The Annals of Statistics
2017, Vol. 45, No. 4, 1516–1541
DOI: 10.1214/16-AOS1498

MINIMAX THEORY OF ESTIMATION OF LINEAR FUNCTIONALS OF THE DECONVOLUTION DENSITY WITH OR WITHOUT SPARSITY

BY MARIANNA PENSKY[1]

University of Central Florida

The present paper considers the problem of estimating a linear functional $\Phi = \int_{-\infty}^{\infty} \varphi(x) f(x)\, dx$ of an unknown deconvolution density f on the basis of n i.i.d. observations, $Y_1, \dots, Y_n$ of $Y = \theta + \xi$, where ξ has a known pdf g, and f is the pdf of θ. The objective of the present paper is to develop the a general minimax theory of estimating Φ, and to relate this problem to estimation of functionals $\Phi_n = n^{-1} \sum_{i=1}^{n} \varphi(\theta_i)$ in indirect observations. In particular, we offer a general, Fourier transform based approach to estimation of Φ (and Φ_n) and derive upper and minimax lower bounds for the risk for an arbitrary square integrable function φ. Furthermore, using technique of inversion formulas, we extend the theory to a number of situations when the Fourier transform of φ does not exist, but Φ can be presented as a functional of the Fourier transform of f and its derivatives. The latter enables us to construct minimax estimators of the functionals that have never been handled before such as the odd absolute moments and the generalized moments of the deconvolution density. Finally, we generalize our results to the situation when the vector $\boldsymbol{\theta}$ is sparse and the objective is estimating Φ (or Φ_n) over the nonzero components only. As a direct application of the proposed theory, we automatically recover multiple recent results and obtain a variety of new ones such as, for example, estimation of the mixing probability density function with classical and Berkson errors and estimation of the $(2M + 1)$-th absolute moment of the deconvolution density.

1. Introduction.

In the present paper, we consider the problem of estimating a linear functional

$$\Phi = \int_{-\infty}^{\infty} \varphi(x) f(x)\, dx \tag{1.1}$$

of an unknown deconvolution density f on the basis of observations

$$Y_i = \theta_i + \xi_i, \qquad i = 1, \dots, n. \tag{1.2}$$

Here, θ_i are i.i.d. random variables with unknown pdf f, and ξ_i are i.i.d. random errors with a known pdf g. The density f is sometimes referred to as the *mixing*

Received May 2015; revised July 2016.
[1]Supported in part by NSF Grants DMS-11-06564 and DMS-14-07475.
MSC2010 subject classifications. Primary 62G20; secondary 62G05, 62G07.
Key words and phrases. Linear functional, minimax lower bound, sparsity, deconvolution.

density. The vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ in (1.2) may be sparse in the sense that, on the average, it has only $n\mu_n$ nonzero elements where $\mu_n \to 0$ as $n \to \infty$.

Note that the problem of estimating Φ in (1.1) appears in many contexts. If $\varphi(x) = \delta(x - x_0)$, then Φ is the value of the unknown deconvolution density f at the point x_0. If $\varphi(x) = \mathbb{I}(x < x_0)$, where $\mathbb{I}(\Omega)$ denotes the indicator of a set Ω, then problem (1.1) reduces to estimation of the mixing distribution function $\Phi = F(x_0)$ at x_0, examined by Dattner, Goldenshluger and Juditsky (2011). If $\varphi(x) = e^{i\omega_0 x}$, then $\Phi = \widehat{f}(\omega_0)$, the characteristic function of the mixing distribution at $\omega = \omega_0$. If $\varphi(x) = f_\eta(x - x_0)$, where f_η is a known pdf, then $\Phi = \Phi(x_0)$ is itself a convolution density at a point x_0, as considered by Delaigle (2007). Finally, if $\varphi(x) = x^k$ or $\varphi(x) = |x|^{2M+1}$, then Φ is, respectively, the kth moment or the $(2M + 1)$-th absolute moment of f.

In addition, the problem of estimating Φ in (1.1) can be related to estimating of functionals in indirect observations

$$\Phi_n = \frac{1}{n} \sum_{i=1}^{n} \varphi(\theta_i). \tag{1.3}$$

Indeed, if θ_i, $i = 1, \dots, n$ are i.i.d. with the pdf f and $\mathbb{E}|\varphi(\theta)| < \infty$, then Φ_n in (1.3) can be viewed as an "estimator" of $\Phi = \mathbb{E}\varphi(\theta)$ in (1.1) on the basis of "observations" θ_i. Moreover, as long as $\mathbb{E}|\varphi(\theta)|^2 < \infty$, one has $\mathbb{E}(\Phi_n - \Phi)^2 \le n^{-1}\mathbb{E}|\varphi(\theta)|^2$. Hence, the minimax risks for estimating Φ_n and Φ are equivalent up to the Cn^{-1} additive term. Therefore, the upper bounds and the minimax lower bounds for the risks of both estimators will coincide up to, at most, a constant factor.

In the case when the vector $\boldsymbol{\theta}$ is nonrandom and sparse, so that it has only k_n nonzero components, observations (1.2) can be viewed as heterogeneous sparse mixture, which was studied by a number of authors [see, e.g., Donoho and Jin (2004), Cai, Jin and Low (2007), Hall and Jin (2010) among others]. If the signal is present ($k_n > 0$) and k_n is known, the problem of interest is to estimate some characteristics of nonzero elements of $\boldsymbol{\theta}$. The latter problem can be summarized as the problem of estimating

$$\Phi_{k_n} = \frac{1}{k_n} \sum_{i=1}^{n} \varphi(\theta_i)\mathbb{I}(\theta_i \neq 0). \tag{1.4}$$

Cai and Low (2011) considered estimation of (1.4) when $\varphi(x) = |x|$, the errors are Gaussian and $k_n = Cn^\nu$, $0 < \nu < 1$. They concluded that consistent estimation is impossible if $\nu \le 1/2$. A question of interest is whether the same will happen in general, or whether this phenomenon is due to the type of the functional (the first absolute moment) or the type of errors (Gaussian) studied in the paper. Note that, in the case when components of $\boldsymbol{\theta}$ are i.i.d. with the pdf f, then the random number of nonzero components k_n is Binomial(n, μ_n) and f can be written as $f(x) = \mu_n f_0(x) + (1 - \mu_n)\delta(x)$, where $f_0(\theta)$ is pdf of the nonzero entries of $\boldsymbol{\theta}$

and $\mu_n = n^{\nu-1}$. If n is large enough, then with high probability k_n is close to $n\mu_n$ and Φ_{k_n} in (1.4) corresponds to

$$\Phi_\mu = \int_{-\infty}^{\infty} \varphi(x) f_0(x)\, dx. \tag{1.5}$$

We propose a general procedure designed for estimating functionals (1.4) and (1.5) in the sparse case for any function $\varphi(x)$ and any kind of error density g. We discover that when $n\mu_n = n^\nu$, then convergence rates are determined by the "effective" sample size $n{\mu_n}^2 = n^{2\nu-1}$. The latter proves that if θ_i are i.i.d, then the conclusion of Cai and Low (2011) that consistent estimation is impossible whenever $\nu \le 1/2$ applies not only to their particular case [$\varphi(x) = |x|$, Gaussian errors] but to any functional and any distribution of errors.

In spite of its great importance, surprisingly, the general problem of estimation of a linear functional of the deconvolution density has not been thoroughly investigated. In the nonsparse case, the problem of estimation of the linear functional (1.1) of the mixing density with a square integrable function φ has been addressed by Butucea and Comte (2009) who derived the upper bounds for the mean squared risk for a variety of estimation scenarios, and constructed adaptive estimators that attain them (up to, at most, a logarithmic factor). However, minimax lower risk bounds have been derived only in the case when $\varphi(x) = \delta(x - x_0)$ due to technical difficulties. Moreover, the general study of Butucea and Comte (2009) does not allow one to apply the theory when the Fourier transform of $\varphi(\theta)$ does not exist. Furthermore, as far as we know, there has never been a study of estimation of a linear functional of the deconvolution density in the sparse setting.

The purpose of the present paper is to fill in the existing gaps and to advance the theory of estimation of linear functionals of the deconvolution density. In particular, the paper accomplishes several key goals: (a) derivation of minimax lower bound for the risk of an estimator of a general linear functional of the deconvolution density; (b) estimation of linear functionals (1.1) when function φ is not necessarily integrable or square integrable using inversion formulas; (c) application of those methodologies to estimation of functionals of the form (1.3); (d) estimation of functionals of the form (1.5) [or (1.4)] in the case when deconvolution density f (or vector $\boldsymbol{\theta}$) is sparse.

The rest of the paper is organized as follows. In Section 2, we study the case when the Fourier transform of function $\varphi(x)$ in (1.1) exists. We refer to this situation as the *standard case* in comparison with the situations considered in Section 3 where Φ is represented as a linear functional of Fourier transform f^* of f using some inversion formula. We establish the minimax lower bounds for the risk for a general function $\varphi(x)$ and compare them with the upper bounds for the risk in the case when f belongs to a Sobolev ball, thus, completing the theory of Butucea and Comte (2009). In Section 3, we expand our approach to incorporate estimation of functionals of the form (1.1) where $\varphi(x)$ does not have the Fourier transform but functional Φ can be represented via the Fourier transform of the deconvolution density or its derivatives, using technique of inversion formulas. Section 4 deals with the situation where $\boldsymbol{\theta}$ or the deconvolution density f is sparse. Section 5 contains a brief description of a finite sample simulation study of estimation of the first absolute moment of the mixing density. The complete report of the simulation study can be found in Section 7.1 of the supplementary material [Pensky (2017)]. Section 5 contains some supplementary statements and some essential proofs. The rest of the proofs have been placed into the supplementary material.

2. The minimax upper and lower bounds for the risk: The standard case. In Section 2, we assume that the functional Φ in (1.1) can be represented as

$$\Phi = \frac{1}{2\pi}\int_{-\infty}^{\infty} f^*(\omega)\varphi^*(-\omega)\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{q^*(\omega)}{g^*(\omega)}\varphi^*(-\omega)\,d\omega, \tag{2.1}$$

where the integral is absolutely convergent. This happens if, for example, $a > 1$ in (2.5), so that $|\varphi^*|$ and $|\varphi|$ are square integrable, however, this is true for a wider variety of functions φ [e.g., $\varphi(x) = \delta(x - x_0)$ considered in Butucea and Comte (2009)]. We refer to this situation as the *standard case* in comparison with the situations considered in Section 3 where Φ cannot be represented in the form (2.1).

2.1. *Notation and assumptions.* For a function $t(x)$, we denote its Fourier transform by

$$t^*(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} t(x)\,dx.$$

Denote the pdf of Y_i by $q(y)$, so that $q^*(\omega) = f^*(\omega)g^*(\omega)$. We assume that the mixing density belongs to the Sobolev ball $f \in \Omega_s(B)$ where

$$\Omega_s(B) = \left\{t^* : \int_{-\infty}^{\infty} \left|t^*(\omega)\right|^2 \left(\omega^2+1\right)^s d\omega \le B^2\right\}, \qquad s \ge 0. \tag{2.2}$$

We introduce the following assumptions on the known functions g and φ.

A1. There exist nonnegative constants C_{g1}, C_{g2}, α, β and γ such that

$$\left|g^*(\omega)\right| \ge C_{g1}\left(\omega^2+1\right)^{-\alpha/2} \exp\left(-\gamma|\omega|^{\beta}\right), \tag{2.3}$$

$$\left|g^*(\omega)\right| \le C_{g2}\left(\omega^2+1\right)^{-\alpha/2} \exp\left(-\gamma|\omega|^{\beta}\right), \tag{2.4}$$

where $\alpha > 0$ and $\beta = 0$ if and only if $\gamma = 0$.

A2. There exist nonnegative constants $C_{\varphi1}$, $C_{\varphi2}$, a, b and d such that

$$\left|\varphi^*(\omega)\right| \ge C_{\varphi1}\left(\omega^2+1\right)^{-a/2} \exp\left(-d|\omega|^{b}\right), \tag{2.5}$$

$$\left|\varphi^*(\omega)\right| \le C_{\varphi2}\left(\omega^2+1\right)^{-a/2} \exp\left(-d|\omega|^{b}\right), \tag{2.6}$$

where $b = 0$ if and only if $d = 0$.

In what follows, we use the symbol C for a generic positive constant, which takes different values at different places and is independent of n. Also, for any positive functions $a(n)$ and $b(n)$, we write $a(n) \asymp b(n)$ if the ratio $a(n)/b(n)$ is bounded above and below by finite positive constants independent of n.

2.2. *The lower bounds for the risk.* We start with the derivation of the minimax lower bounds for the risk of a general linear functional which, to the best of our knowledge, have not been obtained so far. In our derivation we follow Tsybakov (2009). Butucea and Comte (2009) derived those lower bounds only in the simple case when $|\varphi^*(\omega)| = 1$. Denote

$$R_n^{low} = R_n\big(\Omega_s(B)\big) = \inf_{\widetilde{\Phi}} \sup_{f \in \Omega_s(B)} \mathbb{E}(\widetilde{\Phi} - \Phi)^2, \tag{2.7}$$

where $\widetilde{\Phi}$ is any estimator of Φ based on observations $Y_1, \ldots, Y_n$. Then the following theorem is true.

THEOREM 1. *Let g be bounded above and such that g^* is differentiable and*

$$\frac{|(g^*)'(\omega)|}{|g^*(\omega)|} \le C_g\big(1 + |\omega|\big)^{\tau}, \qquad \tau \ge 0, \text{ with } \tau = 0 \text{ if } \gamma = 0. \tag{2.8}$$

Let there exist $\omega_0 \in (0, \infty)$ such that, for $|\omega| > \omega_0$, $\rho(\omega) = \arg(\varphi^*(\omega))$ is twice continuously differentiable with $|\rho^{(j)}(\omega)| \le \rho < \infty$, $j = 0, 1, 2$. Then, under Assumptions A1 *and* A2 [*inequalities* (2.4) *and* (2.5) *only*], *one has the lower bounds for the risk given in Table* 1 *with $U_1 = 2a + 2s(1 - d/\gamma) - 2d\alpha/\gamma + 8\beta - (U_\tau + 1)d/\gamma - 1$, $U_2 = 2a + 2s + 8b - 1$ and $U_\tau = \min(7\beta - 2\tau - 1, 5\beta + 1)$.*

Since any estimator $\widetilde{\Phi}_n$ of Φ_n defined in (1.3) can be viewed as an estimator of Φ, due to inequality

$$\mathbb{E}(\widetilde{\Phi}_n - \Phi_n)^2 \ge 0.5\mathbb{E}(\widetilde{\Phi}_n - \Phi)^2 - 2n^{-1}\|\varphi\|_\infty^2,$$

Theorem 1 immediately provides the lower bounds for the risk of any estimator $\widetilde{\Phi}_n$ of Φ_n based on $Y_1, \ldots, Y_n$.

COROLLARY 1. *Let θ_i, $i = 1, \ldots, n$, in* (1.2) *be i.i.d. with pdf f. If $\varphi(\theta)$ is uniformly bounded $|\varphi(\theta)| \le \|\varphi\|_\infty < \infty$, then under assumptions of Theorem* 1, *for sufficiently large n, one has*

$$R_n\big(\Phi_n, \Omega_s(B)\big) = \inf_{\widetilde{\Phi}_n} \sup_{f \in \Omega_s(B)} \mathbb{E}(\widetilde{\Phi}_n - \Phi_n)^2 \ge C R_n\big(\Omega_s(B)\big), \tag{2.9}$$

where $R_n(\Omega_s(B))$ is given in Theorem 1.

TABLE 1

Asymptotic expressions for the minimax lower bounds R_n^{low} and the upper bounds for the risks R_n^{up} of nonadaptive and $\widehat{R_n^{up}}$ of adaptive estimators

	Cases	Bounds for the risk
1	$b > \beta$	$R_n^{low} \asymp R_n^{up} \asymp \widehat{R_n^{up}} \asymp n^{-1}$
2	$b = \beta, d > \gamma > 0$	$R_n^{low} \asymp R_n^{up} \asymp \widehat{R_n^{up}} \asymp n^{-1}$
3	$b = \beta, d = \gamma, a > \alpha + 1/2$	$R_n^{low} \asymp R_n^{up} \asymp \widehat{R_n^{up}} \asymp n^{-1}$
4	$b = \beta > 0, d = \gamma > 0, a = \alpha + 1/2$	$R_n^{low} \asymp n^{-1}$, $R_n^{up} \asymp \widehat{R_n^{up}} \asymp R_n^{low} \cdot \log\log n$
5	$b = \beta = 0, d = \gamma = 0, a = \alpha + 1/2$	$R_n^{low} \asymp n^{-1}$, $R_n^{up} \asymp \widehat{R_n^{up}} \asymp R_n^{low} \cdot \log n$
6	$b = \beta > 0, d = \gamma > 0, a < \alpha + 1/2$	$R_n^{low} \asymp n^{-1}$, $R_n^{up} \asymp \widehat{R_n^{up}} \asymp R_n^{low} \cdot (\log n)^{\frac{2\alpha-2a+1}{\beta}}$
7	$b = \beta = 0, d = \gamma = 0, a < \alpha + 1/2$	$R_n^{low} \asymp R_n^{up} \asymp n^{-\frac{2s+2a-1}{2s+2\alpha}}$, $\widehat{R_n^{up}} \asymp R_n^{low} \cdot \log n$
8	$b = \beta > 0, \gamma > d > 0$	$R_n^{low} \asymp (\log n)^{-\frac{U_1}{\beta}} n^{-d/\gamma}$, $R_n^{up} \asymp R_n^{low} \cdot (\log n)^{\frac{U_1-U_3}{\beta}}$
		$\widehat{R_n^{up}} \asymp R_n^{up} \cdot (\log n)^{\frac{\Delta U_3}{\beta}}$
9	$\beta > b > 0, d > 0, \gamma > 0$	$R_n^{low} \asymp (\log n)^{-\frac{U_2}{\beta}} \exp(-2d[\frac{\log n}{2\gamma}]^{b/\beta})$,
		$R_n^{up} \asymp R_n^{low} \cdot (\log n)^{\frac{U_2-U_4}{\beta}}$, $\widehat{R_n^{up}} \asymp R_n^{up} \cdot (\log n)^{\frac{\Delta U_4}{\beta}}$
10	$b = d = 0, \beta > 0, \gamma > 0$	$R_n^{low} \asymp R_n^{up} \asymp \widehat{R_n^{up}} \asymp (\log n)^{-\frac{2s+2a-1}{\beta}}$

2.3. *Estimation and the upper bounds for the risk.* Following Butucea and Comte (2009), we estimate Φ in (2.1) by

$$\widehat{\Phi}_h = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\widehat{q^*}(\omega)}{g^*(\omega)} \varphi^*(-\omega) \mathbb{I}\big(|\omega| \le h^{-1}\big)\, d\omega, \tag{2.10}$$

where

$$\widehat{q^*}(\omega) = n^{-1} \sum_{j=1}^{n} e^{i\omega Y_j} \tag{2.11}$$

is the unbiased estimator of $q^*(\omega)$ and $h = 0$ if function $|\varphi^*(\omega)|/|g^*(\omega)|$ has finite L^2-norm. In particular, the upper bound for the risk of the estimator $\widehat{\Phi}_h$ over the Sobolev class $\Omega_s(B)$

$$R_n\big(\widehat{\Phi}_h, \Omega_s(B)\big) = \sup_{f \in \Omega_s(B)} \mathbb{E}(\widehat{\Phi}_h - \Phi)^2$$

is given by the following inequality:

$$\begin{aligned} R_n\big(\widehat{\Phi}_h, \Omega_s(B)\big) \le{}& \frac{\|g\|_\infty}{2\pi n} \int_{-h^{-1}}^{h^{-1}} \frac{|\varphi^*(\omega)|^2}{|g^*(\omega)|^2}\, d\omega \\ &+ \frac{B^2}{4\pi^2} \int_{-\infty}^{\infty} \frac{|\varphi^*(\omega)|^2}{(\omega^2+1)^s} \mathbb{I}\big(|\omega| > h^{-1}\big)\, d\omega, \end{aligned} \tag{2.12}$$

where $\|g\|_\infty = \sup_x g(x)$.

THEOREM 2. *If g is bounded above, then under Assumptions* A1 *and* A2 [*inequalities* (2.3) *and* (2.6) *only*], *one derives the upper bounds for the risk* $R_n^{up} = R_n(\widehat{\Phi}_{h_n}, \Omega_s(B))$ *provided in Table* 1, *where* $U_\tau = \min(7\beta - 2\tau - 1, 5\beta + 1)$, $U_1 = 2a + 2s(1 - d/\gamma) - 2d\alpha/\gamma + 8\beta - (U_\tau + 1)d/\gamma - 1$, $U_2 = 2a + 2s + 8b - 1$, $U_3 = (2s + 2a + b - 1) - d(2\alpha + 2s)/\gamma$ *and* $U_4 = 2s + 2a + b - 1$. *The corresponding values of* $h = h_n$ *are* $h_n = 0$ *for cases* 1, 2 *and* 3; $h_n = n^{-1}$ *for case* 5; $h_n = n^{-\frac{1}{2s+2\alpha}}$ *for case* 7; $h_n = \{[\log n - (t/\beta)\log\log n]/(2\gamma)\}^{-\frac{1}{\beta}}$ *with* $t = 0$ *for cases* 4 *and* 6; $t = 2s + 2\alpha$ *for case* 8, $t = 2s + 2\alpha + b - \beta$ *for case* 9 *and* $h_n = [\log n/(3\gamma)]^{-\frac{1}{\beta}}$ *for case* 10.

Theorem 2 allows us to derive upper bounds for the risk of $\widehat{\Phi}_{\tilde{h}_n}$ when it is used as an estimator of the functional Φ_n defined in (1.3).

COROLLARY 2. *Let* θ_i, $i = 1, \ldots, n$, *in* (1.2) *be i.i.d. with the pdf* f. *If* $\varphi(\theta)$ *is uniformly bounded* $|\varphi(\theta)| \le \|\varphi\|_\infty < \infty$, *then under assumptions of Theorem* 2, *one has* $R_n(\widehat{\Phi}_{\tilde{h}_n}, \Phi_n, \Omega_s(B)) = \sup_{f \in \Omega_s(B)} \mathbb{E}(\widehat{\Phi}_{\tilde{h}_n} - \Phi_n)^2 \le 2R_n(\widehat{\Phi}_{\tilde{h}_n}, \Omega_s(B)) + 2n^{-1}\|\varphi\|_\infty^2 \le CR_n(\widehat{\Phi}_{\tilde{h}_n}, \Omega_s(B))$, *where* $R_n(\widehat{\Phi}_{\tilde{h}_n}, \Omega_s(B))$ *is provided in Theorem* 2.

2.4. *Adaptive estimation*. Note that in the expressions for the optimal value of bandwidth $\tilde{h}_n$ in Theorem 2, parameters a, α, b, d, β and γ are known; the only unknown parameters are s and B. Hence, the only cases for which one needs an adaptive choice of bandwidth are the cases where $\tilde{h}_n$ depends on s. This occurs only in cases 7, 8 and 9. In cases 8 and 9, one can easily provide an alternative value $h = \widehat{h}_n$ that introduces an additional logarithmic factor of n into the expression of the risk. In case 7, one can use the Lepskii method for construction of $\widehat{h}_n$ [see, e.g., Lepskiĭ (1991), Lepski, Mammen and Spokoiny (1997)]. In order to apply the method, consider the set of bandwidths

$$\mathcal{H} = \{h_j = n^{-\frac{1}{2\alpha}} 2^{-j}, j = 0, \ldots, J\} \qquad \text{with } 2^J \le (\log n)^{-1} n^{\frac{2\alpha-1}{2\alpha}}, \tag{2.13}$$

where J is the largest positive integer satisfying inequality above. Denote

$$\widehat{j} = \min\left\{ j : 0 \le j \le J; |\widehat{\Phi}_{h_j} - \widehat{\Phi}_{h_k}| \le \frac{C_\Phi \sqrt{\log n}}{\sqrt{n h_k^{2\alpha - 2a + 1}}}, \forall k, j \le k \le J \right\}, \tag{2.14}$$

where C_Φ is such that

$$C_\Phi \ge 4 \max\left\{ \frac{C_{\varphi 2}}{\pi \sqrt{\log n}}; \frac{C_{\varphi 2}}{C_{g1}} \left(\frac{16}{3\pi} + \frac{2\sqrt{\|g\|_\infty}}{\sqrt{\pi}} \right) \right\} \tag{2.15}$$

and constants $C_{\varphi 2}$ and C_{g1} appear in (2.6) and (2.3), respectively. The following statement provides the minimax upper bounds for the risk when the bandwidth $h = \widehat{h}_n$ is chosen adaptively, without the knowledge of the parameters B and s.

THEOREM 3. *If g is bounded above, then under Assumptions* A1 *and* A2 [*inequalities* (2.3) *and* (2.6) *only*], *one obtains the expressions for* $\widehat{R_n^{up}} \equiv R_n(\widehat{\Phi}_{\widehat{h}}, \Omega_s(B))$ *provided in Table* 1 *with* $\Delta U_3 = 2s(\gamma - d)/\gamma$ *and* $\Delta U_4 = 2(s - s_0)_+$. *The corresponding values of* $h = \widehat{h}_n$ *are* $\widehat{h}_n = 0$ *for cases* 1, 2 *and* 3; $\widehat{h}_n = n^{-1}$ *for case* 5; $\widehat{h}_n = h_{\widehat{j}}$ *with* $\widehat{j}$ *defined in* (2.14) *with* C_Φ *given in* (2.15) *for case* 7; $\widehat{h}_n = \{[\log n - \frac{t}{\beta}\log\log n]/(2\gamma)\}^{-\frac{1}{\beta}}$ *with* $t = 0$ *for cases* 4 *and* 6; $t = 2\alpha$ *for case* 8, $t = 2s_0 + 2\alpha + b - \beta$ *for case* 9 *and and* $\widehat{h}_n = [\log n/(3\gamma)]^{-\frac{1}{\beta}}$ *for case* 10.

REMARK 1 [Comparison with Butucea and Comte (2009)]. Note that Butucea and Comte (2009) derived lower bounds for the risk only in the case when $\varphi(x) = \delta(x - x_0)$ which corresponds to $a = b = d = 0$ in Table 1, hence, comparison of the lower bounds is impossible. On the other hand, they considered a much wider variety of cases, since they assumed that f is such that $\int_{-\infty}^{\infty} |f^*(\omega)|^2(\omega^2 + 1)^s \exp\{2s_1|\omega|^{s_2}\}\, d\omega \le B^2$, while we consider only the case of $s_1 = s_2 = 0$. If $s_1 = s_2 = 0$, our upper bounds R_n^{up} exactly coincide with theirs, and $\widehat{R_n^{up}}$ coincide with theirs up to some logarithmic factors in cases 5, 6 and 7 and coincide exactly in all other cases. In addition, we examined the nonspecific term v_n used in Butucea and Comte (2009) and obtain exact results for the rates of convergence in cases 8 and 9.

As an example of application of the theory above, we solve the problem of pointwise estimation of the mixing density with classical and Berkson errors studied by Delaigle (2007).

2.5. *Pointwise estimation of the deconvolution density with classical and Berkson errors.* Consider the situation where one is interested in estimating the pdf f_ζ of the random variable $\zeta = \theta + \eta$ where θ and η are independent, the pdf f_η of η is known and one has measurements $Y_1, \dots, Y_n$ of random variable $Y = \theta + \xi$ of the form of (1.2) where the pdf g of ξ is known. The model was originally introduced by Berkson (1950) in the regression context and subsequently studied by Delaigle (2007) who obtained the upper bounds for the integrated mean squared risk. In particular, if the pdf $q(y)$ of Y is k times continuously differentiable and is such that q and $q^{(k+1)}$ are square integrable and $q^{(k+1)}$ is bounded, Delaigle (2007) derived an estimator $\widehat{f_\zeta}$ of f_ζ such that for $|\omega| \to \infty$

$$\mathbb{E}\|\widehat{f_\zeta} - f_\zeta\|^2 \le \begin{cases} Cn^{-\min[1, 2k/(2k+1+2b)]}, & \text{if } |f_\eta^*(\omega)/g^*(\omega)| \asymp |\omega|^b, \\ C(\log n)^{-2k/\beta}, & \text{if } |f_\eta^*(\omega)/g^*(\omega)| \asymp |\omega|^b \exp(\gamma|\omega|^\beta), \end{cases}$$

where $\|\cdot\|$ denotes the L^2-norm with respect to the Lebesgue measure and the constant C depends on the density f of each θ.

The theory developed in this paper allows one to construct an estimator of the pdf f_ζ at a point x_0 with no additional effort. Let, as before, f, g and q be the pdfs of θ, ξ and Y, respectively. Then $f_\zeta^* = f^* f_\eta^* = q^* f_\eta^*/g^*$ and

$$f_\zeta(x_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ix_0\omega}\frac{q^*(\omega) f_\eta^*(\omega)}{g^*(\omega)}\,d\omega.$$

Therefore, $\varphi^*(\omega) = e^{ix_0\omega} f_\eta^*(-\omega)$, so that $|\varphi^*(\omega)| = |f_\eta^*(\omega)|$. The estimator of $f_\zeta(x_0)$ is of the form (2.10) and Theorems 1 and 2 give the upper and the minimax lower bounds for the risk of estimating f_ζ at a point x_0. In addition, Theorem 3 provides an adaptive estimator of $f_\zeta(x_0)$ that, to the best of our knowledge, has not been derived so far.

Note that we obtain a wider variety of convergence rates here than Delaigle (2007) who recovered only parametric, polynomial (with $d = 0$) and logarithmic convergence rates. The latter is due to the fact that we impose assumptions on g^* and f_η^* separately while Delaigle (2007) considers only the cases when the absolute value of the ratio $|f_\eta^*/g^*|$ grows polynomially or exponentially as $|\omega| \to \infty$.

3. Estimation of linear functionals by using inversion formulas.

3.1. *Formulation and some inversion formulas.* Estimation of the linear functional Φ in (1.1) relies on the fact that $\varphi \in L^2(-\infty, \infty)$, so its Fourier transform exists. It is easy to see that this condition, however, is not necessary for consistent estimation of Φ. Consider, for example, estimation of $\Phi_m = \int_{-\infty}^{\infty} \theta^m f(\theta)\,d\theta$, the mth moment of $f(\theta)$. Note that if $\psi_m(y)$ is a solution of the equation

$$\int_{-\infty}^{\infty} g(y-\theta)\psi_m(y)\,dy = \theta^m \tag{3.1}$$

then $\Phi_m = \int_{-\infty}^{\infty} \psi_m(y) q(y)\,dy = \mathbb{E}[\psi_m(Y)]$. In order to construct $\psi_m(y)$ satisfying equation (3.1), denote

$$\mu_k = \int_{-\infty}^{\infty} \theta^k g(\theta)\,d\theta, \qquad \nu_k = \int_{-\infty}^{\infty} \theta^k f(\theta)\,d\theta. \tag{3.2}$$

Assume that $\mu_{2m} < \infty$ and $\nu_{2m} < \infty$. Let $c_m = 1$ and c_k, $k = 0, \ldots, m-1$, be solutions of the system of linear equations

$$\sum_{k=j}^{m} \mu_{k-j} c_k = 0, \qquad j = 0, 1, \ldots, m-1.$$

Then it is easy to check that $\psi_m(y) = y^m + \sum_{k=0}^{m-1} c_k y^k$, and under the assumption that $\mu_{2m} < \infty$ and $\nu_{2m} < \infty$, Φ_m can be estimated by $\widehat{\Phi}_m = n^{-1}\sum_{l=1}^{n} \psi_m(Y_l)$, where $\mathbb{E}\widehat{\Phi}_m = \Phi_m$, $\operatorname{Var}[\widehat{\Phi}_m] \le C_m n^{-1}$ and constant C_m depends only on m, μ_{2m} and ν_{2m}.

Note that although we did not use Fourier transform for estimation of Φ_m and Fourier transform of x^m does not exist in the usual sense, it does exist in a sense of generalized functions and is equal to $(-1)^m \delta^{(m)}(\omega)$ where $\delta^{(m)}(\omega)$ is the mth derivative of the Dirac delta function [see, e.g., Zayed (1996)]. However, using the Fourier transform of φ as a generalized function would require f to belong to a so-called test-function space. Those spaces are usually very restrictive, like, for example, commonly used for the Fourier transforms of generalized functions, the space of the Schwartz distributions which consists of all infinitely differentiable functions that vanish outside some compact set [see, e.g., Zayed (1996)]. One, of course, cannot expect the unknown density f to belong to such space and, moreover, this will make any minimax estimation totally irrelevant. For this reason, instead of using the theory of generalized functions, we shall use inversion formulas that mimic generalized functions but do not require restrictive assumptions on the unknown pdf f. Our goal is to represent the functionals of interest as integrals of the Fourier transform of f^* and its derivatives. Dattner, Goldenshluger and Juditsky (2011) used inversion formula of Gil-Pelaez (1951) for estimation of the cumulative distribution function at a point; nevertheless, there are many more possible applications of this technique which is based on the following lemma proved in Section A.2.

LEMMA 1. *If $u(\theta)$ is absolutely integrable, then*

$$\int_{-\infty}^{\infty} \operatorname{sign}(\theta - t) u(\theta)\, d\theta = \frac{2}{\pi} \int_0^{\infty} \frac{1}{\omega} \left[\int_{-\infty}^{\infty} \sin\bigl((\theta - t)\omega\bigr) u(\theta)\, d\theta \right] d\omega. \tag{3.3}$$

Below we consider several examples of applications of Lemma 1.

EXAMPLE 1 (Pointwise estimation of the deconvolution cumulative distribution function). Denote the real and the imaginary part of z by $\Re[z]$ and $\Im[z]$, respectively. Then due to the relation $\mathbb{I}(\theta \le t) = 1/2 - 1/2 \operatorname{sign}(\theta - t)$ and Lemma 1, the cdf $F(t)$ can be represented as $F(t) = 0.5 - 0.5\Phi(t)$ where

$$\begin{aligned} \Phi(t) &= \int_{-\infty}^{\infty} \operatorname{sign}(\theta - t) f(\theta)\, d\theta = \frac{2}{\pi} \int_0^{\infty} \frac{1}{\omega} \left[\int_{-\infty}^{\infty} \sin\bigl((\theta - t)\omega\bigr) f(\theta)\, d\theta \right] d\omega \\ &= \frac{2}{\pi} \int_0^{\infty} \frac{\cos(t\omega)}{\omega} \Im[f^*(\omega)]\, d\omega - \frac{2}{\pi} \int_0^{\infty} \frac{\sin(t\omega)}{\omega} \Re[f^*(\omega)]\, d\omega. \end{aligned}$$

EXAMPLE 2 (Estimation of generalized moments). Consider estimation of functionals of the form (1.1) where $\varphi(\theta) = \theta^m u(\theta)$ with $u(\theta)$ such that $u(\theta) \in L^2(-\infty, \infty)$ but $\theta^m u(\theta) \notin L^2(-\infty, \infty)$. Note that, since $u(\theta) \in L^2(-\infty, \infty)$ implies that $u(\theta)\mathbb{I}(\theta > \theta_0) \in L^2(-\infty, \infty)$ and $u(\theta) \operatorname{sign}(\theta - \theta_0) \in L^2(-\infty, \infty)$ for any θ_0, we are automatically including all square integrable discontinuous functions $u(\theta)$. In order to derive an inversion formula for this functional, denote $f_m(\theta) = \theta^m f(\theta)$ and observe that $f_m^*(\omega) = i^{-m} \frac{d^m}{d\omega^m}[f^*(\omega)]$. Therefore, if $f_m(\theta) \in L^1(-\infty, \infty)$, one has

$$\Phi_u = \int_{-\infty}^{\infty} \theta^m u(\theta) f(\theta)\, d\theta = \frac{i^{-m}}{2\pi} \int_{-\infty}^{\infty} u^*(-\omega) \frac{d^m f^*(\omega)}{d\omega^m}\, d\omega. \tag{3.4}$$

EXAMPLE 3 (Estimation of the $(2M+1)$-th absolute moment of the deconvolution density). Consider estimation of a functional of the form

$$\Phi_{2M+1} = \int_{-\infty}^{\infty} |\theta|^{2M+1} f(\theta)\, d\theta \tag{3.5}$$

under assumption that $\Phi_{2M+1} < \infty$. Since $|\theta|^{2M+1} = \theta^{2M+1}\operatorname{sign}(\theta)$, using Lemma 1, rewrite Φ_{2M+1} as

$$\begin{aligned}\Phi_{2M+1} &= \frac{2}{\pi}\int_0^{\infty} \frac{1}{\omega} \int_{-\infty}^{\infty} \sin(\omega\theta)\theta^{2M+1} f(\theta)\, d\theta\, d\omega \\ &= \frac{2}{i^{2M+1}\pi} \int_0^{\infty} \frac{1}{\omega} \int_{-\infty}^{\infty} \Im\left[\frac{d^{2M+1}}{d\omega^{2M+1}} \int_{-\infty}^{\infty} e^{i\omega\theta} f(\theta)\, d\theta\right] d\omega.\end{aligned}$$

Therefore,

$$\Phi_{2M+1} = (-1)^{M+1} \frac{2}{\pi} \int_0^{\infty} \frac{1}{\omega} \frac{d^{2M+1}\Re[f^*(\omega)]}{d\omega^{2M+1}}\, d\omega. \tag{3.6}$$

3.2. *Construction of the estimators and evaluation of their risks.* Observe that in all three examples above, the linear functionals Φ can be presented as a combination of two integrals

$$\Phi = \int_0^{\infty} \psi_{m1}^*(\omega) \frac{d^m \Re[f^*(\omega)]}{d\omega^m}\, d\omega + \int_0^{\infty} \psi_{m2}^*(\omega) \frac{d^m \Im[f^*(\omega)]}{d\omega^m}\, d\omega, \tag{3.7}$$

where we assume that f satisfies conditions [that, of course, depend on the particular forms of $\psi_{m1}^*(\omega)$ and $\psi_{m2}^*(\omega)$] which guarantee absolute convergence of the integrals in (3.7). In particular, we assume that both $f^*(\omega)$ and $g^*(\omega)$ are m times differentiable. Note that the regular case corresponds to $m = 0$ and $\psi_{m1}^*(\omega) = \psi_{m2}^*(\omega) = \varphi^*(-\omega)/\pi$.

In order to construct an estimator of the functional Φ in (3.7), we partition the area of integration into $\mathcal{A}_1 = [0; 1]$ and $\mathcal{A}_2 = (1, \infty)$ and rewrite Φ as $\Phi = \Phi_1 + \Phi_2$ where

$$\begin{aligned}\Phi_k = &\int_{\mathcal{A}_k} \psi_{m1}^*(\omega) \Re\left[\frac{d^m f^*(\omega)}{d\omega^m}\right] d\omega \\ &+ \int_{\mathcal{A}_k} \psi_{m2}^*(\omega) \Im\left[\frac{d^m f^*(\omega)}{d\omega^m}\right] d\omega, \qquad k = 1, 2.\end{aligned} \tag{3.8}$$

Since in the majority of situations the error density g is symmetric about zero, we assume that g is an even function, so that its Fourier transform $g^*(\omega)$ is real-valued. Otherwise, $1/g^*(\omega) = G_1(\omega) - iG_2(\omega)$ where $G_1(\omega) = \Re[g^*(\omega)/|g^*(\omega)|^2]$ is an even and $G_2(\omega) = \Im[g^*(\omega)/|g^*(\omega)|^2]$ is an odd function of ω and all subsequent calculations can be rewritten accordingly.

If g is an even function, then using the formula 0.42 of Gradshtein and Ryzhik (1980), Φ_1 in (3.8) can be rewritten as

$$\Phi_1 = \sum_{j=0}^{m} \binom{m}{j} \int_0^1 \frac{d^{(m-j)}}{d\omega^{(m-j)}} \left[\frac{1}{g^*(\omega)} \right] \times \left[\psi_{m1}^*(\omega) u_{j1}(\omega) + \psi_{m2}^*(\omega) u_{j2}(\omega) \right] d\omega, \tag{3.9}$$

where

$$u_{j1}(\omega) = \Re\left[\frac{d^j q^*(\omega)}{d\omega^j} \right], \qquad u_{j2}(\omega) = \Im\left[\frac{d^j q^*(\omega)}{d\omega^j} \right]. \tag{3.10}$$

Denote

$$v_{j1}(\omega) = \int_{-\infty}^{\infty} y^j q(y) \cos(\omega y)\, dy, \qquad v_{j2}(\omega) = \int_{-\infty}^{\infty} y^j q(y) \sin(\omega y)\, dy$$

and construct their respective unbiased estimators as

$$\widehat{v_{j1}}(\omega) = n^{-1} \sum_{l=1}^{n} Y_l^j \cos(\omega Y_l), \qquad \widehat{v_{j2}}(\omega) = n^{-1} \sum_{l=1}^{n} Y_l^j \sin(\omega Y_l). \tag{3.11}$$

By taking derivatives of $q^*(\omega)$ under the integral sign, it is easy to check that the unbiased estimators of $u_{j1}(\omega)$ and $u_{j2}(\omega)$ are, respectively, given by

$$\widehat{u_{j1}}(\omega) = \begin{cases} (-1)^{j/2} \widehat{v_{j1}}(\omega), & \text{if } j \text{ is even}, \\ (-1)^{(j+1)/2} \widehat{v_{j2}}(\omega), & \text{if } j \text{ is odd}, \end{cases} \tag{3.12}$$

$$\widehat{u_{j2}}(\omega) = \begin{cases} (-1)^{j/2} \widehat{v_{j2}}(\omega), & \text{if } j \text{ is even}, \\ (-1)^{(j-1)/2} \widehat{v_{j1}}(\omega), & \text{if } j \text{ is odd}. \end{cases} \tag{3.13}$$

Combination of formulae (3.9)–(3.13) imply that Φ_1 can be estimated by

$$\widehat{\Phi}_1 = \sum_{j=0}^{m} \binom{m}{j} \int_0^1 \frac{d^{(m-j)}}{d\omega^{(m-j)}} \left[\frac{1}{g^*(\omega)} \right] \times \left[\psi_{m1}^*(\omega) \widehat{u_{j1}}(\omega) + \psi_{m2}^*(\omega) \widehat{u_{j2}}(\omega) \right] d\omega. \tag{3.14}$$

Denote

$$\sigma_{j1}^2(\omega) = n \operatorname{Var}\left[\widehat{u_{j1}}(\omega) \right], \qquad \sigma_{j2}^2(\omega) = n \operatorname{Var}\left[\widehat{u_{j2}}(\omega) \right] \tag{3.15}$$

and introduce the following assumption:

A3. There exists an absolute constant C_σ such that for any $j = 0, \dots, m$, and $k = 1, 2$, one has

$$\int_0^1 \left[\frac{d^{m-j}}{d\omega^{m-j}} \left(\frac{1}{g^*(\omega)} \right) \right]^2 |\psi^*_{mk}(\omega)|^2 \sigma^2_{jk}(\omega)\, d\omega \le C_\sigma. \tag{3.16}$$

If functions f and g are such that $\mu_{2m} < \infty$, $\nu_{2m} < \infty$, where μ_k and ν_k are defined in (3.2), then under Assumption A3, the values of $\widehat{u_{j1}}(\omega)$ and $\widehat{u_{j2}}(\omega)$ are uniformly bounded, and hence all integrals in (3.14) are absolutely convergent.

In order to estimate Φ_2, using integration by parts, partition Φ_2 in (3.8) as $\Phi_2 = F_m(1) + \Phi_{20}$, where

$$F_m(\omega) = \sum_{k=1}^m (-1)^k \Bigg[\frac{d^{k-1}\psi^*_{m1}(\omega)}{d\omega^{k-1}} \frac{d^{m-k}[\Re(f^*(\omega))]}{d\omega^{m-k}} + \frac{d^{k-1}\psi^*_{m2}(\omega)}{d\omega^{k-1}} \frac{d^{m-k}[\Im(f^*(\omega))]}{d\omega^{m-k}} \Bigg],$$

$$\Phi_{20} = (-1)^m \int_1^\infty \left(\frac{d^m\psi^*_{m1}(\omega)}{d\omega^m} \Re[f^*(\omega)] + \frac{d^m\psi^*_{m2}(\omega)}{d\omega^m} \Im[f^*(\omega)] \right) d\omega.$$

Again, taking into account that $f^*(\omega) = q^*(\omega)/g^*(\omega)$ and using the formula 0.42 of Gradshtein and Ryzhik (1980), rewrite $F_m(1)$ and Φ_{20} as

$$F_m(1) = \sum_{k=1}^m \sum_{j=0}^{m-k} (-1)^k \binom{m-k}{j} [A_{m,j,k,1}(1) u_{j1}(1) + A_{m,j,k,2}(1) u_{j2}(1)],$$

$$\Phi_{20} = (-1)^m \int_1^\infty \left[\frac{d^m\psi^*_{m1}(\omega)}{d\omega^m} u_{01}(\omega) + \frac{d^m\psi^*_{m2}(\omega)}{d\omega^m} u_{02}(\omega) \right] \frac{1}{g^*(\omega)}\, d\omega,$$

where $u_{j1}(\omega)$ and $u_{j2}(\omega)$ are defined in (3.10) and for $l = 1, 2$

$$A_{m,j,k,l}(\omega) = \frac{d^{k-1}\psi^*_{lm}(\omega)}{d\omega^{k-1}} \mathcal{A}_{m-k-j}(\omega) \quad \text{with } \mathcal{A}_t(\omega) = \frac{d^t}{d\omega^t}\left(\frac{1}{g^*(\omega)}\right). \tag{3.17}$$

Therefore, we can estimate Φ_2 by $\widehat{\Phi}_{2h} = \widehat{F_m}(1) + \widehat{\Phi}_{20h}$ where

$$\widehat{F_m}(1) = \sum_{k=1}^m \sum_{j=0}^{m-k} (-1)^k \binom{m-k}{j} \times [A_{m,j,k,1}(1)\widehat{u_{j1}}(1) + A_{m,j,k,2}(1)\widehat{u_{j2}}(1)], \tag{3.18}$$

$$\widehat{\Phi}_{20h} = (-1)^m \int_1^{1/h} \left[\frac{d^m\psi^*_{m1}(\omega)}{d\omega^m} \widehat{u}_{01}(\omega) + \frac{d^m\psi^*_{m2}(\omega)}{d\omega^m} \widehat{u}_{02}(\omega) \right] \times \frac{1}{g^*(\omega)}\, d\omega, \tag{3.19}$$

and $\widehat{u_{j1}}(\omega)$ and $\widehat{u_{j2}}(\omega)$ are defined by (3.12) and (3.13). Finally, we estimate Φ in (3.7) by $\widehat{\Phi}_h = \widehat{\Phi}_1 + \widehat{F_m}(1) + \widehat{\Phi}_{20h}$ where $\widehat{\Phi}_1$, $\widehat{F_m}(1)$ and $\widehat{\Phi}_{20h}$ are evaluated according to (3.14) and (3.18), respectively.

In order to construct an upper bound for the risk of the estimator $\widehat{\Phi}_h$, we consider a class of pdfs

$$\Xi_s(B) = \Bigl\{ f : \sup_{\omega} \bigl[|f^*(\omega)| \bigl(|\omega|^s + 1 \bigr) \bigr] \le B_2 \Bigr\} \tag{3.20}$$

Then the risk of the estimator $\widehat{\Phi}_h$ is given by the following statement.

THEOREM 4. *Assume that f and g are such that $\mu_{2m} < \infty$, $\nu_{2m} < \infty$, where μ_k and ν_k are defined in* (3.2), *and that Assumption* A3 *holds. Let also function g^* be real-valued, satisfy Assumption* A1, *be m times differentiable and such that, for some $C_g > 0$ and $j = 0, \dots, m$,*

$$\left| \frac{1}{g^*(\omega)} \frac{d^j g^*(\omega)}{d\omega^j} \right| \le C_g \bigl(|\omega| + 1 \bigr)^{j\tau}, \qquad \tau \ge 0, \ \text{where } \tau = 0 \text{ if } \gamma = 0. \tag{3.21}$$

Let $\psi^*_{m1}(\omega)$ and $\psi^*_{m2}(\omega)$ be such that for some positive C_ψ and nonnegative d, a_m and b, for any $j = 0, \dots, m$, one has for $|\omega| \ge 1$:

$$\left| \frac{d^j \psi^*_{mk}(\omega)}{d\omega^j} \right| \le C_\psi \bigl(\omega^2 + 1 \bigr)^{-a_m(j+1)/2} \exp\bigl(-d|\omega|^b \bigr), \qquad k = 1, 2. \tag{3.22}$$

Let $\widehat{\Phi}_h = \widehat{\Phi}_1 + \widehat{F_m}(1) + \widehat{\Phi}_{20h}$ where $\widehat{\Phi}_1$, $\widehat{F_m}(1)$ and $\widehat{\Phi}_{20h}$ are defined in (3.14) *and* (3.18), *respectively. Then*

$$\begin{aligned} \mathbb{E}(\widehat{\Phi}_h - \Phi)^2 \le C \biggl[& h^{2A} \exp\Bigl(-\frac{2d}{h^b} \Bigr) \\ & + \int_1^{h^{-1}} \frac{(\omega^2 + 1)^{A_0}}{n} \exp\bigl(-2d\omega^b + 2\gamma\omega^\beta \bigr)\, d\omega \biggr], \end{aligned} \tag{3.23}$$

where $A_0 = \alpha - (m+1)a_m$, $A = (m+1)a_m + s + b - 1$ if $f \in \Xi_s(B)$ and $A = (m+1)a_m + s + (b-1)/2$ if $f \in \Omega_s(B)$. Here, $\Omega_s(B)$ and $\Xi_s(B)$ are defined by (2.2) *and* (3.20), *respectively.*

Using Lemma 2 in Section A.1, one can obtain convergence rates and optimal bandwidth values $\tilde{h}$ for each combination of parameters $b, d, \beta, \gamma, s, m$ and a_m. Moreover, application of an equivalent of Theorem 3 allows one to obtain an adaptive estimator of Φ. Note, however, that one cannot automatically derive the lower bounds for the risk from Theorem 1. Indeed, in addition to $f \in \Omega_s(B)$ or $f \in \Xi_s(B)$, Assumption A3 imposes additional restrictions on f^* that depend, in a nontrivial way, on the shapes of functions ψ^*_{mk}, $k = 1, 2$, thus, modifying the class of functions f. For this reason, one has to derive lower bounds for the minimax risk on a case-by-case basis. As an example of how this can be done, in the next section, we consider estimation of Φ_{2M+1} given by (3.5).

3.3. *Estimation of the* $(2M+1)$*-th absolute moment of the deconvolution density.* According to (3.6), Φ_{2M+1} can be written in the form (3.7) with $\psi^*_{2M+1,1}(\omega) = 2(-1)^{M+1}(\pi\omega)^{-1}$ and $\psi^*_{2M+1,2}(\omega) = 0$. Similar to (3.8), rewrite Φ_{2M+1} as $\Phi_{2M+1} = \Phi_{2M+1,1} + \Phi_{2M+1,2}$ where $\Phi_{2M+1,1}$ and $\Phi_{2M+1,2}$ are the portions of Φ_{2M+1} evaluated over intervals [0, 1] and $(1, \infty)$. Here,

$$\Phi_{2M+1,1} = (-1)^{M+1}\frac{2}{\pi}\sum_{j=0}^{2M+1}\binom{2M+1}{j}\int_0^1 \frac{1}{\omega}\mathcal{A}_{2M+1-j}(\omega)u_{j1}(\omega)\,d\omega,$$

where $u_{j1}(\omega)$ and $\mathcal{A}_t(\omega)$ are defined in (3.10) and (3.17), respectively. Taking into account relations (3.12) and partitioning the sum above into the portions with the even and the odd indices, we obtain the following estimator of $\Phi_{2M+1,1}$:

$$\begin{aligned}\widehat{\Phi}_{2M+1,1} = \frac{2}{\pi}\sum_{k=0}^{M}(-1)^{M+k}\Bigg\{&\binom{2M+1}{2k+1}\int_0^1 \frac{\widehat{v}_{2k+1,2}(\omega)}{\omega}\mathcal{A}_{2(M-k)}(\omega)\,d\omega \\ &-\binom{2M+1}{2k}\int_0^1 \frac{\widehat{v}_{2k,1}(\omega)}{\omega}\mathcal{A}_{2(M-k)+1}(\omega)\,d\omega\Bigg\}.\end{aligned} \tag{3.24}$$

Note that for $\sigma^2_{j1}(\omega)$ defined in (3.15), one has $\sigma^2_{2j,1}(\omega) \asymp \mathrm{Var}[Y^{2j}\cos(\omega Y)] \le \int_{-\infty}^{\infty} y^{4j}q(y)\,dy$ and $\sigma^2_{2j+1,2}(\omega) \asymp \mathrm{Var}[Y^{2j+1}\sin(\omega Y)]$, so that $\sigma^2_{2j+1,2}(\omega) \le \min[\int_{-\infty}^{\infty} y^{4j+2}q(y)\,dy, \omega^2\int_{-\infty}^{\infty} y^{4j+4}q(y)\,dy]$. Hence, condition (3.16) is guaranteed by $\mu_{4M+4} < \infty$ and $\nu_{4M+4} < \infty$ where μ_k and ν_k are defined in (3.2). Consider now $\widehat{\Phi}_{2M+1,2}$. Taking into account that, for any $j = 0, 1, 2, \ldots$ $\frac{d^j\psi^*_{2M+1,1}(\omega)}{d\omega^j} = \frac{2}{\pi}\frac{(-1)^{M+1+j}j!}{\omega^{j+1}}$, apply (3.18) for this particular case to obtain

$$\begin{aligned}\widehat{\Phi}_{2M+1,2,h} = -\frac{2}{\pi}\Bigg\{&\int_1^{1/h}\widehat{v}_{01}(\omega)\omega^{-(2M+2)}\big[g^*(\omega)\big]^{-1}\,d\omega \\ &-(-1)^M\sum_{k=1}^{2M+1}\sum_{j=0}^{2M+1-k}\binom{2M+1-k}{j} \\ &\times(k-1)!\mathcal{A}_{2M+1-k-j}(1)\widehat{u_{j1}}(1)\Bigg\},\end{aligned} \tag{3.25}$$

where $\mathcal{A}_t$ is defined in (3.10). Finally, Φ_{2M+1} can be estimated as $\widehat{\Phi}_{2M+1,h} = \widehat{\Phi}_{2M+1,1} + \widehat{\Phi}_{2M+1,2,h}$ where $\widehat{\Phi}_{2M+1,1}$ and $\widehat{\Phi}_{2M+1,2,h}$ are given by (3.24) and (3.25), respectively. Note that condition (3.22) holds with $m = 2M+1$, $a_m = 1$ and $b = d = 0$.

In particular, if $M = 0$, formulae (3.24) and (3.25) yield estimator for the first absolute moment of f of the form $\widehat{\Phi}_{1,h} = \widehat{\Phi}_{1,1} + \widehat{\Phi}_{1,2,h}$ where

$$\widehat{\Phi}_{1,1} = \frac{2}{\pi}\int_0^1\left(\frac{\widehat{v}_{01}(\omega)(g^*)'(\omega)}{\omega(g^*(\omega))^2} + \frac{\widehat{v}_{12}(\omega)}{\omega g^*(\omega)}\right)d\omega, \tag{3.26}$$

$$\widehat{\Phi}_{1,2,h} = \frac{2}{\pi}\frac{\widehat{v}_{01}(1)}{g^*(1)} - \frac{2}{\pi}\int_1^{1/h}\frac{1}{\omega^2}\frac{\widehat{v}_{01}(\omega)}{g^*(\omega)}\,d\omega. \tag{3.27}$$

In order to derive upper and lower bounds for the minimax risk of the estimator $\widehat{\Phi}_{2M+1,h}$, we introduce the following sets of functions:

$$\begin{aligned}\Xi_s(B_1, B_2) = \biggl\{f : \int_{-\infty}^{\infty}\theta^{4M+4} f(\theta)\,d\theta \le B_1,\\ \sup_\omega\bigl[\bigl|f^*(\omega)\bigr|\bigl(|\omega|^s + 1\bigr)\bigr] \le B_2\biggr\}\end{aligned} \tag{3.28}$$

The theorem below provides upper and lower bounds for the minimax risk of an estimator of Φ_{2M+1} in (3.5) while Corollary 3 produces similar results for the discrete version of the functional $\Phi_{2M+1,n} = n^{-1}\sum_{i=1}^n |\theta_i|^{2M+1}$.

THEOREM 5. *Let $\mu_{4M+4} \le B_g < \infty$ for some positive constant B_g, where μ_k is defined in* (3.2). *Let g^* satisfy condition* (3.21) *and also be such that*

$$\sup_{|\omega|\le 1}\left|\frac{1}{\omega}\frac{d^{2k+1}}{d\omega^{2k+1}}\left(\frac{1}{g^*(\omega)}\right)\right| \le C_{gM}, \qquad 0 \le k \le M. \tag{3.29}$$

If inequality (2.3) *in Assumptions* A1 *holds, then for $h = \tilde{h}_n$, we obtain*

$$R_n^\sharp \le \begin{cases} Cn^{-1}, \tilde{h}_n = 0, & \text{if } 2\alpha - 4M < 3, \gamma = 0,\\ Cn^{-1}\log n, \tilde{h}_n = n^{-\frac{1}{4M+2}}, & \text{if } 2\alpha - 4M = 3, \gamma = 0,\\ Cn^{-\frac{4M+2s+2}{2s+2\alpha-1}}, \tilde{h}_n = n^{-\frac{1}{2s+2\alpha-1}}, & \text{if } 2\alpha - 4M > 3, \gamma = 0,\\ C(\log n)^{-\frac{4M+2s+2}{\beta}}, \widehat{h}_n = \tilde{h}_n^*, & \text{if } \beta > 0, \gamma > 0,\end{cases} \tag{3.30}$$

where $\tilde{h}_n^* = [\log n/(3\gamma)]^{-\frac{1}{\beta}}$. If inequality (2.4) *in Assumption* A1 *holds, then*

$$R_n^\flat \ge \begin{cases} Cn^{-1}, & \text{if } 2\alpha - 4M \le 3, \beta = \gamma = 0,\\ Cn^{-\frac{4M+2s+2}{2s+2\alpha-1}}, & \text{if } 2\alpha - 4M > 3, \beta = \gamma = 0,\\ C(\log n)^{-\frac{4M+2s+2}{\beta}}, & \text{if } \beta > 0, \gamma > 0.\end{cases} \tag{3.31}$$

Here, $R_n^\sharp = R_n(\widehat{\Phi}_{2M+1,\tilde{h}_n}, \Xi_s(B_1, B_2)) = \sup_{f\in\Xi_s(B_1,B_2)}\mathbb{E}(\widehat{\Phi}_{2M+1,\tilde{h}_n} - \Phi_{2M+1})^2$ and $R_n^\flat = R_n(\Xi_s(B_1, B_2)) = \inf_{\widetilde{\Phi}}\sup_{f\in\Omega_s(B_1,B_2)}\mathbb{E}(\widetilde{\Phi} - \Phi_{2M+1})^2$.

Note that the values of $\tilde{h}_n$ are independent of unknown parameters s, B_1 and B_2 if $2\alpha - 4M \le 3$ and $\beta = \gamma = 0$ or if $\beta > 0$ and $\gamma > 0$. If $2\alpha - 4M > 3$ and $\beta = \gamma = 0$, one can find the value of $\widehat{h}_n$ using the Lepskii method similar to the regular case with the price of an extra $\log n$ in the convergence rates.

COROLLARY 3. *Let θ_i, $i = 1, \dots, n$, in (1.2) be i.i.d. with pdf f. If $\Phi_{2M+1,n}$ is defined by formula (1.3) with $\varphi(\theta) = |\theta|^{2M+1}$, then under assumptions of Theorem 5, for sufficiently large n, one has*

$$\sup_{f \in \Xi_s(B)} \mathbb{E}(\widehat{\Phi}_{2M+1,\tilde{h}_n} - \Phi_{2M+1,n})^2 \asymp R_n\big(\widehat{\Phi}_{2M+1,\tilde{h}_n}, \Xi_s(B_1, B_2)\big),$$

$$\inf_{\widetilde{\Phi}_n} \sup_{f \in \Xi_s(B)} \mathbb{E}(\widetilde{\Phi}_n - \Phi_{2M+1,n})^2 \asymp R_n\big(\Xi_s(B_1, B_2)\big),$$

where $R_n(\widehat{\Phi}_{2M+1,\tilde{h}_n}, \Xi_s(B_1, B_2))$ and $R_n(\Xi_s(B_1, B_2))$ are given by (3.30) and (3.31), respectively.

REMARK 2 (The choice of the class of functions). Observe that we derived the lower and the upper bounds for the risk not for the subset of the Sobolev ball $\Omega_s(B_2)$ but rather for the subset $\Xi_s(B_1, B_2)$ of $\Xi_s(B_2)$. This is motivated by our intention to compare our estimator for the first absolute moment with the respective estimator of Cai and Low (2011). One can easily obtain upper and lower bounds for the minimax risk over the set $\Omega_s(B_1, B_2))$ in a very similar manner.

REMARK 3 [Relation to Cai and Low (2011)]. Cai and Low (2011) studied estimation of Φ_n of the form (1.3) with $\varphi(\theta) = |\theta|$ based on data generated by model (1.2) where the errors ξ_i are i.i.d. $\mathcal{N}(0, \sigma^2)$ and there are no probabilistic assumptions on vector $\boldsymbol{\theta}$. They showed that

$$\inf_{\widetilde{\Phi}} \sup_{\Theta_n(M_0)} \mathbb{E}(\widetilde{\Phi}_n - \Phi)^2 \asymp M_0^2 \left(\frac{\log\log n}{\log n}\right)^2, \qquad \inf_{\widetilde{\Phi}} \sup_{\Theta_n(\infty)} \mathbb{E}(\widetilde{\Phi}_n - \Phi)^2 \asymp \frac{1}{\log n},$$

where $\Theta_n(M_0) = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\|_\infty \le M_0\}$. By employing a rather complex procedure based on Chebyshev and Hermite polynomials, they constructed adaptive estimators that attain these convergence rates. With the assumption that θ_i are generated independently from pdf f, the problem reduces to estimation of Φ_1 in (3.5), the first absolute moment of the mixing density. Using formulae (3.26) and (3.27), one can construct an estimator $\widehat{\Phi}_1$ of Φ_1.

Note that, since in the case of Gaussian errors, one has $\alpha = 0$, $\beta = 2$ and $\gamma = \sigma^2/2$, the estimators (3.26) and (3.27) are adaptive if $h = \widehat{h}_n = [\log n/\sigma^2]^{-1/2}$, and Corollary 3 implies that

$$R_n\big(\widehat{\Phi}_{\widehat{h}_n}, \Phi_n, \Xi_s(B_1, B_2)\big) \asymp R_n\big(\Phi_n, \Xi_s(B_1, B_2)\big) \asymp (\log n)^{-(s+1)},$$

where $\Xi_s(B_1, B_2)$ is defined in (3.28). Since f is a pdf, it is absolutely integrable, so that $s \ge 0$. In our setting, the convergence rates of Cai and Low (2011) correspond to "the worst case scenario" where $s = 0$ and f is a combination of delta functions. Since Cai and Low (2011) do not impose any probabilistic assumptions on θ_i, their estimator addresses this "worst-case scenario" but is unable to adapt to a more favorable situation where $|f^*(\omega)| \to 0$ as $|\omega| \to \infty$. In addition, our estimator can be used for any type of error density g. On the other hand, our estimator requires θ_i, $i = 1, \dots, n$, to be i.i.d. random variables which is not necessary in the setup of Cai and Low (2011).

4. The sparse case.

4.1. *Estimation procedure and the upper bounds for the risk.* The objective of this section is to estimate the functional $\Phi_\mu = \int_{-\infty}^{\infty} \varphi(x) f_0(x)\, dx$ defined by (1.5). Here, $f_0(\theta)$ is the pdf of the nonzero entries of $\boldsymbol{\theta}$ and

$$f(x) = \mu_n f_0(x) + (1-\mu_n)\delta(x), \tag{4.1}$$

where $\mu_n = n^{\nu-1}$, $0 < \nu < 1$, is known. Due to (4.1), one has $\Phi = \mu_n \Phi_\mu + (1 - \mu_n)\varphi(0)$, so that the value of Φ_μ can be recovered as

$$\Phi_\mu = {\mu_n}^{-1}\Phi - {\mu_n}^{-1}(1-\mu_n)\varphi(0). \tag{4.2}$$

Therefore, we estimate Φ_μ by

$$\begin{aligned} \widehat{\Phi}_{\mu,h} &= \frac{\widehat{\Phi}_h}{\mu_n} - \frac{1-\mu_n}{\mu_n}[\varphi(0) - \delta_h] \\ &\qquad \text{with } \delta_h = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi^*(-\omega)\mathbb{I}\big(|\omega| > h^{-1}\big)\, d\omega, \end{aligned} \tag{4.3}$$

where $\widehat{\Phi}_h$ is defined in (2.10) and the correction term δ_h is a completely known nonrandom quantity. In order to justify the estimator (4.3), we derive expressions for its variance and bias. Since the second term in (4.3) is nonrandom, $\operatorname{Var}(\widehat{\Phi}_{\mu,h}) = {\mu_n}^{-2}\operatorname{Var}(\widehat{\Phi}_h)$ where $\operatorname{Var}(\widehat{\Phi}_h)$ is bounded by

$$\operatorname{Var}(\widehat{\Phi}_{\mu,h}) \le \frac{\|g\|_\infty}{2\pi n {\mu_n}^2}\int_{-\infty}^{\infty} \frac{|\varphi^*(\omega)|^2}{|g^*(\omega)|^2}\mathbb{I}\big(|\omega| \le h^{-1}\big)\, d\omega.$$

Since $f^*(\omega) = \mu_n f_0^*(\omega) + (1-\mu_n)$, the bias term of $\widehat{\Phi}_{\mu,h}$ is of the form

$$\mathbb{E}\widehat{\Phi}_{\mu,h} - \Phi_\mu = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi^*(-\omega) f_0^*(\omega)\mathbb{I}\big(|\omega| > h^{-1}\big)\, d\omega + \frac{1-\mu_n}{2\pi\mu_n}\Delta(h),$$

where $\Delta(h) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi^*(-\omega)\mathbb{I}(|\omega| \le h^{-1})\, d\omega - \varphi(0) + \delta_h = 0$. Therefore, for any $f_0 \in \Omega_s(B)$, where $\Omega_s(B)$ is defined in (2.2), one derives

$$(\mathbb{E}\widehat{\Phi}_{\mu,h} - \Phi_\mu)^2 \le \frac{B^2}{4\pi^2}\int_{-\infty}^{\infty} \frac{|\varphi^*(\omega)|^2}{(\omega^2+1)^s}\mathbb{I}\big(|\omega| > h^{-1}\big)\, d\omega.$$

Hence,

$$\mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_\mu)^2 \le \frac{\|g\|_\infty}{2\pi n {\mu_n}^2}\int_{-h^{-1}}^{h^{-1}} \frac{|\varphi^*(\omega)|^2}{|g^*(\omega)|^2}\, d\omega + \frac{B^2}{2\pi^2}\int_{h^{-1}}^{\infty} \frac{|\varphi^*(\omega)|^2}{(\omega^2+1)^s}\, d\omega. \tag{4.4}$$

Let $n_\mu = n{\mu_n}^2 = n^{2\nu-1}$ be the new, "effective" sample size. Then, comparing (4.4) with (2.12), one immediately observes that the upper bounds for the risk of the estimator $\widehat{\Phi}_{\mu,h}$ of Φ_μ would coincide with the upper bounds for the risk of the estimator $\widehat{\Phi}_h$ of Φ in the nonsparse case if the sample size n were replaced by the effective sample size n_μ. Denote

$$R_{n,\mu_n}\big(\widehat{\Phi}_{\mu,h}, \Omega_s(B)\big) = \sup_{f_0 \in \Omega_s(B)} \mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_\mu)^2, \tag{4.5}$$

where $\Omega_s(B)$ is defined in (2.2). If $\nu > 1/2$, then $n_\mu = n^{2\nu-1} \to \infty$, so that combination of Theorem 2 and formula (4.4) immediately yields the upper bounds for the risk.

THEOREM 6. *Let g be bounded above and observations be given by model* (1.2) *where f is of the form* (4.1) *with the known $\mu_n = n^{\nu-1}$, $\nu > 1/2$. Then, under Assumptions* A1 *and* A2 [*inequalities* (2.3) *and* (2.6) *only*], *the expressions for the upper bounds for the risks $R^{up}_{n,\mu_n} \equiv R_{n,\mu_n}(\widehat{\Phi}_{\mu,h_n}, \Omega_s(B))$ and $\widehat{R}^{up}_{n,\mu_n} \equiv R_{n,\mu_n}(\widehat{\Phi}_{\mu,\hat{h}_n}, \Omega_s(B))$ of, respectively, the estimator $\widehat{\Phi}_{\mu,h_n}$ and of its adaptive version are provided in Table* 2 *where $U_3 = (2s + 2a + b - 1) - d(2\alpha + 2s)/\gamma$, $\Delta U_3 = 2s(\gamma - d)/\gamma$, $U_4 = 2s + 2a + b - 1$, $\Delta U_4 = 2(s - s_0)_+$ where s_0 is defined in Theorem* 3, *$U_5 = 2a + 2s(1 - d/\gamma) - 2d\alpha/\gamma + 4\beta(\varsigma_0 + 1) - (U_{\tau,\varsigma_0} + 1)d/\gamma - 1$, $U_6 = 4b(\varsigma_0 + 1) + 2a + 2s - 1$ and $U_{\tau,\varsigma_0} = \min(\beta(4\varsigma_0 + 3) - 2\tau\varsigma_0 - 1, \beta(2\varsigma_0 + 3) + 2\varsigma_0 - 1)$.*

The corresponding values of h are $h = h_n$ for the nonadaptive estimator and $h = \widehat{h}_n$ for the adaptive estimator, where $h_n = \hat{h}_n = 0$ for cases 1, 2 *and* 3; *$h_n = \hat{h}_n = n^{1-2\nu}$ for case* 5; *$h_n = n^{\frac{1-2\nu}{2s+2\alpha}}$ and $\hat{h}_n = \widehat{j}$ where $\widehat{j}$ is defined in* (2.14) *with C_Φ given by* (2.15) *and n replaced by $n^{2\nu-1}$ for case* 7; *$h_n = H_n(t)$ and $\widehat{h}_n = H_n(\hat{t})$ where $H_n(x) = \{(2\nu - 1)[\log n - (t/\beta)\log\log n]/(2\gamma)\}^{-\frac{1}{\beta}}$, where $t = \hat{t} = 0$ for cases* 4 *and* 6; *$t = 2s + 2\alpha$ and $\hat{t} = 2\alpha$ for case* 8; *$t = 2s + 2\alpha + b - \beta$ and $\hat{t} = 2s_0 + 2\alpha + b - \beta$ for case* 9 *and $h_n = \hat{h}_n = [(2\nu - 1)\log n/(3\gamma)]^{-\frac{1}{\beta}}$ for case* 10.

4.2. *The lower bounds for the risk.* The upper bounds for the risk in formula (4.4) suggest that, for $0 \le \nu \le 1/2$, one has $n_\mu = n{\mu_n}^2 = n^{2\nu-1} \le 1$ and construction of a consistent estimator is impossible for any functional of the form (1.5). The next proposition shows that this indeed is true in a wide variety of situations. In particular, under mild assumptions, the risk of no estimator can converge to zero at a faster rate than n_μ^{-1}.

THEOREM 7. *Let $f(x)$ be given by* (4.1) *and Φ_μ be defined by* (1.5). *If, for some $C_I > 0$, there exist two pdf's, $f_1(\theta)$ and $f_2(\theta)$, such that*

$$I_k = \int_{-\infty}^{\infty} g^{-1}(x)\bigg[\int_{-\infty}^{\infty} g(x-\theta) f_k(\theta)\, d\theta\bigg]^2 dx \le C_I < \infty, \qquad k = 1, 2, \tag{4.6}$$

TABLE 2
Asymptotic expressions for the minimax lower bounds R^{low}_{n,μ_n} and the risks of adaptive estimators $\widehat{R}^{up}_{n,\mu_n}$

	Cases	Bounds for the risk
1	$b > \beta$	$R^{low}_{n,\mu_n} \asymp R^{up}_{n,\mu_n} \asymp \widehat{R}^{up}_{n,\mu_n} \asymp n^{-(2\nu-1)}$
2	$b = \beta, d > \gamma > 0$	$R^{low}_{n,\mu_n} \asymp R^{up}_{n,\mu_n} \asymp \widehat{R}^{up}_{n,\mu_n} \asymp n^{-(2\nu-1)}$
3	$b = \beta, d = \gamma, a > \alpha + 1/2$	$R^{low}_{n,\mu_n} \asymp R^{up}_{n,\mu_n} \asymp \widehat{R}^{up}_{n,\mu_n} \asymp n^{-(2\nu-1)}$
4	$b = \beta > 0, d = \gamma > 0,$ $a = \alpha + 1/2$	$R^{low}_{n,\mu_n} \asymp n^{-(2\nu-1)}$, $R^{up}_{n,\mu_n} \asymp \widehat{R}^{up}_{n,\mu_n} \asymp R^{low}_{n,\mu_n} \cdot \log\log n$
5	$b = \beta = 0, d = \gamma = 0,$ $a = \alpha + 1/2$	$R^{low}_{n,\mu_n} \asymp n^{-(2\nu-1)}$ $R^{up}_{n,\mu_n} \asymp \widehat{R}^{up}_{n,\mu_n} \asymp R^{low}_{n,\mu_n} \cdot \log n$
6	$b = \beta > 0, d = \gamma > 0,$ $a < \alpha + 1/2$	$R^{low}_{n,\mu_n} \asymp n^{-(2\nu-1)}$ $R^{up}_{n,\mu_n} \asymp \widehat{R}^{up}_{n,\mu_n} \asymp R^{low}_{n,\mu_n} \cdot (\log n)^{\frac{2\alpha-2a+1}{\beta}}$
7	$b = \beta = 0, d = \gamma = 0,$ $a < \alpha + 1/2$	$R^{low}_{n,\mu_n} \asymp R^{up}_{n,\mu_n} \asymp n^{-\frac{(2\nu-1)(2s+2a-1)}{2s+2\alpha}}$, $\widehat{R}^{up}_{n,\mu_n} \asymp R^{low}_{n,\mu_n} \cdot \log n$
8	$b = \beta > 0, \gamma > d > 0$	$R^{low}_{n,\mu_n} \asymp (\log n)^{-\frac{U_5}{\beta}} n^{-\frac{d(2\nu-1)}{\gamma}}$ $R^{up}_{n,\mu_n} \asymp R^{low}_{n,\mu_n} \cdot (\log n)^{\frac{U_5-U_3}{\beta}}$ $\widehat{R}^{up}_{n,\mu_n} \asymp R^{up}_{n,\mu_n} \cdot (\log n)^{\frac{\Delta U_3}{\beta}}$
9	$\beta > b > 0, d > 0, \gamma > 0$	$R^{low}_{n,\mu_n} \asymp (\log n)^{-\frac{U_6}{\beta}} \exp(-2d[\frac{(2\nu-1)\log n}{2\gamma}]^{b/\beta})$, $R^{up}_{n,\mu_n} \asymp R^{low}_{n,\mu_n} \cdot (\log n)^{\frac{U_6-U_4}{\beta}}$, $\widehat{R}^{up}_{n,\mu_n} \asymp R^{up}_{n,\mu_n} \cdot (\log n)^{\frac{\Delta U_4}{\beta}}$
10	$b = d = 0, \beta > 0, \gamma > 0$	$R^{low}_{n,\mu_n} \asymp R^{up}_{n,\mu_n} \asymp \widehat{R}^{up}_{n,\mu_n} \asymp (\log n)^{-\frac{2s+2a-1}{\beta}}$

and

$$\Delta_{12} = \int_{-\infty}^{\infty} \varphi(\theta)\big[f_1(\theta) - f_2(\theta)\big]\, d\theta \neq 0, \tag{4.7}$$

then for some positive constants C_{01}, C_{02} and p_0 independent of n, and any estimator $\widetilde{\Phi}_\mu$ of Φ_μ based on observations $Y_1, \dots, Y_n$ one has

$$\inf_{\widetilde{\Phi}_\mu} \sup_{f_1, f_2} \mathbb{P}\big\{(\widetilde{\Phi}_\mu - \Phi_\mu)^2 \geq C_{01} \min\big(1, n_\mu^{-1}\big)\big\} \geq p_0, \tag{4.8}$$

where $n_\mu = n{\mu_n}^2$. In particular, $g(x) = \mathcal{N}(x|0, \sigma^2)$ is a Gaussian pdf and $\int_{-\infty}^{\infty} |\varphi(x)| g(x)\, dx < \infty$, then the lower bound (4.8) holds.

Theorem 7 implies that when sparsity level is high, that is, $\nu > 1/2$, consistent estimation of Φ_μ is impossible for all pdfs g satisfying condition (4.6). Moreover, Theorem 7 does not require function φ to be integrable or square integrable, so one can apply this theorem easily to a wide variety of functionals. The quantity $n_\mu^{-1} = (n\mu_n{}^2)^{-1}$ acts as the parametric convergence rate that cannot be improved.

One would like to derive the lower bounds for the risk for any combination of function $\varphi(\theta)$ and $g(\theta)$. Unfortunately, in some of the cases, we need an additional assumption that $g(\theta)$ has polynomial descent as $|\theta| \to \infty$.

A4. Let g be bounded above and such that $|g(\theta)| \ge C_{g1}(\theta^2+1)^{-\varsigma}$. Let function g^* be ς_0 times continuously differentiable, where ς_0 is the closest integer no less than ς, and satisfy the following condition:

$$\frac{|d^l g^*(\omega)|}{d\omega^l} \le C_{g2}|g^*(\omega)|\bigl(1+|\omega|\bigr)^{\tau l}, \qquad l = 1, 2, \dots, \varsigma_0, \text{ where } \tau = 0 \text{ if } \gamma = 0. \tag{4.9}$$

Let there exist $\omega_0 \in (0, \infty)$ such that function $\rho(\omega) = \arg(\varphi^*(\omega))$ is ς_0 times continuously differentiable for $|\omega| \ge \omega_0$, with $|\rho^{(j)}(\omega)| \le \rho < \infty$, $j = 0, 1, 2, \dots, \varsigma_0$.

Denote

$$R^{low}_{n,\mu_n} \equiv R_{n,\mu_n}\bigl(\Omega_s(B)\bigr) = \inf_{\widetilde{\Phi}_\mu} \sup_{f_0 \in \Omega_s(B)} \mathbb{E}(\widetilde{\Phi}_\mu - \Phi_\mu)^2, \tag{4.10}$$

where $\widetilde{\Phi}_\mu$ is any estimator of Φ_μ based on observations $Y_1, \dots, Y_n$ and $\Omega_s(B)$ is defined in (2.2). Then the following theorem is true.

THEOREM 8. *Let $f(x)$ be given by (4.1) where $\mu_n = n^{\nu-1}$ with $\nu > 1/2$ and Φ be defined by (1.5). Let Assumptions* A1 *and* A2 [*inequalities* (2.4) *and* (2.5) *only*] *hold. Then the lower bounds for the risks R^{low}_{n,μ_n} are provided in Table* 2 *above, where the lower bounds for the risk in cases* 7, 8 *and* 9 *are valid under Assumption* A4. *Here, $U_5 = 2a + 2s(1 - d/\gamma) - 2d\alpha/\gamma + 4\beta(\varsigma_0 + 1) - (U_{\tau,\varsigma_0} + 1)d/\gamma - 1$, $U_6 = 4b(\varsigma_0 + 1) + 2a + 2s - 1$ and $U_{\tau,\varsigma_0} = \min(\beta(4\varsigma_0 + 3) - 2\tau\varsigma_0 - 1, \beta(2\varsigma_0 + 3) + 2\varsigma_0 - 1)$.*

4.3. *Estimation of discrete functionals.* Observe that for $f(\theta)$ given by formula (4.1), there are two discrete functionals associated with Φ_μ defined in (1.5)

$$\Phi_{n\mu_n} = \frac{1}{n\mu_n}\sum_{i=1}^n \varphi(\theta_i)\mathbb{I}(\theta_i \ne 0), \qquad \Phi_{k_n} = \frac{1}{k_n}\sum_{i=1}^n \varphi(\theta_i)\mathbb{I}(\theta_i \ne 0), \tag{4.11}$$

where the last functional coincides with (1.4) that was considered by Cai and Low (2011) in the case of nonrandom θ_i, $i = 1, \dots, n$. Note that since $(n\mu_n)^{-1} = n^{-(1-\nu)}(n\mu_n{}^2)^{-1}$, one has

$$\begin{aligned} \mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_{n\mu_n})^2 &\le \mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_\mu)^2 + \frac{2n^{-(1-\nu)}}{n\mu_n{}^2}\int_{-\infty}^{\infty}\varphi^2(\theta) f_0(\theta)\,d\theta, \\ \mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_{n\mu_n})^2 &\ge \frac{\mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_\mu)^2}{2} - \frac{n^{-(1-\nu)}}{n\mu_n{}^2}\int_{-\infty}^{\infty}\varphi^2(\theta) f_0(\theta)\,d\theta. \end{aligned} \tag{4.12}$$

Since $n^{-(1-\nu)} \to 0$ as $n \to \infty$, similar to Corollaries 1 and 2, the minimax lower and the upper bounds for the risks of the estimators of Φ_μ and $\Phi_{n\mu_n}$ are of the same asymptotic order.

Not that in the case when θ_i are i.i.d. with the pdf f given by (4.1), k_n is the Binomial (n, μ_n) random variable, so k_n is close to $n\mu_n$ with high probability. Therefore, estimators of Φ_{k_n} exhibit similar behavior to estimators of Φ_μ and $\Phi_{n\mu_n}$. In particular, the following statement is true.

COROLLARY 4. *Let θ_i, $i = 1, \dots, n$, in (1.2) be i.i.d. with pdf f defined in (4.1) and $n\mu_n = n^\nu$. If g, φ and f_j, $j = 1, 2$, satisfy conditions of Theorem 7 where f_j, $j = 1, 2$, are such that $\int_{-\infty}^{\infty} \varphi^2(\theta) f_j(\theta)\, d\theta < \infty$ and n is large enough, then for some positive constants $\tilde{C}_0$ and $\tilde{p}_0$ independent of n one has*

$$\inf_{\widetilde{\Phi}_{k_n}} \sup_{f_1, f_2} \mathbb{P}\{(\widetilde{\Phi}_{k_n} - \Phi_{k_n})^2 \geq \tilde{C}_0 \min(1, (n\mu_n{}^2)^{-1})\} \geq \tilde{p}_0.$$

Here, $\widetilde{\Phi}_{k_n}$ is any estimator of Φ_{k_n} based on $Y_1, \dots, Y_n$ and k_n.

Recall that, in the case of Gaussian errors, Cai and Low proved that consistent estimation of Φ_{k_n} is impossible if $\nu \leq 1/2$ and $\varphi(x) = x$ [Cai and Low (2004)] or $\varphi(x) = |x|$ [Cai and Low (2011)]. Note that the lower bounds for the risk in the case when the values $\theta_1, \dots, \theta_n$ are unconstrained are higher than in the case where θ_i, $i = 1, \dots, n$, are i.i.d., so it is logical to conclude that if the lower bounds for the risk in the latter case are bounded above by a constant, the same is true for the lower bounds in the situation where $\theta_1, \dots, \theta_n$ can be generated by any mechanism. Therefore, Corollary 4 generalizes conclusions of Cai and Low (2004, 2011) to essentially any error distribution g and any function φ with the finite second moment.

REMARK 4 [Relation to Collier, Comminges and Tsybakov (2015)]. Recently, Collier, Comminges and Tsybakov (2015) considered estimation of functionals of the form $\tilde{\Phi}_n = \sum_{i=1}^n \varphi(\theta_i)$ with $\varphi(x) = x$ or $\varphi(x) = x^2$ where θ_i, $i = 1, \dots, n$, form an arbitrary sparse sequence, $Y_i = \theta_i + \sigma\xi_i$ and ξ_i are i.i.d. standard normal variables. The results of Collier, Comminges and Tsybakov (2015) are nonasymptotic. They are stated for the functional $\tilde{\Phi}_{k_n}$ but after division by k_n they obviously apply to Φ_{k_n}. Therefore, they imply the same consequences for Φ_{k_n} as those proved above when being considered as asymptotic.

5. Simulation study.

In order to evaluate small sample properties of the estimators presented in the paper, we carried out a limited simulation study. In particular, we compared our estimator $\widehat{\Phi}_{1,h} = \widehat{\Phi}_{1,1} + \widehat{\Phi}_{1,2,h}$ of the first absolute moment of the mixing density Φ_1, where $\widehat{\Phi}_{1,1}$ and $\widehat{\Phi}_{1,2,h}$ are defined in (3.26) and (3.27), respectively, with the estimator $\widehat{\Phi}_{CL}$ of Cai and Low (2011) which is based on approximation of the absolute value by combination of Chebyshev and Hermite polynomials:

$$\widehat{\Phi}_{CL} = \sum_{k=1}^{K_*} G_{2k}^* M_0^{1-2k} B_{2k}. \tag{5.1}$$

Here, M_0 is such that $|\theta_i| \le M_0$, $i = 1, \dots, n$, G_{2k}^* is the coefficient for θ^{2k} in the expansion of $|\theta|$ by Chebyshev polynomials and $B_{2k} = n^{-1} \sum_{i=1}^n H_{2k}(y_i)$, where $H_{2k}(x)$ are Hermite polynomials [with respect to $\exp(-x^2/2)$] of degree $2k$. We considered three choices for the mixing density f: f is Gaussian $\mathcal{N}(m_\theta, \sigma_\theta^2)$, f is uniform on the interval $[a, b]$ and f is a is combination of delta functions

$$f(x) = \sum_{k=1}^{K} b_k \delta(x - a_k). \tag{5.2}$$

Note that, in the latter case, variable θ is discrete and does not have a pdf in a regular sense. Complete description of the simulation study can be found in Section 7.1 of the supplementary material.

Simulations confirm that, if one uses the exact value of M_0 in (5.1), estimators $\widehat{\Phi}_{\widehat{h}_n}$ and $\widehat{\Phi}_{CL}$ have very similar precisions: the difference between the average errors of the two estimators is smaller than respective standard deviations. As it was expected, estimator $\widehat{\Phi}_{CL}$ performs better when θ is a discrete random variable while the set up where θ is a continuous random variable benefits $\widehat{\Phi}_{\widehat{h}_n}$. Also, though $\widehat{\Phi}_{\widehat{h}_n}$ is designed to estimate Φ while $\widehat{\Phi}_{CL}$ is intended to estimate Φ_n, the error Δ_n turns out to be smaller for $\widehat{\Phi}_{\widehat{h}_n}$ while $\widehat{\Phi}_{CL}$ is somewhat more accurate as an estimator of Φ. The latter confirms that, if θ_i are i.i.d. random variables, the problems are equivalent up to a small additive error. However, the advantage of $\widehat{\Phi}_{\widehat{h}_n}$ is that it is adaptive in a finite sample setting: in all three cases, we use a date-driven value $\hat{h}$ of h. On the other hand, our simulations show that estimator $\widehat{\Phi}_{CL}$ is very sensitive to the choice of M_0 and works only in asymptotic setting. Indeed, the value $\widehat{M_0} = \sqrt{2 \log n}$ suggested in Cai and Low (2011) requires n to be large (e.g., $n > 355{,}000$ if $M_0 = 5$). If $\widehat{M_0} < M_0$, precision of the estimator deteriorates very significantly. The latter demonstrated the advantage of the estimator $\widehat{\Phi}_{\widehat{h}_n}$ in comparison with $\widehat{\Phi}_{CL}$.

APPENDIX

A.1. Upper bounds for the risks: Supplementary lemma. Note that the upper bounds for the risks of the estimators in the paper can usually be bounded by $H(n, h) \equiv H(n, h; A_1, A_2, b, d, \beta, \gamma)$ where

$$\begin{aligned} H(n, h) &= h^{2A_1} \exp(-2dh^{-b}) \\ &\quad + n^{-1} \int_0^{1/h} (\omega^2 + 1)^{A_2} \exp(2\gamma \omega^\beta - 2d\omega^b)\, d\omega. \end{aligned} \tag{A.1}$$

The following statement (which is proved in Section 7.3 of the supplemental material [Pensky (2017)]) provides upper bounds for expression (A.1).

LEMMA 2. *Let* $\Delta_n \equiv \Delta_n(A_1, A_2, b, d, \beta, \gamma) = H(n, \tilde{h}_n; A_1, A_2, b, d, \beta, \gamma)$ *and* $H(n, h) \equiv H(n, h; A_1, A_2, b, d, \beta, \gamma)$.

Then, for $h_n^*(t) = [\frac{1}{2\gamma}(\log n - (t/\beta)\log\log n)]^{-\frac{1}{\beta}}$ *and* $V_1 = 2A_1^*(d-\gamma) + d(2A_2 + 1 - \beta)$ *and* $V_2 = 2\min(A_1, A_3)$ *with* $A_1^* \le A_1$ *one has*

(1) $\Delta_n \asymp n^{-1}, \tilde{h}_n = 0$	*if* $b > \beta$,
(2) $\Delta_n \asymp n^{-1}, \tilde{h}_n = 0$	*if* $b = \beta, d > \gamma > 0$,
(3) $\Delta_n \asymp n^{-1}, \tilde{h}_n = 0$	*if* $b = \beta, d = \gamma, A_2 < -1/2$,
(4) $\Delta_n \asymp n^{-1}\log\log n, \tilde{h}_n = h_n^*(0)$	*if* $b = \beta > 0, d = \gamma > 0$, $A_2 = -1/2$,
(5) $\Delta_n \asymp n^{-1}\log n, \tilde{h}_n = n^{-1}$	*if* $b = \beta = 0, d = \gamma = 0$, $A_2 = -1/2$,
(6) $\Delta_n \asymp n^{-1}(\log n)^{\frac{2A_2+1}{b}}, \tilde{h}_n = h_n^*(0)$	*if* $b = \beta > 0, d = \gamma > 0$, $A_2 > -1/2$,
(7) $\Delta_n \asymp n^{-\frac{2A_1}{2A_1+2A_2+1}}, \tilde{h}_n = n^{-\frac{1}{2A_1+2A_2+1}}$	*if* $b = \beta = 0, d = \gamma = 0$, $A_2 > -1/2$,
(8) $\Delta_n \asymp (\log n)^{-\frac{V_1}{\beta\gamma}} n^{-d/\gamma}$, $\tilde{h}_n = h_n^*(2A_1^* + 2A_2 + 1 - \beta)$	*if* $b = \beta > 0, \gamma > d > 0$,
(9) $\Delta_n \asymp (\log n)^{-\frac{V_2}{\beta}} \exp\left(-\frac{d}{\gamma}[\log n]^{b/\beta}\right)$, $\tilde{h}_n = h_n^*(2A_3 + 2A_2 + 1 - \beta)$	*if* $\beta > b > 0, d > 0, \gamma > 0$,
(10) $\Delta_n \asymp (\log n)^{-\frac{2A_1}{\beta}}, \tilde{h}_n = [\log n/(3\gamma)]^{-1/\beta}$	*if* $b = d = 0, \beta > 0, \gamma > 0$.

PROOF OF THEOREM 2. Validity of this theorem follows immediately from Lemma 2 with $A_1 = s + a - 1/2 + b/2\mathbb{I}(d > 0)$ and $A_2 = \alpha - a$. $\square$

A.2. Proofs of statements in Section 3.

PROOF OF LEMMA 1. The statement is based on formulas 3.721.1, 8.230 and 8.231 of Gradshtein and Ryzhik (1980) which imply that for any $0 \le \lambda_1 \le \lambda_2 < \infty$ one has

$$\frac{2}{\pi}\int_0^\infty \frac{\sin(ax)}{x}\,dx = \operatorname{sign}(a), \qquad \left|\int_{\lambda_1}^{\lambda_2} \frac{\sin(ax)}{x}\,dx\right| \le \pi, \tag{A.2}$$

where sign(a) is the sign of a. Therefore, for any $u(\theta)$, it follows from (A.2) and the Lebesgue dominated convergence theorem that

$$\begin{aligned}\int_{-\infty}^\infty \operatorname{sign}(\theta - t)u(\theta)\,d\theta &= \frac{2}{\pi}\int_{-\infty}^\infty u(\theta)\int_0^\infty \frac{\sin((\theta-t)\omega)}{\omega}\,d\omega\,d\theta \\ &= \frac{2}{\pi}\lim_{\substack{\lambda_1\to 0\\ \lambda_2\to\infty}}\int_{-\infty}^\infty u(\theta)\int_{\lambda_1}^{\lambda_2}\frac{\sin((\theta-t)\omega)}{\omega}\,d\omega\,d\theta.\end{aligned}$$

Since in the last integral $|\sin((\theta - t)\omega)/\omega| \le \lambda_1^{-1}$, Fubini's theorem yields

$$\int_{-\infty}^{\infty} \operatorname{sign}(\theta - t) u(\theta)\, d\theta = \frac{2}{\pi} \lim_{\substack{\lambda_1 \to 0 \\ \lambda_2 \to \infty}} \int_{\lambda_1}^{\lambda_2} \frac{1}{\omega} \int_{-\infty}^{\infty} \sin\bigl((\theta - t)\omega\bigr) u(\theta)\, d\theta\, d\omega,$$

which completes the proof. $\square$

PROOF OF THEOREM 4. For derivation of (3.23), observe that $\mathbb{E}\widehat{\Phi}_1 = \Phi_1$ and, moreover, due to assumption A1 and conditions (3.21)–(3.16), one has $\operatorname{Var}(\widehat{\Phi}_1) \le Cn^{-1}$. For the same reason, the values of $A_{m,j,k,l}$, $l = 1, 2$, are uniformly bounded, so that $\mathbb{E}\widehat{F_m(1)} = F_m(1)$ and $\operatorname{Var}[\widehat{F_m(1)}] \le Cn^{-1}$. Observe that under assumption (3.22), one has

$$\operatorname{Var}[\widehat{\Phi}_{20h}] \asymp n^{-1} \int_1^{h^{-1}} (\omega^2 + 1)^{\alpha - (m+1)a_m} \exp\bigl(-2d\omega^b + 2\gamma\omega^\beta\bigr)\, d\omega$$

$$\mathbb{E}\widehat{\Phi}_{20h} - \Phi_{20} \asymp \int_{h^{-1}}^{\infty} (\omega^2 + 1)^{-(m+1)a_m} \exp\bigl(-2d\omega^b\bigr) \bigl|f^*(\omega)\bigr|\, d\omega.$$

In order to complete the proof, note that $\mathbb{E}\widehat{\Phi}_h - \Phi = \mathbb{E}\widehat{\Phi}_{20h} - \Phi_{20}$ and $\operatorname{Var}[\widehat{\Phi}_h] \le \operatorname{Var}[\widehat{\Phi}_{20h}] + Cn^{-1}$ and use the Cauchy–Schwarz inequality if $f \in \Omega_s(B)$ or upper bounds for $|f^*|$ if $f \in \Xi_s(B)$. $\square$

Acknowledgments. The author is very grateful to Alexandre Tsybakov, Oleg Lepski and Alexander Goldenshluger for extremely valuable suggestions and discussions.

SUPPLEMENTARY MATERIAL

Supplementary Material (DOI: 10.1214/16-AOS1498SUPP; .pdf). Supplement contains complete description of the simulation study (Section 7.1), as well as Proofs of Theorem 1, Lemma 2, Theorem 3, Theorems 5–8 and Corollary 4.

REFERENCES

BERKSON, J. (1950). Are there two regression problems? *J. Amer. Statist. Assoc.* **45** 164–180.

BUTUCEA, C. and COMTE, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli* **15** 69–98. MR2546799

CAI, T. T., JIN, J. and LOW, M. G. (2007). Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.* **35** 2421–2449. MR2382653

CAI, T. T. and LOW, M. G. (2004). Minimax estimation of linear functionals over nonconvex parameter spaces. *Ann. Statist.* **32** 552–576. MR2060169

CAI, T. T. and LOW, M. G. (2011). Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *Ann. Statist.* **39** 1012–1041. MR2816346

COLLIER, O., COMMINGES, L. and TSYBAKOV, A. B. (2015). Minimax estimation of linear and quadratic functionals on sparsity classes. Available at arXiv:1502.00665.

DATTNER, I., GOLDENSHLUGER, A. and JUDITSKY, A. (2011). On deconvolution of distribution functions. *Ann*. *Statist*. **39** 2477–2501. MR2906875

DELAIGLE, A. (2007). Nonparametric density estimation from data with a mixture of Berkson and classical errors. *Canad*. *J*. *Statist*. **35** 89–104. MR2345376

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann*. *Statist*. **32** 962–994. MR2065195

GIL-PELAEZ, J. (1951). Note on the inversion theorem. *Biometrika* **38** 481–482. MR0045992

GRADSHTEIN, I. S. and RYZHIK, I. M. (1980). *Tables of Integrals*, *Series*, *and Products*. Academic Press, New York.

HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann*. *Statist*. **38** 1686–1732. MR2662357

LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann*. *Statist*. **25** 929–947. MR1447734

LEPSKIĬ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatn*. *Primen*. **36** 645–659. MR1147167

PENSKY, M. (2017). Supplement to "Minimax theory of estimation of linear functionals of the deconvolution density with or without sparsity." DOI:10.1214/16-AOS1498SUPP.

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. MR2724359

ZAYED, A. I. (1996). *Handbook of Function and Generalized Function Transformations*. CRC Press, Boca Raton, FL. MR1392476

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CENTRAL FLORIDA
ORLANDO, FLORIDA 32816-1354
USA
E-MAIL: Marianna.Pensky@ucf.edu