirical information contained in the data, we consider $(Y_{j,1}, Y_{j,2})'$ and its characteristic function $f^{ft}_{(Y_{j,1},Y_{j,2})}(t_1,t_2) = f^{ft}_X(t_1+t_2)f^{ft}_\varepsilon(t_1)f^{ft}_\varepsilon(t_2)$; the information about $f^{ft}_X(t)$ can be obtained from $f^{ft}_{(Y_{j,1},Y_{j,2})}(\tau, t-\tau)$ if $f^{ft}_\varepsilon(\tau)$ and $f^{ft}_\varepsilon(t-\tau)$ are both nonzero. There is an important difference between the replicated measurement setting and the standard deconvolution model as the parameter τ can be selected for any t. In the Fourier domain, deconvolving the density is carried out by dividing the empirically accessible function $f^{ft}_{(Y_{j,1},Y_{j,2})}(\tau, t-\tau)$ by $f^{ft}_\varepsilon(\tau)f^{ft}_\varepsilon(t-\tau)$. Intuitively, the accuracy of the estimator increases when $|f^{ft}_\varepsilon(\tau)f^{ft}_\varepsilon(t-\tau)|$ becomes larger. Therefore, it seems advantageous to select τ so that the latter term is maximized. Division by zero can also be avoided by an appropriate choice of τ.

To derive a general framework for studying unconventional error distributions we assume that

$$\rho(t) \geq C(1+|t|)^{-\alpha}, \qquad \text{for some } C>0,\ \ \alpha \geq 0\,, \tag{1.2}$$

where $\rho(t) := \sup_\tau \left|f^{ft}_\varepsilon(\tau)f^{ft}_\varepsilon(t-\tau)\right|$. In examples (a)–(d), (1.2) is satisfied; more precisely, in example (a), we have $\alpha = 1$; in example (b), we have $\alpha = 0$; in example (c), α is equal to the smoothness degree of the ordinary smooth density plus 1 or 0 for continuous and discrete uniform distributions, respectively; in example (d), α appears in the definition of the error density. Note that (1.2) is implied by the existence of an increasing sequence $(T_k)_k \uparrow \infty$ with $T_0 = 0$ and $T_{k+1} - T_k < \delta S_\varepsilon$, for all k with a fixed $\delta \in (0,1)$ and $S_\varepsilon = \min\{t \geq 0 \,:\, f^{ft}_\varepsilon(t) = 0\}$, so that

$$|f^{ft}_\varepsilon(T_k)| \geq C_\varepsilon\big(1+T_k\big)^{-\alpha} \tag{1.3}$$

holds for all integers $k \geq 0$. When f^{ft}_ε has no zeros we set $S_\varepsilon = \infty$. It can be shown that f_ε as in (d) satisfies (1.3) with $S_\varepsilon = 1/2$ and $T_k = 5k/2$; thus, it satisfies condition (1.2). These conditions are significantly weaker than the usual version of ordinary smooth densities in Fan (1991), where the inequality in (1.3) is assumed to be valid on the whole real line instead on the discrete set

$\{T_k\}$. Ordinary smooth error densities with parameter α are thus included in our framework.

2. Density Estimation

On the basis of the observations $(Y_{1,1}, Y_{1,2})', \ldots, (Y_{n,1}, Y_{n,2})'$ obeying model (1.1), we take the estimator of the lth derivative of f_X ($l \in \mathbb{N} \cup \{0\}$) to be

$$\widehat{f}_{X,n}^{(l)}(x) = \mathrm{Re}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)(-it)^l K^{ft}(th_n)\frac{\widehat{\varphi}_{Y,n}\{\tau(t), t-\tau(t)\}}{f_\varepsilon^{ft}\{\tau(t)\}f_\varepsilon^{ft}\{t-\tau(t)\}}dt\right), \tag{2.1}$$

where $\widehat{\varphi}_{Y,n}(s,t) = n^{-1}\sum_{j=1}^n \exp(isY_{j,1} + itY_{j,2})$, K is a kernel function, $h = h_n > 0$ is a bandwidth parameter, and τ is a function which satisfies $|f_\varepsilon^{ft}\{\tau(t)\}f_\varepsilon^{ft}\{t-\tau(t)\}| \geq \rho(t)/2$, and ρ as defined at (1.2). We recall that K^{ft} denotes the Fourier transform of K and, to simplify notation, we write $\widehat{f}_{X,n}(x) := \widehat{f}_{X,n}^{(0)}(x)$.

If one is sure that this supremum is also a maximum, e.g., if $|f_\varepsilon^{ft}(t)| \to 0$ as $|t| \to \infty$, then one can select τ such that

$$|f_\varepsilon^{ft}\{\tau(t)\}f_\varepsilon^{ft}\{t-\tau(t)\}| = \sup_s |f_\varepsilon^{ft}(s)f_\varepsilon^{ft}(t-s)| = \rho(t). \tag{2.2}$$

In examples (a) to (d), $\tau(t)$ in (2.2) has to be determined numerically, as no explicit formula is available. However, an appropriate grid search can often be restricted to certain compact intervals. In example (a), define $g_t(s) = |\sin\{(b-a)s/2\}\sin\{(b-a)(t-s)/2\}|$ and $h_t(s) = |s(t-s)|$. By definition, $\tau(t)$ is the maximizer of $g_t(s)/h_t(s)$. The target function g_t/h_t is symmetric about $t/2$, the function g_t is $2\pi/(b-a)$-periodic and, for $t > 0$, h_t is monotonously decreasing on $[t/2, t]$ and increasing on $[t, \infty)$. For $t > 0$, it suffices to search for $\tau(t)$ on the interval $[\max\{t/2, t-2\pi/(b-a)\}, t+2\pi/(b-a)]$.

In order to establish rates of convergence for the pointwise risk (MSE) at a specific point $x_0 \in \mathbb{R}$, local smoothness assumptions in x_0 are required; they are given by

$$\begin{aligned} f_X \in \mathcal{H}(x_0, \beta, L) := \big\{&f : f \text{ is a density with } \lfloor\beta\rfloor \text{ derivatives and} \\ &\left|f^{(\lfloor\beta\rfloor)}(x_0) - f^{(\lfloor\beta\rfloor)}(y)\right| \leq L|x_0 - y|^{\beta-\lfloor\beta\rfloor}, \text{ for all } y \in \mathbb{R}\big\}, \end{aligned}$$

where $\lfloor\beta\rfloor$ is the largest integer strictly less than β. Of course, we have to assume that l in (2.1) does not exceed $\lfloor\beta\rfloor$. Then, we have a result about upper bounds on the MSE.

Theorem 2.1. *Let K be a kernel of order $\geq \lfloor\beta\rfloor - l$, $l \in \{0, \ldots, \lfloor\beta\rfloor\}$ with*

$$\int_{-\infty}^{\infty} |K^{ft}(t)|^2|t|^{2\alpha+2l}dt < \infty.$$

Under (1.2) *with* $h_n \asymp n^{-1/(2\beta+2\alpha+1)}$, *we have*

$$\sup_{f_X \in \mathcal{H}(x_0,\beta,L)} \mathbb{E}_{f_X} \left| \widehat{f}^{(l)}_{X,n}(x_0) - f^{(l)}_X(x_0) \right|^2 = O\left\{ n^{-2(\beta-l)/(2\beta+2\alpha+1)} \right\}.$$

These rates coincide with those derived for ordinary smooth error densities with the Fourier-tail behaviour described in Fan (1991). Therefore, the repeated measurement model (1.1) allows one to relax the assumption of ordinary smooth f_ε to (1.2), while keeping the convergence rates.

This result can be extended to $L_2(\mathbb{R})$-risk (also known as MISE) when local smoothness class $\mathcal{H}(x_0,\beta,L)$ is changed to the Sobolev class $\mathcal{S}(\beta,L)$, where

$$f_X \in \mathcal{S}(\beta,L) := \Big\{ f : f \text{ is a density and } \int_{-\infty}^{\infty} (1+|t|)^{2\beta} |f^{ft}(t)|^2 dt \le L \Big\},$$

$\beta > 0$. (In the case of integer β this is a usual Sobolev class; otherwise we call it generalized Sobolev class.)

As an analogue to Theorem 2.1, we have the following result.

Theorem 2.2. *Suppose* $\sup_{t\neq 0}\{|K^{ft}(t)-1|\,|t|^{l-\beta}\} < \infty$, $l \in \{0,\ldots,\lfloor\beta\rfloor\}$, *and*

$$\int_{-\infty}^{\infty} |K^{ft}(t)|^2 |t|^{2\alpha+2l} dt < \infty.$$

Under (1.2) *with* $h_n \asymp n^{-1/(2\beta+2\alpha+1)}$, *we have*

$$\sup_{f_X \in \mathcal{S}(\beta,L)} \mathbb{E}_{f_X} \|\widehat{f}^{(l)}_{X,n} - f^{(l)}_X\|^2_{L_2(\mathbb{R})} = O\left\{ n^{-2(\beta-l)/(2\beta+2\alpha+1)} \right\}.$$

This theorem yields coincidence of the rates when considering the MISE as well, see e.g., Fan (1993) for a similar integrated risk under ordinary smooth f_ε.

As a fully data-driven choice of the bandwidth h, cross-validation methods may be considered, involving minimization of the empirically accessible

$$\mathrm{CV}(h) = \int \left| \widehat{f}_{X,n}(x;h) \right|^2 dx - \frac{1}{\pi} \mathrm{Re} \int K^{ft}(th) |\rho(t)|^{-2} \widehat{\psi}_{Y,n}\{\tau(t), t-\tau(t)\} dt, \quad (2.3)$$

where $\widehat{\psi}_{Y,n}(s,t) := \{n(n-1)\}^{-1} \sum\sum_{j\neq k} \exp\big\{ is(Y_{j,1}-Y_{k,1}) + it(Y_{j,2}-Y_{k,2}) \big\}$. Cross-validation has been proposed in density deconvolution without replications by Stefanski and Carroll (1990); a rigorous study is provided in Hesse (1999).

Finally, we mention that our method is also applicable to the supersmooth case of densities with exponential Fourier tails. Our method also attains the optimal logarithmic convergence rates; and it addresses those f_ε, that are convolutions of supersmooth densities and uniform densities or, to be more general, densities with exponential Fourier tails and zeros in the Fourier domain. The ridge-parameter estimator of Hall and Meister (2007) also attains these logarithmic rates, and without replicates.

If a different number of measurements is given for each observation, say k_j measurements for the jth observation, we may generalize the definition of estimator (2.1). First, given the data $Y_{j,k}$, $j = 1, \ldots, n$, $k = 1, \ldots, k_j$, we set

$$\tilde{\rho}_j(t) = \sup_{\tau_1 + \cdots + \tau_{k_j} = t} \Big| \prod_{k=1}^{k_j} f_\varepsilon^{ft}(\tau_k) \Big|.$$

Analogously, we choose a function $\tilde{\tau}_j(t) = \big(\tilde{\tau}_{j,1}(t), \ldots, \tilde{\tau}_{j,k_j}(t)\big)$ such that $\sum_{k=1}^{k_j} \tilde{\tau}_{j,k}(t) = t$ and

$$\Big| \prod_{k=1}^{k_j} f_\varepsilon^{ft}(\tilde{\tau}_{j,k}) \Big| \geq \frac{\tilde{\rho}_j(t)}{2}.$$

Then, instead of estimator (2.1), we suggest

$$\widetilde{f}_X^{(l)}(x) = \mathrm{Re}\Big(\frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)(-it)^l K^{ft}(th_n) \Big[\sum_{j=1}^{n} \Big| \prod_{k=1}^{k_j} f_\varepsilon^{ft}\{\tilde{\tau}_{j,k}(t)\} \Big|^2 \Big]^{-1}$$
$$\sum_{j=1}^{n} \exp\Big\{ i \sum_{k=1}^{k_j} \tilde{\tau}_{j,k}(t) Y_{j,k} \Big\} \prod_{k=1}^{k_j} f_\varepsilon^{ft}\{-\tilde{\tau}_{j,k}(t)\} dt \Big)$$

as the estimator of $f_X^{(l)}(x)$. This estimator is inspired by the heteroscedastic estimator introduced in Delaigle and Meister (2008). It is still well-defined if we have just one measurement ($k_j = 1$) for some of the data. The investigation of the asymptotic properties of this procedure is left open for future research.

3. Distribution Estimation

In this section, we study the asymptotic risk for estimating

$$P_{a,b} = \int_a^b f_X(x) dx$$

for some arbitrary but fixed $a < b$. The related problem of estimating the cumulative distribution function was studied in Zhang (1990), Fan (1991), and Hall and Lahiri (2008) for ordinary smooth f_ε. Considering $P_{a,b}$ as the $L_2(\mathbb{R})$-inner product of f_X and the indicator function $\mathbb{1}_{[a,b]}$ we obtain by the Plancherel isometry that

$$\int f(x)\overline{g(x)} dx = \frac{1}{2\pi} \int f^{ft}(t) \overline{g^{ft}(t)} dt$$

for all $f, g \in L_2(\mathbb{R})$ (see e.g., Yosida (1968)), and that

$$P_{a,b} = \int \mathbb{I}_{[a,b]}(x) f_X(x) dx = \frac{1}{2\pi} \int \Psi_{a,b}(-t) f_X^{ft}(t) dt,$$

where $\Psi_{a,b}(t) = \mathbb{I}_{[a,b]}^{ft} = \big\{\exp(ibt) - \exp(iat)\big\}/(it)$. This motivates an estimator for $P_{a,b}$ as

$$\widehat{P}_{a,b,n} = \frac{1}{2\pi} \int \Psi_{a,b}(-t) K^{ft}(th_n) \frac{\widehat{\varphi}_{Y,n}\{\tau(t), t-\tau(t)\}}{f_\varepsilon^{ft}\{\tau(t)\}\, f_\varepsilon^{ft}\{t-\tau(t)\}} dt. \tag{3.1}$$

Its asymptotic performance is studied in the next theorem. Here we need global Hölder smoothness, setting

$$\mathcal{H}(\beta, L) = \bigcap_{x \in \mathbb{R}} \mathcal{H}(x, \beta, L).$$

Theorem 3.1. *Suppose K is a kernel of order $\lfloor\beta\rfloor + 1$, $\beta \geq 0$ and $\int_{-\infty}^{\infty} |K^{ft}(t)|^2 (1+|t|)^{2\alpha-2} dt < \infty$. Under* (1.2) *with*

$$h_n \asymp \begin{cases} O\{n^{-1/(2\beta+2)}\}, & \text{if } \alpha < \frac{1}{2}, \\ \Big\{\frac{\log(n)}{n}\Big\}^{1/(2\beta+2)}, & \text{if } \alpha = \frac{1}{2}, \\ n^{-1/(2\beta+2\alpha+1)}, & \text{if } \alpha > \frac{1}{2}, \end{cases}$$

one has

$$\sup_{f_X \in \mathcal{H}(\beta,L)} \mathbb{E}_{f_X} |\widehat{P}_{a,b,n} - P_{a,b}|^2 = \begin{cases} O(n^{-1}), & \text{if } \alpha < \frac{1}{2}, \\ O\Big\{\frac{\log(n)}{n}\Big\}, & \text{if } \alpha = \frac{1}{2}, \\ O\{n^{-2(\beta+1)/(2\beta+2\alpha+1)}\}, & \text{if } \alpha > \frac{1}{2}. \end{cases}$$

Note that no smoothness assumptions on f_X (i.e., $\beta = 0$) is needed to obtain the rate n^{-1} or $n^{-2/(2\alpha+1)}$ in cases $\alpha < 1/2$ or $\alpha > 1/2$, respectively. However, smoothness restrictions on f_X accelerate the rate of convergence in the latter case. Therefore, as in the case where $X_1, \ldots, X_n$ are directly observed, the problem of distribution estimation is different from density estimation, as considered in the previous section, and the rates can be arbitrarily slow in absence of smoothness.

For $\alpha > 1/2$, the bandwidth selection in Theorem 3.1 that leads to optimal rates converges to zero at the same rate as the rate-optimal bandwidth in density estimation with respect to the MISE; see Theorem 2.2. The cross-validation bandwidth minimizes the MISE asymptotically in standard density deconvolution (see Hesse (1999)). Therefore, in practical applications it seems appropriate to use cross-validation in order to select the bandwidth in distribution estimation as well.

4. Support Estimation

We consider the problem of estimating the support of f_X. We restrict our consideration to that part of the support that is included in $[0,1]$, denoted by S_X. If f_X vanishes outside $[0,1]$ then S_X becomes the support of f_X. However, non-compactly supported f_X are also included in our framework. If a different interval is of interest, the data can be transformed to $[0,1]$ with no loss.

Our assumptions are the following.

(S1) $S_X = [a,b] \subseteq [0,1]$, where $a < b$.

(S2) $f_X(x) = 0$ for all $x \in [0,1]\backslash S_X$ and $f_X(x) \geq c_f > 0$ for all $x \in S_X$, where c_f is a known constant.

(S3) $f_X(x) \leq c_{\max} < \infty$ for all $x \in [0,1]$.

Note that f_X is bounded away from zero on S_X and may have jump discontinuities at its endpoints. No smoothness conditions on f_X are imposed. Condition **(S1)** restricts the complexity of the support inside $[0,1]$ but no other assumptions on the support are required. For deconvolution support estimation for standard error densities, see e.g., Goldenshluger and Tsybakov (2004), Delaigle and Gijbels (2006), and Meister (2006).

To construct an estimator of S_X, we consider the function $T(A) := \int_A f_X(x)dx$ $-c_f\lambda(A)/2$, where λ denotes Lebesgue-Borel measure and $A \subseteq [0,1]$. We notice that T takes its maximum at $A = S_X$. Functions such as T are referred to as excess mass functions in the literature, see e.g., Hartigan (1987), Müller and Sawitzki (1991), Nolan (1991), Polonik (1995) and Cuevas, Manteiga and Rodriguez Casal (2006). Excess mass has been used to estimate level sets of a density, i.e., those regions where the density has some given lower bound, but are also applicable to support estimation for multivariate data. So far most considerations have been restricted to cases where the data are measured without any error.

We introduce equidistant grid points $a_j = jh_n$, and the set of intervals $G_n := \{(a_j, a_k) : j,k = 0,\ldots,[1/h_n]\}$. As an estimator of $\int_A f_X(x)dx$ for $A \in G_n$, we employ $\widehat{P}_{A,n} := \widehat{P}_{a_j,a_k,n}$, where $A = (a_j, a_k)$ and $\widehat{P}_{a,b,n}$ is defined as in the previous section. As an empirically accessible version of T, we take

$$\widehat{T}_n(A) := \widehat{P}_{A,n} - \frac{c_f}{2}\lambda(A).$$

Then, our estimator $\widehat{S}_{X,n}$ of S_X is taken to be that element of G_n that maximizes $\widehat{T}_n$ within the collection G_n. In the case of several maximizing sets, we choose just one of them as $\widehat{S}_{X,n}$.

The densities satisfying **(S1)**−**(S3)** are collected in the class $\mathcal{S}_{c_{\max},c_f}$. The convergence rates are given as follows.

Theorem 4.1. *Under* (1.2), *if*

$$\int_{-\infty}^{\infty} |K^{ft}(t)|^2(1+|t|)^{2\alpha-2}dt < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} |K^{ft}(t)|(1+|t|)^{\alpha-1}dt,$$

$$h_n \asymp \begin{cases} n^{-1}, & \text{if } \alpha < \frac{1}{2}, \\ \left\{\frac{\log(n)}{n}\right\}^{1/2}, & \text{if } \alpha = \frac{1}{2}, \\ n^{-1/(2\alpha+1)}, & \text{if } \alpha > \frac{1}{2}, \end{cases}$$

then

$$\sup_{f_X \in \mathcal{S}_{c_{\max},c_f}} \mathbb{E}_{f_X}\left\{\lambda(\widehat{S}_{X,n}\Delta S_X)\right\} = \begin{cases} O\{n^{-1/2}\}, & \text{if } \alpha < \frac{1}{2}, \\ O\left\{\sqrt{\frac{\log(n)}{n}}\right\}, & \text{if } \alpha = \frac{1}{2}, \\ O\{n^{-1/(2\alpha+1)}\}, & \text{if } \alpha > \frac{1}{2}, \end{cases}$$

where $A\Delta B = (A \setminus B) \cup (B \setminus A)$ *denotes the symmetric difference of the sets* A *and* B*.*

The selection of the bandwidth according to Theorem 4.1 does not require knowledge of the smoothness degree β. However, the constant c_f must be known in order to construct the estimator $\widehat{S}_{X,n}$; it is sufficient to know just a lower bound on the restriction of f_X to its support, not necessarily the infimum. Choosing c_f based on the data and constructing an adaptive estimator is a difficult problem. In the error-free setting, Cuevas and Fraiman (1997) study support estimation based on the empirical level sets of a kernel density estimator; therein, the constant c_f is treated as a smoothing parameter.

In the current problem, it is also possible to derive a reasonable data-driven selector as

$$\widehat{c}_f = \sup\left\{c \in [0,1] \,:\, \widehat{Q}_n(c) \le \widehat{Q}_n(0) + \rho_n\right\},$$

where

$$\widehat{Q}_n(c) = \int_0^1 \max\left\{|\widehat{f}_{X,n}(x)|, c\right\}|\widehat{f}_{X,n}(x)|dx\,,$$

and the density estimator $\widehat{f}_{X,n}$ is defined at (2.1). Its bandwidth can be chosen according to the cross-validation method introduced in (2.3). We may select the sequence $(\rho_n)_n \asymp (\log n)^{-1}$ arbitrarily. This empirical selector of c_f is motivated by the fact that $Q(c)$, the term we obtain by replacing $\widehat{f}_{X,n}$ by the true density f_X, is a constant function on $[0, c_{0,f}]$, where $c_{0,f}$ is equal to the largest constant c_f that satisfies **(S2)** and, for $c > c_{0,f}$, the function $Q(c)$ is increasing. We can prove that $P\big(\widehat{c}_f \in [c_{0,f}-\delta, c_{0,f}+\delta]\big)$ converges to one for any $\delta > 0$ if f_X has a bounded, continuous and integrable derivative except at the boundaries a and b. We propose to set c_f equal to the empirically accessible quantity $\widehat{c}_f$ in the construction of $\widehat{S}_{X,n}$.

5. Lower Asymptotic Risk Bounds

We establish lower asymptotic risk bounds in order to show that our estimators attain the optimal convergence rates with respect to an arbitrary estimator based on the given data structure. To derive these bounds, we specify appropriate finite-dimensional subclasses of the considered nonparametric distribution classes; these subclasses are comprehensive enough to capture the difficulty of the nonparametric estimation problem. In particular, the lower asymptotic risk bounds derived for these subclasses coincide with the upper bounds established for our estimators.

In the two cases of pointwise density estimation and probability estimation it suffices to consider one-dimensional subexperiments. To generate potential densities for the X_j's, we start with a density of the so-called Pearson type VII,

$$f_0(x) = \frac{\Gamma(r)}{\sqrt{\pi}\Gamma(r-1/2)} \frac{c}{\{1+(cx)^2\}^r}, \tag{5.1}$$

where $1/2 < r < 1$. Two competing densities are now obtained as

$$f_{X,\theta}(x) = f_0(x) + \theta c \delta_n^\beta H(\frac{x}{\delta_n}), \tag{5.2}$$

where $\theta \in \{0,1\}$, $\delta_n \asymp n^{-1/(2\beta+2\alpha+1)}$ and the function H satisfies

(A1) (i) $H^{(l)}(0) \neq 0$, for $l = 0, \ldots, \lfloor \beta \rfloor$, and $\int_{-\infty}^{0} H(x)dx = 0$,

(ii) $H \in \mathcal{H}(\beta, L)$,

(iii) $\int_{-\infty}^{\infty} H(x)dx = 0$,

(iv) $H^{ft}(t) = 0$, if $|t| \notin [1,2]$,

(v) $\sup_t |(H^{ft})^{(j)}(t)| < \infty$ $(j = 0,1,2)$,

(vi) $H(x) = O(|x|^{-2r})$.

It follows from the arguments in Fan (1991) that a function H with these properties exists and that $f_{X,0}, f_{X,1} \in \mathcal{H}(\beta, L)$ if the constant c in (5.1) and (5.2) is sufficiently small.

For the error distribution we assume a bounded density f_ε and a characteristic function satisfying

(A2) (i) $\int |(f_\varepsilon^{ft})^{(j_1)}(\tau)(f_\varepsilon^{ft})^{(j_2)}(t-\tau)|^2 d\tau =: \overline{\rho}^{(j_1,j_2)}(t) = O\left\{(1+|t|)^{-2\alpha}\right\}$, $j_1, j_2 = 0, 1$,

(ii) $f_\varepsilon(x) \geq C/(1+x^2)^{r'}$, where $r' = 2 - r$.

Remark 1. Condition **(A2)**(i) differs from (1.2), it follows from $(f_\varepsilon^{ft})^{(j)}(t) \leq C(1+|t|)^{-\alpha}$, for all $t \in \mathbb{R}, j \in \{0,1\}$, for some $C > 0$. Such a condition is in Fan (1991, 1993). In general **(A2)** is needed to bound the Hellinger distance by the $L_2(\mathbb{R})$-distance of the densities so that Parseval's identity may be applied.

Lemma 5.1. *Suppose that* **(A1)** *and* **(A2)** *hold. If* $g_\theta(y,z) = \int f_{X,\theta}(x) f_\varepsilon(y-x) f_\varepsilon(z-x)dx$, $\theta = 0, 1$, *then* $He^2(g_0, g_1) = O(n^{-1})$, *where* $He(f,g) = \{(1/2)\int(\sqrt{f} - \sqrt{g})^2\}^{1/2}$ *denotes Hellinger distance.*

On the basis of this lemma, we can now derive a lower asymptotic risk bound.

Theorem 5.1. Suppose that **(A1)** and **(A2)** hold. Then

(i) $\inf_{\widetilde{f}^{(l)}_{X,n}} \max_{\theta\in\{0,1\}} \mathbb{E}_\theta \left|\widetilde{f}^{(l)}_{X,n}(0) - f^{(l)}_{X,\theta}(0)\right|^2 \geq Cn^{-2(\beta-l)/(2\beta+2\alpha+1)}$,

(ii) $\inf_{\widetilde{P}_{a,b,n}} \max_{\theta\in\{0,1\}} \mathbb{E}_\theta \left(\widetilde{P}_{a,b,n} - P_{a,b,\theta}\right)^2 \geq \begin{cases} Cn^{-1}, & \textit{if } \alpha \leq \frac{1}{2}, \\ Cn^{-2(\beta+1)/(2\beta+2\alpha+1)}, & \textit{if } \alpha > \frac{1}{2}. \end{cases}$

In the case of density estimation in L_2, different subclasses of target densities must be used in order to establish optimal lower bounds. We consider densities of the form

$$f_{X,\boldsymbol{\theta}}(x) = f_0(x) + \sum_{j=1}^{M_n} \theta_j c \delta_n^\beta H(\frac{x}{\delta_n} - j), \tag{5.3}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{M_n})' \in \{0,1\}^{M_n}$, $\delta_n \asymp n^{-1/(2\beta+2\alpha+1)}$ and $M_n = [1/\delta_n]$. The function H is assumed to satisfy

(A1') (i) $0 < \int |H^{(l)}|^2(x)dx < \infty$, for $l = 0, \ldots, \lfloor\beta\rfloor$,

(ii) $H \in \mathcal{S}(\beta, L)$,

(iii) $\int_{-\infty}^{\infty} H(x)dx = 0$,

(iv) $H^{ft}(t) = 0$, if $|t| \notin [1,2]$,

(v) $\sup_t |(H^{ft})^{(j)}(t)| < \infty$ $(j = 0, 1, 2)$,

(vi) $H(x) = O(|x|^{-2r})$.

It follows again from the arguments in Fan (1991) that such a function H exists and that $f_{X,\boldsymbol{\theta}} \in \mathcal{S}(\beta, L)$ for all $\boldsymbol{\theta} \in \{0,1\}^{M_n}$, if the constant c in (5.3) is sufficiently small.

Lemma 5.2. *Suppose that* **(A1')** *and* **(A2)** *hold. If* $g_{\boldsymbol{\theta}}(y,z) = \int f_{X,\boldsymbol{\theta}}(x) f_\varepsilon(y-x) f_\varepsilon(z-x)dx$, $\boldsymbol{\theta} \in \{0,1\}^{M_n}$, *then*

$$\max_{\boldsymbol{\theta}\in\{0,1\}^{M_n}, 1\leq j\leq M_n} He^2\left\{g_{(\theta_1,\ldots,\theta_{M_n})}, g_{(\theta_1,\ldots,\theta_{j-1},1-\theta_j,\theta_{j+1},\ldots,\theta_{M_n})}\right\} = O(n^{-1}).$$

This lemma leads to a lower asymptotic bound for the MISE.

Theorem 5.2. *Suppose that* **(A1')** *and* **(A2)** *hold. Then*

$$\inf_{\widetilde{f}_{X,n}^{(l)}} \max_{\boldsymbol{\theta}\in\{0,1\}^{M_n}} \mathbb{E}_{\boldsymbol{\theta}} \left\| \widetilde{f}_{X,n}^{(l)} - f_{X,\boldsymbol{\theta}}^{(l)} \right\|_{L_2(\mathbb{R})}^2 \geq C n^{-2(\beta-l)/(2\beta+2\alpha+1)}.$$

With respect to the convergence rates in support estimation (Section), let

$$f_n(x) = \frac{1}{2(1-\varepsilon_n/8)} \mathbb{1}_{[\varepsilon_n/8,1]}(x) + \frac{1}{2c(1+x^2)} \mathbb{1}_{\mathbb{R}\setminus[0,1]}(x)\,,$$

with $c := \int_{\mathbb{R}\setminus[0,1]}(1+x^2)^{-1}dx$, for some sequence $(\varepsilon_n)_n \downarrow 0$ still to be determined. Let $Q(x) = \{1-\cos(x)\}^2/(\pi x^2)^2$, with Fourier transform Q^{ft} twice continuously differentiable and supported on $[-2,2]$, and consider the densities

$$\tilde{f}_n(x) = f_n(x) + \cos(\frac{2\pi x}{\varepsilon_n}) Q(\frac{x}{\varepsilon_n})$$

with Fourier transforms

$$\tilde{f}_n^{ft}(t) = f_n^{ft}(t) + \frac{\varepsilon_n}{2}\, Q^{ft}\big\{\varepsilon_n(t - \frac{2\pi}{\varepsilon_n})\big\} + \frac{\varepsilon_n}{2}\, Q^{ft}\big\{\varepsilon_n(t + \frac{2\pi}{\varepsilon_n})\big\}.$$

We easily verify that all f_n, $\tilde{f}_n$ are contained in $\mathcal{S}_{c_{\max},c_f}$ for some appropriate parameters $c_f, c_{\max}$. We use $(f_n)_n$ and $(\tilde{f}_n)_n$ as the competing sequences for f_X. We show the lower bound matching Theorem 4.1 for $\alpha > 1/2$.

Theorem 5.3. *Suppose that* **(A2)** *holds with $r = 1$. Then, for any estimator $\widetilde{S}_{X,n}$ of S_X, we have*

$$\max_{f_X \in \{f_n, \tilde{f}_n\}} \mathbb{E}\Big\{ \lambda\big(\widetilde{S}_{X,n} \Delta S_X\big)\Big\} \geq C n^{-1/(2\alpha+1)}.$$

6. Simulations

To illustrate the potential of our density estimator, we present the results of a small simulation study. The implementation was done on the basis of the statistical software package *R*; see R Development Core Team (2007).

We considered the estimator $\widehat{f}_{X,n} = \widehat{f}_{X,n}^{(0)}$ defined by (2.1) with the sinc kernel $K^{ft}(t) = \mathbb{1}_{[-1,1]}(t)$. The bandwidth was chosen by the cross validation criterion (2.3). We were particularly interested in cases where traditional kernel deconvolution estimators fail. Accordingly, we chose a uniform distribution on

$[-1,1]$ for the errors, one that has a characteristic function $f_\varepsilon^{ft}(t) = \sin(t)/t$, with isolated zeros at $t = k\pi$, $k = \pm 1, \pm 2, \ldots$.

Suppose we have data $Y_{j,k} = X_j + \varepsilon_{j,k}$ with two replications, $k \in \{1,2\}$. As an alternative to our estimator, one could perhaps average these data first, $\bar{Y}_j = (Y_{j,1} + Y_{j,2})/2$, and apply some deconvolution technique afterward. Since the characteristic function of the errors of the averaged data has again isolated zeros, we take as an alternative estimator the one proposed by Hall and Meister (2007):

$$\widetilde{f}_{X,n}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\widehat{f}_{\bar{Y}}^{ft}(t) f_{\bar{\varepsilon}}^{ft}(-t) |f_{\bar{\varepsilon}}^{ft}(t)|^2}{\Big[\max\{|f_{\bar{\varepsilon}}^{ft}(t)|, h(t)\}\Big]^3} \exp(-itx) dt,$$

where $\widehat{f}_{\bar{Y}}^{ft}(t) = n^{-1} \sum_{j=1}^n \exp(it\bar{Y}_j)$ and $h(t) = \kappa t^2$. The isolated zeros of $f_{\bar{\varepsilon}}^{ft}$ are taken into account by the ridge function h, where the parameter κ is chosen as a minimizer of the cross validation function

$$\begin{aligned} CV(\kappa) = & \int_{-\infty}^{\infty} |\widetilde{f}_{X,n}(x)|^2 dx \\ & - \frac{1}{\pi n(n-1)} \int_{-\infty}^{\infty} \frac{|f_{\bar{\varepsilon}}^{ft}(t)|^2}{(\max\{|f_{\bar{\varepsilon}}^{ft}(t)|, \kappa t^2\})^3} \sum_{j \neq k} \exp\{-i(\bar{Y}_j - \bar{Y}_k)\} dt. \end{aligned}$$

Note that $f_{\bar{\varepsilon}}^{ft}$ has isolated zeros at $t = 2k\pi$, for $k = \pm 1, \pm 2, \ldots$. To demonstrate possible problems with these zeros we chose a target density f_X as

$$f_X(u) = \frac{\{1 - \cos(u)\}\{1 + \cos(2\pi u)\}}{\pi u^2}.$$

The characteristic function

$$f_X^{ft}(t) = (1 - |t|)_+ + \frac{\{(1 - |t + 2\pi|)_+ + (1 - |t - 2\pi|)_+\}}{2},$$

has considerable mass around the points $\pm 2\pi$; see Figure 1.

In view of the difficulty of the estimation problem, we expect that relatively large sample sizes are required to obtain good results. Therefore, we tried a sample size of $n = 1,000$. Table 1 shows estimates of the mean squared error of both estimators based on $N = 1,000$ Monte Carlo replications.

Figure 2 shows the target density (thick solid line), and shows estimates corresponding to a moderately favorable sample (with the 250th smallest sum of ISEs, dashed line), an average sample (with the 500th smallest sum of ISEs, dotted line), and a moderately unfavorable sample (with the 750th smallest sum of ISEs, dot-dashed line).

These plots indicate that $\widehat{f}_{X,n}$ estimates the target quantity reasonably well in many instances, while $\widetilde{f}_{X,n}$ tends to approximate a different density,

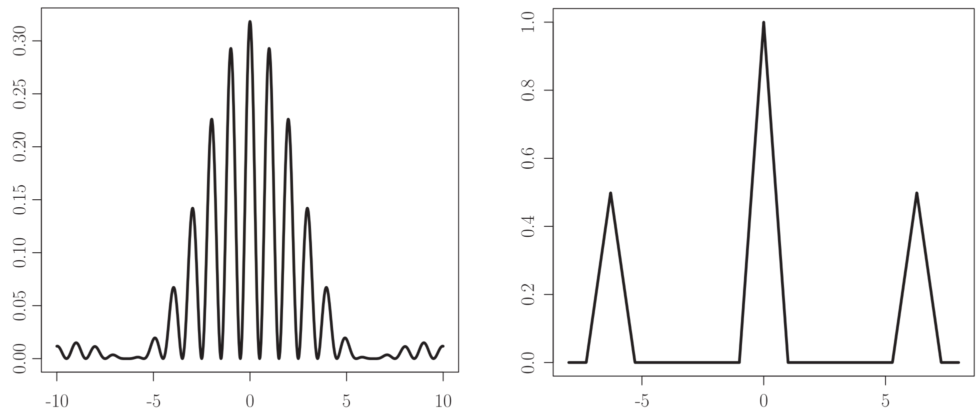

Figure 1. Left: target density f_X, right: characteristic function f_X^{ft}.

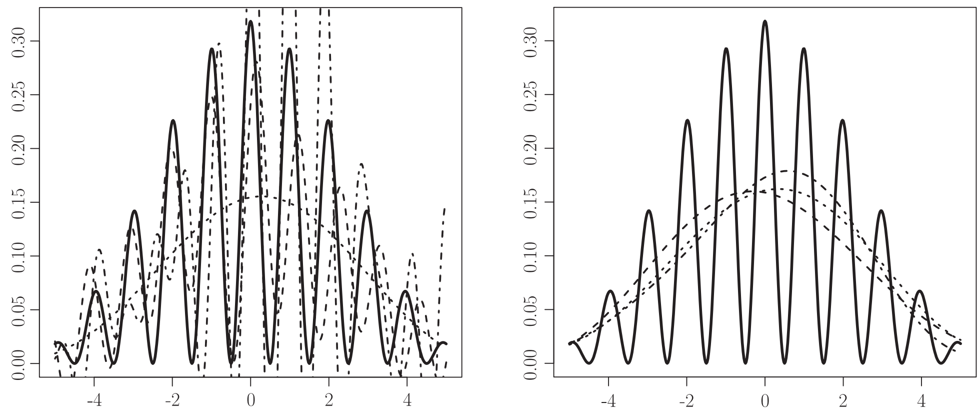

Figure 2. Left: target density f_X (solid) and three of our estimates, right: target density f_X (solid) and three estimates from Hall and Meister (2007).

Table 1. MSE, $n = 1,000$.

	MSE, $n = 1,000$
our estimator $\widehat{f}_{X,n}$	0.0506
estimator $\widetilde{f}_{X,n}$ in Hall and Meister (2007)	0.1067

$f_{X,wrong}(u) = (1 - \cos(u))/(\pi u^2)$. Figure 3, which contains corresponding estimates of the characteristic function, $\widehat{f}_{X,n}^{ft}$ and $\widetilde{f}_{X,n}^{ft}$, reveals what happens in both cases. Here we indicate the real parts of the estimates by dashed lines and the imaginary parts by dotted lines. While $\widehat{f}_{X,n}^{ft}$ tends to estimate f_X^{ft}, $\widetilde{f}_{X,n}^{ft}$ merely approximates $(1-|t|)_+$, the characteristic function of the density $f_{X,wrong}$. This is of course the expected effect of the ridging that was introduced by Hall and Meister (2007) in order to deal with zeros of the characteristic function of

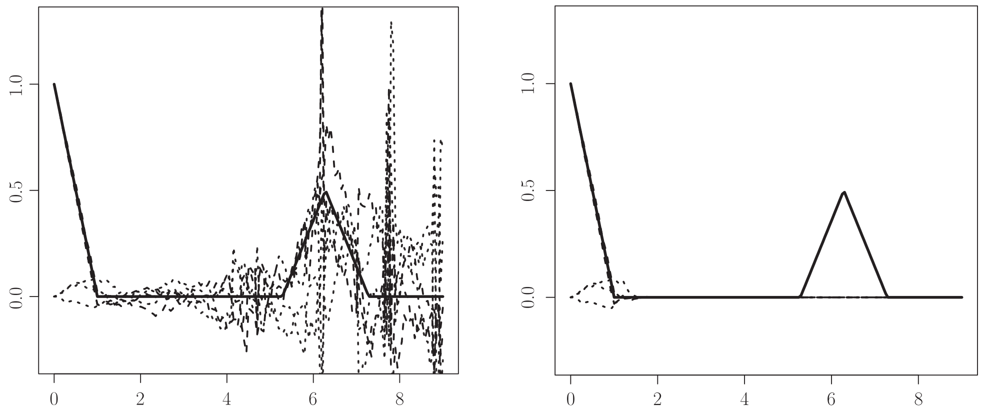

Figure 3. Left: char. function f_X^{ft} (solid) and three of our estimates, right: char. function f_X^{ft} (solid) and three estimates from Hall and Meister (2007).

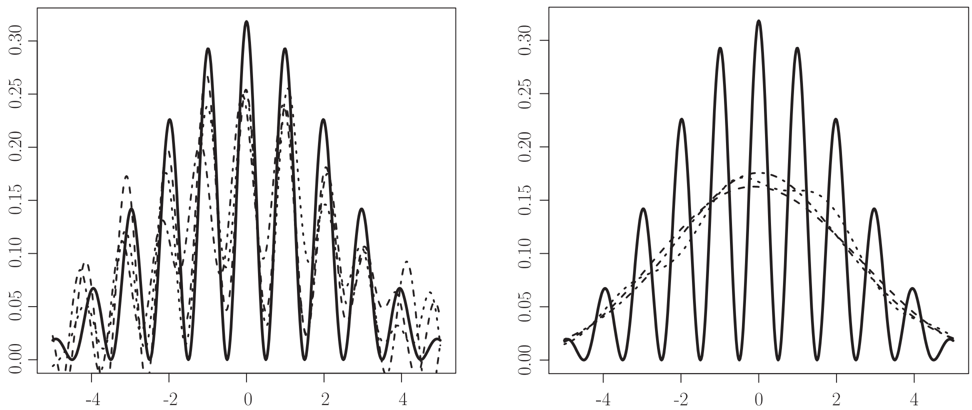

Figure 4. Left: target density f_X (solid) and three of our estimates, right: target density f_X (solid) and three estimates from Hall and Meister (2007).

Table 2. MSE, $n = 3,000$.

	MSE, n=3,000
our estimator $\widehat{f}_{X,n}$	0.0189
estimator $\widetilde{f}_{X,n}$ in Hall and Meister (2007)	0.0697

the error density.

These effects become even more pronounced for large sample sizes; see Table 2 and Figures 4 and 5 for analogous results with $n = 3,000$.

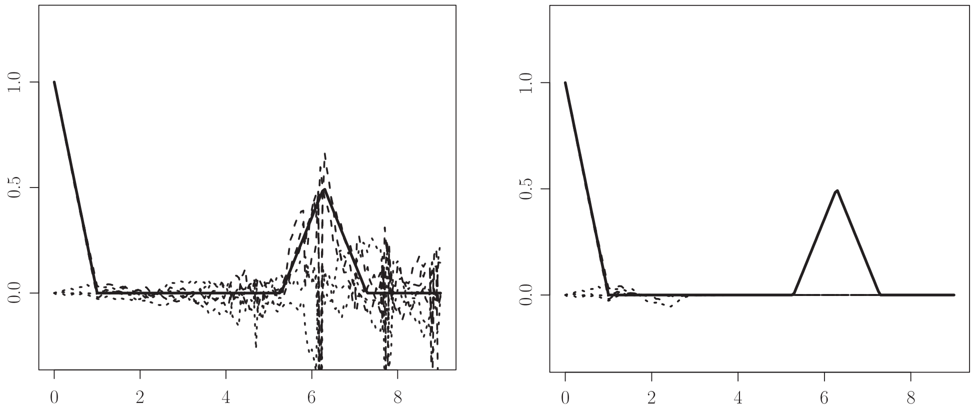

Figure 5. Left: char. function f_X^{ft} (solid) and three of our estimates, right: char. function f_X^{ft} (solid) and three estimates from Hall and Meister (2007).

7. Proofs

Proof of Theorem 2.1. We have that

$$\begin{aligned}\mathbb{E}_{f_X}\widehat{f}^{(l)}_{X,n}(x_0) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-itx_0)(-it)^l K^{ft}(th_n)\mathbb{E}_{f_X}\frac{\widehat{\varphi}_{Y,n}\{\tau(t),t-\tau(t)\}}{f_\varepsilon^{ft}\{\tau(t)\}f_\varepsilon^{ft}\{t-\tau(t)\}}dt\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-itx_0)(-it)^l K^{ft}(th_n)f_X^{ft}(t)dt\\ &= \int_{-\infty}^{\infty}\frac{1}{h_n}K\left(\frac{x_0-y}{h_n}\right)f_X^{(l)}(y)dy.\end{aligned}$$

This implies that

$$\begin{aligned}\mathbb{E}_{f_X}\widehat{f}^{(l)}_{X,n}(x_0)-f_X^{(l)}(x_0) &= \int_{-\infty}^{\infty}\frac{1}{h_n}K\left(\frac{x_0-y}{h_n}\right)\left\{f_X^{(l)}(y)-f_X^{(l)}(x_0)\right\}dy\\ &= O\left(h_n^{\beta-l}\right). \qquad (7.1)\end{aligned}$$

Furthermore, we find

$$\begin{aligned}\mathrm{Var}\left\{\widehat{f}^{(l)}_{X,n}(x_0)\right\} &\le \frac{1}{n(2\pi)^2}\mathbb{E}_{f_X}\Bigg|\int_{-\infty}^{\infty}\exp(-itx_0)(-it)^l K^{ft}(th_n)\\ &\qquad\times\frac{\exp(itX_1)\exp\{i\tau(t)\varepsilon_{11}\}\exp[i\{t-\tau(t)\}\varepsilon_{12}]}{f_\varepsilon^{ft}\{\tau(t)\}f_\varepsilon^{ft}\{t-\tau(t)\}}dt\Bigg|^2\\ = \frac{1}{n(2\pi)^2}\mathbb{E}_{f_X}\int_{-\infty}^{\infty}\Big|\int_{-\infty}^{\infty}&\exp(ity)\exp(-itx_0)(-it)^l K^{ft}(th_n)\\ &\qquad\times\frac{\exp\{i\tau(t)\varepsilon_{11}\}\exp[i\{t-\tau(t)\}\varepsilon_{12}]}{f_\varepsilon^{ft}\{\tau(t)\}f_\varepsilon^{ft}\{t-\tau(t)\}}dt\Big|^2 f_X(y)dy\end{aligned}$$

$$\leq \frac{\|f_X\|_\infty}{n(2\pi)^2} \mathbb{E}_{f_X} \int_{-\infty}^{\infty} \Big| \int_{-\infty}^{\infty} \exp(ity)\exp(-itx_0)(-it)^l K^{ft}(th_n) \times \frac{\exp\{i\tau(t)\varepsilon_{11}\}\exp[i\{t-\tau(t)\}\varepsilon_{12}]}{f_\varepsilon^{ft}\{\tau(t)\} f_\varepsilon^{ft}\{t-\tau(t)\}} dt \Big|^2 dy.$$

Applying Plancherel's identity, and taking into account that $\sup_{f\in\mathcal{H}(\beta,L)} \|f\|_\infty < \infty$ (see, for example, Lemma 1 in Bickel and Ritov (1988)), we obtain that

$$\begin{aligned}
&\mathrm{Var}\left\{\widehat{f}_{X,n}^{(l)}(x_0)\right\} \\
&\leq \frac{C}{n}\mathbb{E}_{f_X} \int_{-\infty}^{\infty} \left| \exp(-itx_0)(-it)^l K^{ft}(th_n) \frac{\exp\{i\tau(t)\varepsilon_{11}\}\exp[i\{t-\tau(t)\}\varepsilon_{12}]}{f_\varepsilon^{ft}\{\tau(t)\} f_\varepsilon^{ft}\{t-\tau(t)\}} \right|^2 dt \\
&\leq \frac{C}{n} \int_{-\infty}^{\infty} |t|^{2l} |K^{ft}(th_n)|^2 \rho(t)^{-2} dt \\
&= O\left(n^{-1} h_n^{-2\alpha-2l-1}\right),
\end{aligned}$$

which completes the proof.

Proof of Theorem 2.2. As in (7.1), we have that

$$\mathbb{E}_{f_X} \widehat{f}_{X,n}^{(l)}(x) - f_X^{(l)}(x) = \int_{-\infty}^{\infty} \frac{1}{h_n} K\left(\frac{x-y}{h_n}\right) \left\{ f_X^{(l)}(y) - f_X^{(l)}(x) \right\} dy,$$

which implies by Parseval's identity that

$$\begin{aligned}
\left\| \mathbb{E}_{f_X} \widehat{f}_{X,n}^{(l)} - f_X^{(l)} \right\|_{L_2(\mathbb{R})}^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \{K^{ft}(th_n) - 1\}(-it)^l f_X^{ft}(t) \right|^2 dt \\
&\leq C \int_{-\infty}^{\infty} |K^{ft}(th_n)-1|^2 |t|^{2(l-\beta)}\ |t|^{2\beta} |f_X^{ft}(t)|^2 dt \\
&= O\left[\sup_t \{ |K^{ft}(th_n)-1|^2 |t|^{2(l-\beta)} \} \right] \\
&= O\left\{ h_n^{2(\beta-l)} \right\}.
\end{aligned}$$

Furthermore, we have, again by Parseval's identity, that

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathrm{Var}\left\{\widehat{f}_{X,n}^{(l)}(x)\right\} dx &= \int_{-\infty}^{\infty} \mathbb{E}_{f_X} \left| \widehat{f}_{X,n}^{(l)}(x) - \mathbb{E}_{f_X} \widehat{f}_{X,n}^{(l)}(x) \right|^2 dx \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbb{E}_{f_X} \left| K^{ft}(th_n)(-it)^l \{\widehat{\varphi}_{X,n}(t) - f_X^{ft}(t)\} \right|^2 dt \\
&\leq \frac{C}{n} \int_{-\infty}^{\infty} \left| K^{ft}(th_n)(-it)^l / \rho(t) \right|^2 dt
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{C}{n}\int_{-\infty}^{\infty}|K^{ft}(th_n)|^2(1+|t|)^{2\alpha+2l}dt \\
&= O\left(n^{-1}h_n^{-2\alpha-2l-1}\right).
\end{aligned}
$$

Proof of Theorem 3.1. We have that

$$
\begin{aligned}
\mathbb{E}_{f_X}\widehat{P}_{a,b,n} &= \frac{1}{2\pi}\int \Psi_{a,b}(-t)K^{ft}(th_n)f_X^{ft}(t)dt \\
&= \int_a^b\int \frac{1}{h_n}K\left(\frac{y}{h_n}\right)f_X(x-y)dydx \\
&= \int \frac{1}{h_n}K\left(\frac{y}{h_n}\right)\int_a^b f_X(x-y)dxdy \\
&= \int \frac{1}{h_n}K\left(\frac{y}{h_n}\right)\{F_X(b-y)-F_X(a-y)\}dy.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}_{f_X}\widehat{P}_{a,b,n} - P_{a,b} &= \int \frac{1}{h_n}K\left(\frac{y}{h_n}\right)\{F_X(b-y)-F_X(b)\}dy \\
&\quad - \int \frac{1}{h_n}K\left(\frac{y}{h_n}\right)\{F_X(a-y)-F_X(a)\}dy \\
&= O\left(h_n^{\beta+1}\right). \qquad (7.2)
\end{aligned}
$$

The variance of $\widehat{P}_{a,b,n}$ can be estimated as

$$
\begin{aligned}
&\mathrm{Var}\,(\widehat{P}_{a,b,n}) \\
&\leq \frac{1}{n}\mathbb{E}_{f_X}\left|\int \Psi_{a,b}(-t)K^{ft}(th_n)\frac{\exp(itX_1)\exp\{i\tau(t)\varepsilon_{11}\}\exp[i\{t-\tau(t)\}\varepsilon_{12}]}{f_\varepsilon^{ft}\{\tau(t)\}f_\varepsilon^{ft}\{t-\tau(t)\}}dt\right|^2 \\
&= \frac{1}{n}\mathbb{E}\int\left|\int \exp(ity)\Psi_{a,b}(-t)K^{ft}(th_n)\frac{\exp\{i\tau(t)\varepsilon_{11}\}\exp[i\{t-\tau(t)\}\varepsilon_{12}]}{f_\varepsilon^{ft}\{\tau(t)\}f_\varepsilon^{ft}\{t-\tau(t)\}}dt\right|^2 f_X(y)dy \\
&\leq \frac{\|f_X\|_\infty}{n}\int\left|\Psi_{a,b}(-t)K^{ft}(th_n)/\rho(t)\right|^2dt \\
&= O\left(\frac{\|f_X\|_\infty}{n}\int\left|K^{ft}(th_n)(1+|t|)^{\alpha-1}\right|^2dt\right) \\
&= \begin{cases} O(n^{-1}) & \text{if } \alpha<\frac{1}{2}, \\ O\{n^{-1}\log(\frac{1}{h_n})\} & \text{if } \alpha=\frac{1}{2}, \\ O(n^{-1}h_n^{1-2\alpha}) & \text{if } \alpha>\frac{1}{2}. \end{cases} \qquad (7.3)
\end{aligned}
$$

Proof of Theorem 4.1. Let $a^0 = a^0(f_X), b^0 = b^0(f_X) \in \{a_j : j = 0, \ldots, [1/h_n]\}$ be such that $\lambda(S_X \Delta [a^0, b^0])$ is minimal. It is clear that

$$\sup_{f_X \in \mathcal{S}_{c_{\max}, c_f}} \lambda(S_X \Delta [a^0, b^0]) \leq h_n.$$

It follows from the definition of the set $\mathcal{S}_{c_{\max}, c_f}$ that estimates hold uniformly in $f_X \in \mathcal{S}_{c_{\max}, c_f}$. First, there exists some constant $C_1 < \infty$ such that, for all $0 \leq j \leq k \leq [1/h_n]$,

$$\begin{aligned} &\left\{ P_{a^0,b^0} - \frac{c_f(b^0 - a^0)}{2} \right\} - \left\{ P_{a_j,a_k} - \frac{c_f(a_k - a_j)}{2} \right\} \\ &\qquad \geq \frac{c_f}{2} \lambda\left([a^0, b^0] \Delta [a_j, a_k]\right) - C_1 h_n =: d(j,k). \end{aligned}$$

Note that $\widehat{P}_{a,b,n} - \mathbb{E}\widehat{P}_{a,b}$ can be written as $\sum_{j=1}^n Z_{n,j}$, with

$$Z_{n,j} = \frac{1}{n} \frac{1}{2\pi} \int \Psi_{a,b}(-t) K^{ft}(th_n) \left(\frac{\exp[i\tau(t)Y_{j,1} + i\{t - \tau(t)\}Y_{j,2}]}{f_\varepsilon^{ft}\{\tau(t)\}\, f_\varepsilon^{ft}\{t - \tau(t)\}} - f_X^{ft}(t) \right) dt\ .$$

Then

$$\begin{aligned} \operatorname{ess\,sup} |Z_{n,j}| &= O\left\{ n^{-1} \int |K^{ft}(th_n)| (1 + |t|)^{\alpha - 1} dt \right\} \\ &= O\left\{ n^{-1} h_n^{-\alpha} \right\}, \end{aligned}$$

which implies, in conjunction with (7.3), that

$$\mathbb{E}Z_{n,j}^4 \leq \mathbb{E}Z_{n,j}^2 \operatorname{ess\,sup} |Z_{n,j}|^2 = \begin{cases} O(n^{-4} h_n^{-2\alpha}) & \text{if } \alpha < \frac{1}{2}, \\ O\{n^{-4} \log(\frac{1}{h_n}) h_n^{-1}\} & \text{if } \alpha = \frac{1}{2}, \\ O(n^{-4} h_n^{-2\alpha}) & \text{if } \alpha > \frac{1}{2}. \end{cases}$$

Therefore, again in conjunction with (7.3),

$$\begin{aligned} \mathbb{E}\left(\widehat{P}_{a,b,n} - \mathbb{E}\widehat{P}_{a,b,n}\right)^4 &= \mathbb{E}\Big(\sum_{j=1}^n Z_{n,j}\Big)^4 \\ &= 3\Big\{\operatorname{Var}\big(\sum_{j=1}^n Z_{n,j}\big)\Big\}^2 + n\left\{\mathbb{E}Z_{n,1}^4 - 3(\mathbb{E}Z_{n,1}^2)^2\right\} \\ &= \begin{cases} O(n^{-2} + n^{-3} h_n^{-2\alpha}) & \text{if } \alpha < \frac{1}{2} \\ O[n^{-2}\{\log(n)\}^2 + n^{-3}\log(n) h_n^{-1}] & \text{if } \alpha = \frac{1}{2} \\ O(n^{-2} h_n^{2-4\alpha} + n^{-3} h_n^{-2\alpha}) & \text{if } \alpha > \frac{1}{2} \end{cases} \end{aligned}$$

$$= \begin{cases} O(n^{-2}) & \text{if } \alpha < \frac{1}{2} \\ O[n^{-2}\{\log(n)\}^2] & \text{if } \alpha = \frac{1}{2} \\ O(n^{-2}h_n^{2-4\alpha}) & \text{if } \alpha > \frac{1}{2} \end{cases}$$
$$= O(h_n^4).$$

This implies, in conjunction with (7.2), that $\mathbb{E}_{f_X}\left(\widehat{P}_{a,b,n} - P_{a,b}\right)^4 = O(h_n^4)$ holds uniformly in a, b. Therefore, for all j, k with $(c_f/2)\lambda([a^0, b^0]\Delta[a_j, a_k]) > C_1 h_n$,

$$\begin{aligned} P_{f_X}\left(\widehat{S}_{X,n} = [a_j, a_k]\right) &\le P_{f_X}\left\{\widehat{P}_{a^0,b^0,n} - \frac{c_f(b^0 - a^0)}{2} \le \widehat{P}_{a_j,a_k,n} - \frac{c_f(a_k - a_j)}{2}\right\} \\ &\le P_{f_X}\left\{P_{a^0,b^0} - \widehat{P}_{a^0,b^0,n} \ge \frac{d(j,k)}{2}\right\} \\ &\quad + P_{f_X}\left\{P_{a_j,a_k} - \widehat{P}_{a_j,a_k,n} \ge \frac{d(j,k)}{2}\right\} \\ &\le C\,\frac{h_n^4}{d(j,k)^4}. \end{aligned}$$

This, however, implies

$$\begin{aligned} \mathbb{E}_{f_X}\left(\widehat{S}_{X,n}\Delta S_X\right) &\le h_n + \mathbb{E}_{f_X}\left(\widehat{S}_{X,n}\Delta[a^0, b^0]\right) \\ &\le h_n + \sum_{j,k} \lambda([a_j, a_k]\Delta[a^0, b^0])\; P_{f_X}(\widehat{S}_{X,n} = [a_j, a_k]) \\ &= O(h_n), \end{aligned}$$

uniformly in $f_X \in \mathcal{S}_{c_{\max},c_f}$.

Proof of Lemma 5.1. Recall that, for $(Y_{j,1}, Y_{j,2})' = (X_j + \varepsilon_{j,1}, X_j + \varepsilon_{j,2})'$,

$$\begin{aligned} g_0(y,z) &= \int f_0(x) f_\varepsilon(y-x) f_\varepsilon(z-x)dx, \\ g_1(y,z) &= g_0(y,z) + c\delta_n^\beta \int H(\frac{x}{\delta_n}) f_\varepsilon(y-x) f_\varepsilon(z-x)dx. \end{aligned}$$

We can show that

$$g_0(y,z) \ge \frac{C}{(1+y^2)(1+z^2)}. \tag{7.4}$$

Actually, assume without loss of generality that $|y| > |z|$. Then

$$g_0(y,z) \ge \int \frac{C_r}{(1+x^2)^r} \frac{C}{\{1+(y-x)^2\}^{r'}} \frac{C}{\{1+(z-x)^2\}^{r'}} dx$$

$$
\begin{aligned}
&= \int \frac{C_r}{\{1+(y-x)^2\}^r} \frac{C}{(1+x^2)^{r'}} \frac{C}{\{1+(z-y+x)^2\}^{r'}} dx \\
&\geq \frac{C}{(1+y^2)^r \{1+(z-y)^2\}^{r'}} \\
&\geq \frac{C}{(1+y^2)^r (1+z^2)^{r'}} \geq \frac{C}{(1+y^2)(1+z^2)}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
2\mathrm{He}^2(g_1, g_0) &\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{|g_1(y,z) - g_0(y,z)|^2}{g_0(y,z)} dy\, dz \\
&\leq C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1+y^2)(1+z^2)\, |g_1(y,z) - g_0(y,z)|^2 \, dy\, dz \\
&\leq C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g_1(y,z) - g_0(y,z)|^2 \, dy\, dz \\
&\quad + C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^2 \, |g_1(y,z) - g_0(y,z)|^2 \, dy\, dz \\
&\quad + C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z^2 \, |g_1(y,z) - g_0(y,z)|^2 \, dy\, dz \\
&\quad + C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^2 z^2 \, |g_1(y,z) - g_0(y,z)|^2 \, dy\, dz \\
&=: C \, \{I_1 + \cdots + I_4\}.
\end{aligned}
$$

Since $g_1(y,z) - g_0(y,z) = c\delta_n^\beta \int H(x/\delta_n) f_\varepsilon(y-x) f_\varepsilon(z-x) dx$ has Fourier transform $c\delta_n^\beta \{H(\cdot/\delta_n)\}^{ft}(t_1+t_2) f_\varepsilon^{ft}(t_1) f_\varepsilon^{ft}(t_2) = c\delta_n^{\beta+1} H^{ft}\{(t_1+t_2)\delta_n\} f_\varepsilon^{ft}(t_1) f_\varepsilon^{ft}(t_2)$, by Parseval's identity,

$$
\begin{aligned}
I_1 &= c^2 \delta_n^{2\beta+2} \iint \left| H^{ft}\{(t_1+t_2)\delta_n\} f_\varepsilon^{ft}(t_1) f_\varepsilon^{ft}(t_2) \right|^2 dt_1 dt_2 \\
&= c^2 \delta_n^{2\beta+2} \iint \left| H^{ft}(t\delta_n) f_\varepsilon^{ft}(\tau) f_\varepsilon^{ft}(t-\tau) \right|^2 d\tau dt \\
&= c^2 \delta_n^{2\beta+1} \int_{\{t:\, 1 \leq |t| \leq 2\}} \int_{-\infty}^{\infty} \left| H^{ft}(t) f_\varepsilon^{ft}(\tau) f_\varepsilon^{ft}(\frac{t}{\delta_n} - \tau) \right|^2 d\tau dt \\
&= O\left(\delta_n^{2\beta+1} \int_{\{t:\, 1 \leq |t| \leq 2\}} \int_{-\infty}^{\infty} \left| f_\varepsilon^{ft}(\tau) f_\varepsilon^{ft}(\frac{t}{\delta_n} - \tau) \right|^2 d\tau dt \right) \\
&= O\left(\delta_n^{2\beta+2\alpha+1} \right) = O\left(n^{-1} \right). \qquad (7.5)
\end{aligned}
$$

Similarly, since $y\{g_1(y,z) - g_0(y,z)\}$ has Fourier transform $c\delta_n^{\beta+1}(1/i)(\partial/\partial t_1)$

$H^{ft}\{(t_1+t_2)\delta_n\}f_\varepsilon^{ft}(t_1)f_\varepsilon^{ft}(t_2)$, again by Parseval's identity,

$$
\begin{aligned}
I_2 &= c^2\delta_n^{2\beta+2}\iint\left|\frac{\partial H^{ft}\{(t_1+t_2)\delta_n\}f_\varepsilon^{ft}(t_1)f_\varepsilon^{ft}(t_2)}{\partial t_1}\right|^2 dt_1dt_2\\
&= c^2\delta_n^{2\beta+2}\iint\Big|\delta_n(H^{ft})'\{(t_1+t_2)\delta_n\}f_\varepsilon^{ft}(t_1)f_\varepsilon^{ft}(t_2)\\
&\quad + H^{ft}\{(t_1+t_2)\delta_n\}(f_\varepsilon^{ft})'(t_1)f_\varepsilon^{ft}(t_2)\Big|^2 dt_1dt_2\\
&\le 2c^2\delta_n^{2\beta+3}\int_{\{t:\,1\le|t|\le 2\}}\int_{-\infty}^{\infty}\left|(H^{ft})'(t)f_\varepsilon^{ft}(\tau)f_\varepsilon^{ft}(\frac{t}{\delta_n}-\tau)\right|^2 d\tau dt\\
&\quad + 2c^2\delta_n^{2\beta+1}\int_{\{t:\,1\le|t|\le 2\}}\int_{-\infty}^{\infty}\left|H^{ft}(t)(f_\varepsilon^{ft})'(\tau)f_\varepsilon^{ft}(\frac{t}{\delta_n}-\tau)\right|^2 d\tau dt\\
&= O\left\{\delta_n^{2\beta+3}\int_{\{t:\,1\le|t|\le 2\}}\int_{-\infty}^{\infty}\left|f_\varepsilon^{ft}(\tau)f_\varepsilon^{ft}(\frac{t}{\delta_n}-\tau)\right|^2 d\tau dt\right\}\\
&\quad + O\left\{\delta_n^{2\beta+1}\int_{\{t:\,1\le|t|\le 2\}}\int_{-\infty}^{\infty}\left|(f_\varepsilon^{ft})'(\tau)f_\varepsilon^{ft}(\frac{t}{\delta_n}-\tau)\right|^2 d\tau dt\right\}\\
&= O\left(\delta_n^{2\beta+2\alpha+3}+\delta_n^{2\beta+2\alpha+1}\right) = O\left(n^{-1}\right).
\end{aligned}
\tag{7.6}
$$

Analogously,

$$I_3 = O\left(n^{-1}\right). \tag{7.7}$$

Finally, since $yz\{g_1(y,z) - g_0(y,z)\}$ has Fourier transform $c\delta_n^{\beta+1}i^{-2}$ $(\partial^2/\partial t_1\partial t_2)H^{ft}\{(t_1+t_2)\delta_n\}f_\varepsilon^{ft}(t_1)f_\varepsilon^{ft}(t_2)$, we obtain in complete analogy to (7.6) that

$$
\begin{aligned}
I_4 &= O\left\{\delta_n^{2\beta+5}\int_{\{t:\,1\le|t|\le 2\}}\int_{-\infty}^{\infty}\left|f_\varepsilon^{ft}(\tau)f_\varepsilon^{ft}(\frac{t}{\delta_n}-\tau)\right|^2 d\tau dt\right\}\\
&\quad + O\left\{\delta_n^{2\beta+3}\int_{\{t:\,1\le|t|\le 2\}}\int_{-\infty}^{\infty}\left|(f_\varepsilon^{ft})'(\tau)f_\varepsilon^{ft}(\frac{t}{\delta_n}-\tau)\right|^2 d\tau dt\right\}\\
&\quad + O\left\{\delta_n^{2\beta+1}\int_{\{t:\,1\le|t|\le 2\}}\int_{-\infty}^{\infty}\left|(f_\varepsilon^{ft})'(\tau)(f_\varepsilon^{ft})'(\frac{t}{\delta_n}-\tau)\right|^2 d\tau dt\right\}\\
&= O\left(n^{-1}\right),
\end{aligned}
\tag{7.8}
$$

which completes this proof.

Proof of Theorem 5.1. Let $g_\theta^{(n)}(\boldsymbol{y}) = \prod_{j=1}^n g_\theta(y_j)$, $\theta = 0,1$, be the two possible densities of $(Y_{1,1}, Y_{1,2}\ldots, Y_{n,1}, Y_{n,2})'$, where $\boldsymbol{y} = (y_1,\ldots,y_n)'$. It follows from Lemma 5.1 and standard arguments that

$$\int_{\mathbb{R}^{2n}} \min\{g_0^{(n)}(\boldsymbol{y}), g_1^{(n)}(\boldsymbol{y})\} d\boldsymbol{y} \geq C, \tag{7.9}$$

for some $C > 0$. Therefore, for any estimator $\widetilde{f}_{X,n}^{(n)}(0)$ of $f_{X,\theta}^{(l)}(0)$,

$$\begin{aligned}
&\max_{\theta\in\{0,1\}} \mathbb{E}_\theta \left| \widetilde{f}_{X,n}^{(n)}(0) - f_{X,\theta}^{(l)}(0) \right|^2 \\
&\quad \geq \frac{1}{2} \int_{\mathbb{R}^{2n}} \Big\{ \big| \widetilde{f}_{X,n}^{(n)}(0;\boldsymbol{y}) - f_{X,0}^{(l)}(0) \big|^2 \\
&\qquad\qquad + \big| \widetilde{f}_{X,n}^{(n)}(0;\boldsymbol{y}) - f_{X,1}^{(l)}(0) \big|^2 \Big\} \min\{g_0^{(n)}(\boldsymbol{y}), g_1^{(n)}(\boldsymbol{y})\} d\boldsymbol{y} \\
&\quad \geq \frac{1}{4} \left| f_{X,0}^{(l)}(0) - f_{X,1}^{(l)}(0) \right|^2 \int_{\mathbb{R}^{2n}} \min\{g_0^{(n)}(\boldsymbol{y}), g_1^{(n)}(\boldsymbol{y})\} d\boldsymbol{y}.
\end{aligned}$$

Since, by construction, $|f_{X,0}^{(l)}(0) - f_{X,1}^{(l)}(0)| \geq C n^{-(\beta-l)/(2\beta+2\alpha+1)}$, we obtain assertion (i).

In complete analogy to the above calculations we can show that

$$\max_{\theta\in\{0,1\}} \mathbb{E}_\theta \left(\widetilde{P}_{a,b,n} - P_{a,b,\theta} \right)^2 \geq \frac{1}{4} \left(P_{a,b,0} - P_{a,b,1} \right)^2 \int_{\mathbb{R}^{2n}} \min\{g_0^{(n)}(\boldsymbol{y}), g_1^{(n)}(\boldsymbol{y})\} d\boldsymbol{y},$$

which proves (ii).

Proof of Theorem 5.2. Denote by $g_{\boldsymbol{\theta}}^{(n)}(\boldsymbol{y}) = \prod_{j=1}^n g_{\boldsymbol{\theta}}(y_j)$, $\boldsymbol{\theta} \in \{0,1\}^{M_n}$, the possible densities of $(Y_{1,1}, Y_{1,2}, \ldots, Y_{n,1}, Y_{n,2})'$. We obtain from Lemma 5.2 that

$$\min_{\boldsymbol{\theta}\in\{0,1\}^{M_n}, 1\leq j\leq M_n} \int_{\mathbb{R}^{2n}} \min\{g^{(n)}_{(\theta_1,\ldots,\theta_{M_n})}(\boldsymbol{y}), g^{(n)}_{(\theta_1,\ldots,\theta_{j-1},1-\theta_j,\theta_{j+1},\ldots,\theta_{M_n})}(\boldsymbol{y})\} d\boldsymbol{y} \geq C,$$

for some $C > 0$. Now for any estimator $\widetilde{f}_{X,n}^{(l)}$ of $f_X^{(l)}$,

$$\begin{aligned}
&\max_{\boldsymbol{\theta}\in\{0,1\}^{M_n}} \mathbb{E}_{\boldsymbol{\theta}} \left\| \widetilde{f}_{X,n}^{(l)} - f_X^{(l)} \right\|_{L_2(\mathbb{R})}^2 \\
&\geq \sum_{j=1}^{M_n} \int_{(j-1)\delta_n}^{j\delta_n} \mathbb{E}_{\boldsymbol{\theta}} \left| \widetilde{f}_{X,n}^{(l)}(x) - f_{X,\boldsymbol{\theta}}^{(l)}(x) \right|^2 dx \\
&\geq \sum_{j=1}^{M_n} \inf_{(\theta_1,\ldots,\theta_{j-1},\theta_{j+1},\ldots,\theta_{M_n})\in\{0,1\}^{M_n-1}} \\
&\quad \frac{1}{2} \Bigg\{ \int_{(j-1)\delta_n}^{j\delta_n} \mathbb{E}_{(\theta_1,\ldots,\theta_{j-1},0,\theta_{j+1},\ldots,\theta_{M_n})} \left| \widetilde{f}_{X,n}^{(l)}(x) - f^{(l)}_{X,(\theta_1,\ldots,\theta_{j-1},0,\theta_{j+1},\ldots,\theta_{M_n})}(x) \right|^2 dx
\end{aligned}$$

$$+\int_{(j-1)\delta_n}^{j\delta_n} \mathbb{E}_{(\theta_1,\ldots,\theta_{j-1},1,\theta_{j+1},\ldots,\theta_{M_n})} \left|\widetilde{f}_{X,n}^{(l)}(x) - f_{X,(\theta_1,\ldots,\theta_{j-1},1,\theta_{j+1},\ldots,\theta_{M_n})}^{(l)}(x)\right|^2 dx \Bigg\}$$
$$\geq Cn^{-2(\beta-l)/(2\beta+2\alpha+1)}.$$

Proof of Theorem 5.3. We have

$$\max_{f_X\in\{f_n,\tilde{f}_n\}} \mathbb{E}\Big\{\lambda\big(\widetilde{S}_{X,n}\Delta S_X\big)\Big\}$$
$$\geq \frac{1}{2}\lambda\big(S_n\Delta\widetilde{S}_n\big)\int\cdots\int \min\{g(y_1)\cdots g(y_n)\,,\,\tilde{g}(y_1)\cdots\tilde{g}(y_n)\}dy_1\cdots dy_n\,,$$

where S_n and $\widetilde{S}_n$ denote the intersection of $[0,1]$ with the support of f_n and $\tilde{f}_n$, respectively, and $g(y_{j,1},y_{j,2}) = \int f_n(x)f_\varepsilon(y_{j,1}-x)f_\varepsilon(y_{j,2}-x)dx$, with $\tilde{g}$ obtained by replacing f_n by $\tilde{f}_n$. Using LeCam's inequality (see e.g., Devroye (1987, p.7)), we obtain that

$$\max_{f_X\in\{f_n,\tilde{f}_n\}} \mathbb{E}\Big\{\lambda\big(\widehat{S}_{X,n}\Delta S_X\big)\Big\} \geq \text{const.}\,\lambda\big(S_n\Delta\tilde{S}_n\big) \geq \text{const.}\,\varepsilon_n. \tag{7.10}$$

if the Hellinger distance $\mathrm{He}\,(g,\tilde{g})$ satisfies

$$\mathrm{He}^2(g,\tilde{g}) = O(n^{-1}). \tag{7.11}$$

To show (7.11), we apply **(A2)**, part (ii), to get

$$\begin{aligned} g(y) &\geq C^2\int_{\{x:\,|x|\leq 1\}} f_n(x)\big\{1+(y_1-x)^2\big\}^{-1}\big\{1+(y_2-x)^2\big\}^{-1}dx \\ &\geq \text{const.}\int_{\{x:\,|x|\leq 1\}} f_n(x)\big(1+2y_1^2+2x^2\big)^{-1}\big(1+2x^2+2y_2^2\big)^{-1}dx \\ &\geq \text{const.}\,\big(1+y_1^2\big)^{-1}\big(1+y_2^2\big)^{-1}\,, \end{aligned}$$

leading to

$$\begin{aligned} &\mathrm{He}^2(g,\tilde{g}) \\ &\leq \int_{\{y:\,g(y)>0\}} \big|g(y)-\tilde{g}(y)\big|^2[g(y)]^{-1}dy \\ &\leq \text{const.}\int_{\mathbb{R}^2}\big|g(y)-\tilde{g}(y)\big|^2(1+y_1^2+y_2^2+y_1^2y_2^2)dy \\ &\leq \text{const.}\max_{(j_1,j_2)\in\{0,1\}^2}\int_{\mathbb{R}^2}\Big|\frac{\partial^{j_1+j_2}}{\partial t_1^{j_1}\partial t_2^{j_2}}g^{ft}(t) - \frac{\partial^{j_1+j_2}}{\partial t_1^{j_1}\partial t_2^{j_2}}\tilde{g}^{ft}(t)\Big|^2 dt \\ &\leq O(\varepsilon_n^2)\max_{(j_1,j_2)\in\{0,1\}^2,\,j_3\in\{0,1,2\}} \end{aligned}$$

$$
\begin{aligned}
&\int_{\mathbb{R}^2} \Big|(Q^{ft})^{(j_3)}\big\{\varepsilon_n(t_1+t_2+\frac{2\pi}{\varepsilon_n})\big\}(f_\varepsilon^{ft})^{(j_1)}(t_1)(f_\varepsilon^{ft})^{(j_2)}(t_2)\Big|^2 dt \\
&+ O(\varepsilon_n^2) \max_{(j_1,j_2)\in\{0,1\}^2, j_3\in\{0,1,2\}} \\
&\int_{\mathbb{R}^2} \Big|(Q^{ft})^{(j_3)}\big\{\varepsilon_n(t_1+t_2-\frac{2\pi}{\varepsilon_n})\big\}(f_\varepsilon^{ft})^{(j_1)}(t_1)(f_\varepsilon^{ft})^{(j_2)}(t_2)\Big|^2 dt \\
\le\ & O(\varepsilon_n^2) \max_{(j_1,j_2)\in\{0,1\}^2} \int_{[-2/\varepsilon_n\pm 2\pi/\varepsilon_n, 2/\varepsilon_n\pm 2\pi/\varepsilon_n]} \int \Big|(f_\varepsilon^{ft})^{(j_1)}(s)(f_\varepsilon^{ft})^{(j_2)}(t-s)\Big|^2 ds dt \\
=\ & O\big(\varepsilon_n^{1+2\alpha}\big),
\end{aligned}
$$

where we have used Parseval's identity and **(A2)**, part (i). Therefore, (7.11) is satisfied under the selection $\varepsilon_n = n^{-1/(2\alpha+1)}$, and we can establish the desired rate by (7.10).

Acknowledgement

We thank three anonymous referees and an associate editor for their helpful comments and suggestions.

References

Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: Sharp best order convergence estimates. *Sankhyā Ser. A* **50**, 381-393.

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., Maca, J. D. and Ruppert, D. (1999). Nonparametric regression in the presence of measurement error. *Biometrika* **86**, 541-554.

Cuevas, A. and Fraiman, R. (1997). A plug-in approach to support estimation. *Ann. Statist.* **25**, 2300-2312.

Cuevas, A., Manteiga, W. and Rodriguez Casal, A. (2006). Plug-in estimation of general level sets. *Austral. N. Z. J. Statist.* **48**, 7-19.

Delaigle, A. and Gijbels, I. (2006). Estimation of boundary and discontinuity points in deconvolution problems. *Statist. Sinica* **16**, 773-788.

Delaigle, A., Hall, P. and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.* **36**, 665-685.

Delaigle, A. and Meister, A. (2008). Density estimation with heteroscedastic error. *Bernoulli* **14**, 562-579.

Devroye, L. (1987). *A Course in Density Estimation.* Birkhäuser, Boston.

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

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

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

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

Goldenshluger, A. and Tsybakov, A. (2004). Estimating the endpoint of a distribution in presence of additive observation errors, *Statist. Probab. Lett.* **68**, 39-49.

Hall, P. and Lahiri, S.N. (2008). Estimation of distributions, moments and quantiles in deconvolution problems, *Ann. Statist.* **36**, 2110-2134.

Hall, P. and Meister, A. (2007). A ridge-parameter approach to deconvolution. *Ann. Statist.* **35**, 1535-1558.

Hall, P. and Yao, Q. (2003). Inference in components of variance models with low replication. *Ann. Statist.* **31**, 414-441.

Hartigan, J. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82**, 267-270.

Hesse, C. H. (1999). Data-driven deconvolution. *J. Nonparam. Statist.* **10**, 343-373.

Horowitz, J. L. and Markatou, M. (1996). Semiparametric estimation of regression models for panel data. *Rev. Econom. Stud.* **63**, 145-168.

Johnstone, I., Kerkyacharian, G., Picard, D. and Raimondo, M. (2004). Wavelet deconvolution in a periodic setting. *J. Roy. Statist. Soc. Ser. B* **66**, 547-573, with discussion pp. 627-652.

Kerkyacharian, G., Picard, D. and Raimondo, M. (2007). Adaptive boxcar deconvolution on full Lebesgue measure sets. *Statist. Sinica* **17**, 317-340.

Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivariate Anal.* **65**, 139-165.

Meister, A. (2005). Non-estimability in spite of identifiability in density deconvolution. *Math. Meth. Statist.* **14**, 479-487.

Meister, A. (2006). Support estimation via moment estimation in presence of noise. *Statistics* **40**, 259-275.

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

Müller, D. and Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86**, 738-746.

Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Stat.* **7**, 307-330.

Neumann, M. H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.* **98**, 1955-1968.

Nolan, D. (1991). The excess-mass ellipsoid. *J. Multivariate Anal.* **39**, 348-371.

Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters: An excess mass approach. *Ann. Statist.* **23**, 855-881.

R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`.

Schennach, S. M. (2004a). Estimation of nonlinear models with measurement error. *Econometrica* **72**, 33-75.

Schennach, S. M. (2004b). Nonparametric regression in the presence of measurement error. *Econom. Theory* **20**, 1046-1093.

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

Sun, J., Morrison, H., Harding, P. and Woodroofe, M. (2003). Mixtures and bumps: errors in measurement. Preprint, available at `http://sun.cwru.edu/~jiayang/india.ps`.

Yosida, K. (1968). *Functional Analysis.* Springer, Berlin.

Zhang, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806-830.

Institut für Mathematik, Universität Rostock, Ulmenstrasse 69, 18057 Rostock, Germany.

E-mail: alexander.meister@uni-rostock.de

Institut für Stochastik, Friedrich-Schiller-Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany.

E-mail: michael.neumann@uni-jena.de

(Received July 2008; accepted June 2009)