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

BY MARIANNA PENSKY

University of Central Florida

The present paper considers the problem of estimating a linear functional
\[ \Phi_1 = \int_{-\infty}^{\infty} \varphi(x) f(x) \, dx \]
of an unknown deconvolution density \( f \) on the basis of \( n \) i.i.d. observations, \( Y_1, \ldots, Y_n \) of \( Y = \theta + \xi \), where \( \xi \) has a known pdf \( g \), and \( f \) is the pdf of \( \theta \). The objective of the present paper is to develop a general minimax theory of estimating \( \Phi_1 \), 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 \( \Phi \) (and \( \Phi_n \)) and derive upper and minimax lower bounds for the risk for an arbitrary square integrable function \( \varphi \). Furthermore, using technique of inversion formulas, we extend the theory to a number of situations when the Fourier transform of \( \varphi \) does not exist, but \( \Phi_1 \) 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 \( \theta \) is sparse and the objective is estimating \( \Phi \) (or \( \Phi_n \)) 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
\[ \Phi = \int_{-\infty}^{\infty} \varphi(x) f(x) \, dx \]
of an unknown deconvolution density \( f \) on the basis of observations
\[ Y_i = \theta_i + \xi_i, \quad i = 1, \ldots, n. \]
Here, \( \theta_i \) are i.i.d. random variables with unknown pdf \( f \), and \( \xi_i \) are i.i.d. random errors with a known pdf \( g \). The density \( f \) is sometimes referred to as the mixing
density. The vector $\mathbf{\theta} = (\theta_1, \ldots, \theta_n)$ 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 $\Phi$ in (1.1) appears in many contexts. If $\varphi(x) = \delta(x - x_0)$, then $\Phi$ is the value of the unknown deconvolution density $f$ at the point $x_0$. If $\varphi(x) = \mathbb{I}(x < x_0)$, then $\Phi = \hat{f}(\omega_0)$, the characteristic function of the mixing distribution at $\omega = \omega_0$. If $\varphi(x) = f_\eta(x - x_0)$, where $f_\eta$ is a known pdf, then $\Phi = \Phi_1(x_0)$ 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 $\Phi$ is, respectively, the $k$th moment or the $(2M+1)$-th absolute moment of $f$.

In addition, the problem of estimating $\Phi_1$ in (1.1) can be related to estimating functionals in indirect observations

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

Indeed, if $\theta_i$, $i = 1, \ldots, n$ are i.i.d. with the pdf $f$ and $\mathbb{E}|\varphi(\theta)| < \infty$, then $\Phi_n$ in (1.3) can be viewed as an “estimator” of $\Phi = \mathbb{E}\varphi(\theta)$ in (1.1) on the basis of “observations” $\theta_i$. Moreover, as long as $\mathbb{E}|\varphi(\theta)|^2 < \infty$, one has $\mathbb{E}(\Phi_n - \Phi)^2 \leq n^{-1} \mathbb{E}|\varphi(\theta)|^2$. Hence, the minimax risks for estimating $\Phi_n$ and $\Phi$ 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 $\mathbf{\theta}$ 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 $\mathbf{\theta}$. The latter problem can be summarized as the problem of estimating

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

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 $\nu \leq 1/2$. 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 $\mathbf{\theta}$ are i.i.d. with the pdf $f$, then the random number of nonzero components $k_n$ is Binomial($n, \mu_n$) 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 $\mathbf{\theta}$.
and \( \mu_n = n^{\nu-1} \). If \( n \) is large enough, then with high probability \( k_n \) is close to \( n \mu_n \) and \( \Phi_{k_n} \) in (1.4) corresponds to

\[
(1.5) \quad \Phi_\mu = \int_{-\infty}^{\infty} \varphi(x) f_0(x) \, dx.
\]

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 \( n \mu_n \) is large enough, then convergence rates are determined by the “effective” sample size \( n \mu_n^2 = n^{2\nu-1} \). The latter proves that if \( \theta_i \) are i.i.d., then the conclusion of Cai and Low (2011) that consistent estimation is impossible whenever \( \nu \leq 1/2 \) 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 \( \varphi \) 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 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 \( \varphi \) 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 \( \theta \)) 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 \( \Phi \) is represented as a linear functional of Fourier transform \( f^\ast \) 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 \( \Phi \) 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 $\theta$ 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 $\Phi$ in (1.1) can be represented as

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

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 $\varphi$ [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 $\Phi$ 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

$$\Omega_s(B) = \left\{ t^* : \int_{-\infty}^{\infty} |t^*(\omega)|^2 (\omega^2 + 1)^s \, d\omega \leq B^2 \right\}, \quad s \geq 0.$$

We introduce the following assumptions on the known functions $g$ and $\varphi$.

A1. There exist nonnegative constants $C_{g1}$, $C_{g2}$, $\alpha$, $\beta$ and $\gamma$ such that

$$|g^*(\omega)| \geq C_{g1}(\omega^2 + 1)^{-\alpha/2} \exp(-\gamma |\omega|^\beta),$$

$$|g^*(\omega)| \leq C_{g2}(\omega^2 + 1)^{-\alpha/2} \exp(-\gamma |\omega|^\beta),$$

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

$$|\varphi^*(\omega)| \geq C_{\varphi1}(\omega^2 + 1)^{-a/2} \exp(-d |\omega|^b),$$

$$|\varphi^*(\omega)| \leq C_{\varphi2}(\omega^2 + 1)^{-a/2} \exp(-d |\omega|^b),$$

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

\begin{equation}
R_n^{\text{low}} = R_n(\Omega_1(B)) = \inf_{\tilde{\Phi}} \sup_{f \in \Omega_1(B)} E(\tilde{\Phi} - \Phi)^2, 
\end{equation}

where $\tilde{\Phi}$ is any estimator of $\Phi$ 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

\begin{equation}
\frac{|(g^*)'(\omega)|}{|g^*(\omega)|} \leq C_g (1 + |\omega|)^\tau, \quad \tau \geq 0, \text{ with } \tau = 0 \text{ if } \gamma = 0.
\end{equation}

Let there exist $\omega_0 \in (0, \infty)$ such that, for $|\omega| > \omega_0$, $\rho(\omega) = \arg(\varphi^*(\omega))$ is twice continuously differentiable with $|\rho^{(j)}(\omega)| \leq \rho < \infty$, $j = 0, 1, 2$. 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 $\tilde{\Phi}_n$ of $\Phi_n$ defined in (1.3) can be viewed as an estimator of $\Phi$, due to inequality

\[ E(\tilde{\Phi}_n - \Phi_n)^2 \geq 0.5E(\tilde{\Phi}_n - \Phi)^2 - 2n^{-1}\|\varphi\|_\infty^2, \]

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

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

\begin{equation}
R_n(\Phi_n, \Omega_1(B)) = \inf_{\tilde{\Phi}_n} \sup_{f \in \Omega_1(B)} E(\tilde{\Phi}_n - \Phi_n)^2 \geq CR_n(\Omega_1(B)),
\end{equation}

where $R_n(\Omega_1(B))$ is given in Theorem 1.
where \(S_{\Omega_1s}(B)\) is given by the following inequality:

\[ q \text{ is the unbiased estimator of } \frac{1}{n} \sum_{j=1}^{n} e^{i\omega Y_j} \]

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 \(\hat{\Phi}_h\) over the Sobolev class \(\Omega_s(B)\)

\[ R_n(\hat{\Phi}_h, \Omega_s(B)) = \sup_{f \in \Omega_s(B)} \mathbb{E}(\hat{\Phi}_h - \Phi)^2 \]

is given by the following inequality:

\[
R_n(\hat{\Phi}_h, \Omega_s(B)) \leq \frac{\|g\|_{\infty}^2}{2\pi n} \int_{-h}^{h} \frac{|\varphi^*(\omega)|^2}{|g^*(\omega)|^2} d\omega
+ \frac{B^2}{4\pi^2} \int_{-\infty}^{\infty} \frac{|\varphi^*(\omega)|^2}{(\omega^2 + 1)^{\alpha}} d\omega,
\]

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 $R_n^{up} = R_n(\hat{\Phi}_{hn}, \Omega_s(B))$ provided in Table 1, where $U_\tau = \min(7\beta - 2\tau - 1, 5\beta + 1)$, $U_1 = 2a + 2s(1 - d/\gamma) - 2d\alpha/\gamma + 8\beta - (U_\tau + 1)d/\gamma - 1$, $U_3 = (2s + 2a + b - 1) - d(2\alpha + 2s)/\gamma$ 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; $h_n = n^{-\frac{1}{2\alpha}}$ 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 $\hat{\Phi}_{hn}$ when it is used as an estimator of the functional $\Phi_n$ defined in (1.3).

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

2.4. Adaptive estimation. Note that in the expressions for the optimal value of bandwidth $\hat{h}_n$ in Theorem 2, parameters $a, \alpha, b, d, \beta$ and $\gamma$ 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 $h_n$ depends on $s$. This occurs only in cases 7, 8 and 9. In cases 8 and 9, one can easily provide an alternative value $h = \tilde{h}_n$ that introduces an additional logarithmic factor of $n$ into the expression of the risk. In case 7, one can use the Lepski method for construction of $\tilde{h}_n$ [see, e.g., Lepskii (1991), Lepski, Mammen and Spokoiny (1997)]. In order to apply the method, consider the set of bandwidths

$$H = \{h_j = n^{-\frac{1}{2\alpha}} 2^{-j}, j = 0, \ldots, J\}$$

with $2 \leq (\log n)^{-1} n^{\frac{2\alpha - 1}{2\alpha}}$, where $J$ is the largest positive integer satisfying inequality above. Denote

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

where $C_\Phi$ is such that

$$C_\Phi \geq 4 \max\left\{\frac{C_{\varphi^2}}{\pi \sqrt{\log n}}; \frac{C_{\varphi^2}}{C_{g_1}} \left(\frac{16}{3\pi} + \frac{2\sqrt{\|g\|_\infty}}{\sqrt{\pi}}\right)\right\}$$

and constants $C_{\varphi^2}$ and $C_{g_1}$ appear in (2.6) and (2.3), respectively. The following statement provides the minimax upper bounds for the risk when the bandwidth $h = \tilde{h}_n$ 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 $\hat{R}_n^{up} = R_n(\hat{\Phi}_h, \Omega_s(B))$ provided in Table 1 with $\Delta U_3 = 2s(\gamma - d)/\gamma$ and $\Delta U_4 = 2(s - s_0)$. The corresponding values of $h = \hat{h}_n$ are $\hat{h}_n = 0$ for cases 1, 2 and 3; $\hat{h}_n = n^{-1}$ for case 5; $\hat{h}_n = h_j^*$ with $j$ defined in (2.14) with $C_\Phi$ given in (2.15) for case 7; $\hat{h}_n = [\log n - \frac{1}{\beta} \log \log n]/(2\gamma))^{-\frac{1}{\beta}}$ 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 $\hat{h}_n = [\log n/(3\gamma)]^{-\frac{1}{\beta}}$ 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 $f^\infty_\infty |f^*(\omega)|^2(\omega^2 + 1)^s \exp[2s_1|\omega|^{s_2}] d\omega \leq B^2$, 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 $R_n^{up}$ 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 $\nu_n$ 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 $f_\xi$ of the random variable $\xi = \theta + \eta$ where $\theta$ and $\eta$ are independent, the pdf $f_\eta$ of $\eta$ is known and one has measurements $Y_1, \ldots, Y_n$ of random variable $Y = \theta + \xi$ of the form of (1.2) where the pdf $g$ of $\xi$ 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 $\hat{f}_\xi$ of $f_\xi$ such that for $|\omega| \to \infty$

$$\mathbb{E}\|\hat{f}_\xi - f_\xi\|^2 \leq \begin{cases} Cn^{-\min\{1.2k/(2k+1+2b)\}}, & \text{if } |f^*_\eta(\omega)/g^*(\omega)| \asymp |\omega|^b, \\ C(\log n)^{-2k/\beta}, & \text{if } |f^*_\eta(\omega)/g^*(\omega)| \asymp |\omega|^b \exp(\gamma|\omega|^{\beta}), \end{cases}$$

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

LINEAR FUNCTIONALS OF THE DECONVOLUTION DENSITY 1523
The theory developed in this paper allows one to construct an estimator of the pdf $f_{\xi}$ at a point $x_0$ with no additional effort. Let, as before, $f$, $g$ and $q$ be the pdfs of $\theta$, $\xi$ and $Y$, respectively. Then $f^*_{\xi} = f^* f^*_{\eta} = q^* f^*_{\eta}/g^*$ and

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

Therefore, $\varphi^*(\omega) = e^{ix_0\omega} f^*_{\eta}(-\omega)$, so that $|\varphi^*(\omega)| = |f^*_{\eta}(\omega)|$. The estimator of $f^*_{\xi}(x_0)$ is of the form (2.10) and Theorems 1 and 2 give the upper and the minimax lower bounds for the risk of estimating $f^*_{\xi}$ at a point $x_0$. In addition, Theorem 3 provides an adaptive estimator of $f^*_{\xi}(x_0)$ 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 $f^*_{\eta}$ 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 $\Phi$ 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 $\Phi$. Consider, for example, estimation of $\Phi_m = \int_{-\infty}^{\infty} \theta^m f(\theta) d\theta$, the $m$th moment of $f(\theta)$. Note that if $\psi_m(y)$ is a solution of the equation

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

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

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

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

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

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 $\nu_{2m} < \infty$, $\Phi_m$ can be estimated by $\hat{\Phi}_m = n^{-1} \sum_{l=1}^{n} \psi_m(Y_l)$, where $\mathbb{E} \hat{\Phi}_m = \Phi_m$, $\text{Var}[\hat{\Phi}_m] \leq C_m n^{-1}$ and constant $C_m$ depends only on $m$, $\mu_{2m}$ and $\nu_{2m}$.
Note that although we did not use Fourier transform for estimation of $\Phi_m$ 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 $m$th derivative of the Dirac delta function [see, e.g., Zayed (1996)]. However, using the Fourier transform of $\varphi$ 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

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

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 $\mathbb{P}(\theta \leq t) = 1/2 - 1/2 \text{sign}(\theta - t)$ and Lemma 1, the cdf $F(t)$ can be represented as $F(t) = 0.5 - 0.5\Phi(t)$ where

$$
\Phi(t) = \int_{-\infty}^{\infty} \text{sign}(\theta - t) f(\theta) d\theta = \frac{2}{\pi} \int_0^\infty \frac{1}{\omega} \left[ \int_{-\infty}^{\infty} \sin((\theta - t)\omega) f(\theta) d\theta \right] d\omega
$$

$$
= \frac{2}{\pi} \int_0^\infty \frac{\cos(t\omega)}{\omega} \Re[f^*(\omega)] d\omega - \frac{2}{\pi} \int_0^\infty \frac{\sin(t\omega)}{\omega} \Im[f^*(\omega)] d\omega.
$$

**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 $u(\theta)\mathbb{P}(\theta > \theta_0) \in L^2(-\infty, \infty)$ and $u(\theta)\text{sign}(\theta - \theta_0) \in L^2(-\infty, \infty)$ for any $\theta_0$, 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 \( |\theta|^{2M+1} = \theta^{2M+1} \text{sign}(\theta) \), using Lemma 1, rewrite \( \Phi_{2M+1} \) as

\[
\Phi_{2M+1} = \frac{2}{\pi} \int_{0}^{\infty} \frac{1}{\omega} \int_{-\infty}^{\infty} \sin(\omega\theta) \theta^{2M+1} f(\theta) d\theta d\omega \\
= \frac{2}{i^{2M+1} \pi} \int_{0}^{\infty} \frac{1}{\omega} \int_{-\infty}^{\infty} \theta^{2M+1} \left[ \frac{d^{2M+1}}{d\omega^{2M+1}} \int_{-\infty}^{\infty} e^{i\omega\theta} f(\theta) d\theta \right] d\omega.
\]

Therefore,

\[
\Phi_{2M+1} = (-1)^{M+1} \frac{2}{\pi} \int_{0}^{\infty} \frac{1}{\omega} \frac{d^{2M+1} \eta[f^*(\omega)]}{d\omega^{2M+1}} d\omega. \tag{3.6}
\]

### 3.2. Construction of the estimators and evaluation of their risks

Observe that in all three examples above, the linear functionals \( \Phi \) can be presented as a combination of two integrals

\[
\Phi = \int_{0}^{\infty} \psi_{m1}(\omega) \frac{d^m}{d\omega^m}[f^*(\omega)] d\omega + \int_{0}^{\infty} \psi_{m2}(\omega) \frac{d^m}{d\omega^m}[f^*(\omega)] d\omega, \tag{3.7}
\]

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 \( \Phi \) in (3.7), we partition the area of integration into \( A_1 = [0; 1] \) and \( A_2 = (1, \infty) \) and rewrite \( \Phi \) as \( \Phi = \Phi_1 + \Phi_2 \) where

\[
\Phi_k = \int_{A_k} \psi_{m1}(\omega) \eta[f^*(\omega)] d\omega + \int_{A_k} \psi_{m2}(\omega) \eta[f^*(\omega)] d\omega, \quad k = 1, 2. \tag{3.8}
\]
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 $\omega$ 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), $\Phi_1$ in (3.8) can be rewritten as

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

where

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

Denote

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

and construct their respective unbiased estimators as

$$\hat{v}_{j1}(\omega) = \int_{-\infty}^{\infty} y^j q(y) \cos(\omega y) dy, \quad \hat{v}_{j2}(\omega) = \int_{-\infty}^{\infty} y^j q(y) \sin(\omega y) dy$$

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

$$\hat{u}_{j1}(\omega) = \begin{cases} (-1)^{j/2} \hat{v}_{j1}(\omega), & \text{if } j \text{ is even,} \\ (-1)^{(j+1)/2} \hat{v}_{j2}(\omega), & \text{if } j \text{ is odd,} \end{cases}$$

$$\hat{u}_{j2}(\omega) = \begin{cases} (-1)^{j/2} \hat{v}_{j2}(\omega), & \text{if } j \text{ is even,} \\ (-1)^{(j-1)/2} \hat{v}_{j1}(\omega), & \text{if } j \text{ is odd.} \end{cases}$$

Combination of formulae (3.9)–(3.13) imply that $\Phi_1$ can be estimated by

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

Denote

$$\sigma_{j1}^2(\omega) = n \text{Var}[\hat{u}_{j1}(\omega)] \quad \text{and} \quad \sigma_{j2}^2(\omega) = n \text{Var}[\hat{u}_{j2}(\omega)]$$

and introduce the following assumption:
A3. There exists an absolute constant $C_\sigma$ such that for any $j = 0, \ldots, m$, and $k = 1, 2$, one has

\begin{equation}
\int_0^1 \left| \frac{d^{m-j}}{d\omega^{m-j}} \left( \frac{1}{g^*(\omega)} \right) \right|^2 \left| \psi_{mk}(\omega) \right|^2 \sigma_{jk}^2(\omega) d\omega \leq C_\sigma.
\end{equation}

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

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

\begin{equation}
F_m(\omega) = \sum_{k=1}^{m} (-1)^k \left[ \frac{d^{k-1}\psi_{m1}(\omega)}{d\omega^{k-1}} \left( \frac{d^m f^*(\omega)}{d\omega^m} - \psi_{mk}(\omega) \right) \right] \left( -1 \right)^{m-k} \sum_{j=0}^{m-k} (-1)^j \left[ \frac{d^{k-j}}{d\omega^{k-j}} \psi_{mk}(\omega) \right] \left( \frac{d^m \psi_{mk}(\omega)}{d\omega^m} \right),
\end{equation}

\begin{equation}
\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.
\end{equation}

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 $\Phi_{20}$ as

\begin{align*}
F_m(1) &= \sum_{k=1}^{m} \sum_{j=0}^{m-k} (-1)^k \left( \frac{m-k}{j} \right) \left[ A_{m,j,k,1}(1) u_{j1}(1) + A_{m,j,k,2}(1) u_{j2}(1) \right], \\
\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,
\end{align*}

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

\begin{equation}
A_{m,j,k,l}(\omega) = \frac{d^{k-1}\psi_{lm}(\omega)}{d\omega^{k-1}} A_{m-k-j}(\omega)
\end{equation}

with $\mathcal{A}_l(\omega) = \frac{d^l}{d\sigma^l} \left( \frac{1}{g^*(\omega)} \right)$.

Therefore, we can estimate $\Phi_2$ by $\hat{F}_2h = \hat{F}_m(1) + \hat{F}_{20h}$ where

\begin{align}
\hat{F}_m(1) &= \sum_{k=1}^{m} \sum_{j=0}^{m-k} (-1)^k \left( \frac{m-k}{j} \right) \times \left[ A_{m,j,k,1}(1) \hat{u}_{j1}(1) + A_{m,j,k,2}(1) \hat{u}_{j2}(1) \right], \\
\hat{F}_{20h} &= (-1)^m \int_1^{1/h} \left[ \frac{d^m \psi_{m1}(\omega)}{d\omega^m} \hat{u}_{01}(\omega) + \frac{d^m \psi_{m2}(\omega)}{d\omega^m} \hat{u}_{02}(\omega) \right] \times \frac{1}{g^*(\omega)} d\omega.
\end{align}
LINEAR FUNCTIONALS OF THE DECONVOLUTION DENSITY 1529

and \( \mu_j \) and \( \nu_j \) are defined by (3.12) and (3.13). Finally, we estimate \( \Phi \) in (3.7) by \( \hat{\Phi}_1 + \hat{F}_m(1) + \hat{\Phi}_{20h} \) where \( \hat{\Phi}_1, \hat{F}_m(1) \) and \( \hat{\Phi}_{20h} \) are evaluated according to (3.14) and (3.18), respectively.

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

\[
\Xi_s(B) = \left\{ f : \sup_\omega \left| f^*(\omega) \left( |\omega|^s + 1 \right) \right| \leq B_2 \right\}
\]

Then the risk of the estimator \( \hat{\Phi}_h \) is given by the following statement.

**Theorem 4.** Assume that \( f \) and \( g \) are such that \( \mu_{2m} < \infty, \nu_{2m} < \infty \), where \( \mu_k \) and \( \nu_k \) 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 \( j = 0, \ldots, m \),

\[
C_g (|\omega| + 1)^j \tau \leq |g^*(\omega)| \quad \tau \geq 0, \text{ where } \tau = 0 \text{ if } \gamma = 0.
\]

Let \( \psi_{m1}^* \) and \( \psi_{m2}^* \) be such that for some positive \( C_\psi \) and nonnegative \( d, a_m \) and \( b \), for any \( j = 0, \ldots, m \), one has for \( |\omega| \geq 1 \):

\[
\left| d^j \psi_{mk}^* (\omega) \right| \leq C_\psi (\omega^2 + 1)^{-a_m(j+1)/2} \exp(-d|\omega|^b), \quad k = 1, 2.
\]

Let \( \hat{\Phi}_h = \hat{\Phi}_1 + \hat{F}_m(1) + \hat{\Phi}_{20h} \) where \( \hat{\Phi}_1, \hat{F}_m(1) \) and \( \hat{\Phi}_{20h} \) are defined in (3.14) and (3.18), respectively. Then

\[
\mathbb{E}(\hat{\Phi}_h - \Phi)^2 \leq C \left[ h^{2A} \exp \left( -\frac{2d}{h^b} \right) \right]
\]

\[
+ \int_1^{h^{-1}} \left( \omega^2 + 1 \right)^{A_0} \exp \left( -2d\omega^b + 2\gamma\omega^\beta \right) d\omega,
\]

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 (2.20), respectively.

Using Lemma 2 in Section A.1, one can obtain convergence rates and optimal bandwidth values \( \hat{h} \) 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 \( \Phi \). 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 \( \psi_{mk}^* \), \( k = 1, 2 \), 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 \( \Phi_{2M+1} \) given by (3.5).
3.3. Estimation of the \((2M+1)\)-th absolute moment of the deconvolution density. According to (3.6), \(\Phi_{2M+1}\) can be written in the form (3.7) with 
\(\psi_{2M+1,1}^*(\omega) = 2(-1)^{M+1}(\pi\omega)^{1-1}\) and \(\psi_{2M+1,2}^*(\omega) = 0\). Similar to (3.8), rewrite \(\Phi_{2M+1}\) 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 \(\Phi_{2M+1}\) evaluated over intervals \([0, 1]\) and \((1, \infty)\). Here,

\[
\Phi_{2M+1} = (-1)^M + \frac{2}{\pi} \sum_{j=0}^{2M+1} \left( \frac{2M+1}{2k+1} \right) \int_0^1 \frac{1}{\omega} A_{2M+1-j}(\omega) u_j(\omega) d\omega,
\]

where \(u_j(\omega)\) and \(A_j(\omega)\) 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 \(\Phi_{2M+1,1}\):

\[
\hat{\Phi}_{2M+1,1} = \frac{2}{\pi} \sum_{k=0}^{M} (-1)^{M+k} \left\{ \left( \frac{2M+1}{2k} \right) \int_0^1 \frac{\hat{v}_{2k+1,2}(\omega)}{\omega} A_{2(M-k)}(\omega) d\omega \right. \\
\left. - \left( \frac{2M+1}{2k} \right) \int_0^1 \frac{\hat{v}_{2k,1}(\omega)}{\omega} A_{2(M-k)}(\omega) d\omega \right\}.
\]

Note that for \(\sigma_j^2(\omega)\) defined in (3.15), one has \(\sigma_{2j,1}^2(\omega) \propto \text{Var}[Y^{2j+1}\cos(\omega Y)] \leq \int_{-\infty}^{\infty} y^{4j} q(y) dy\) and \(\sigma_{2j+1,2}^2(\omega) \propto \text{Var}[Y^{2j+1}\sin(\omega Y)]\), so that \(\sigma_{2j+1,2}^2(\omega) \leq \min[\int_{-\infty}^{\infty} y^{4j+2} q(y) dy, \omega^2 \int_{-\infty}^{\infty} y^{4j+4} q(y) dy]\). Hence, condition (3.16) is guaranteed by \(\mu_{4M+4} < \infty\) and \(v_{4M+4} < \infty\) where \(\mu_k\) and \(v_k\) are defined in (3.2). Consider now \(\hat{\Phi}_{2M+1,2}\). Taking into account that, for any \(j = 0, 1, 2, \ldots\)

\[
\frac{d_j \psi_{2M+1,1}^*}{d\omega} = \frac{2}{\pi} \frac{(-1)^{M+1+j}}{\omega^{j+1}},
\]

apply (3.18) for this particular case to obtain

\[
\hat{\Phi}_{2M+1,2,h} = \frac{2}{\pi} \int_1^{1/h} \hat{v}_{01}(\omega) \omega^{-2M-2} [g^*(\omega)]^{-1} d\omega \\
- (-1)^M \sum_{k=1}^{2M+1} \sum_{j=0}^{M+1-k} \left( \frac{2M+1-k}{j} \right) \\
\times (k-1)! A_{2M+1-j}(1) u_j(1),
\]

where \(A_j\) is defined in (3.10). Finally, \(\Phi_{2M+1}\) can be estimated as \(\hat{\Phi}_{2M+1} = \hat{\Phi}_{2M+1,1} + \hat{\Phi}_{2M+1,2,h}\) where \(\hat{\Phi}_{2M+1,1}\) and \(\hat{\Phi}_{2M+1,2,h}\) 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 \(\hat{\Phi}_{1,1} = \hat{\Phi}_{1,1} + \hat{\Phi}_{1,2,h}\) where

\[
\hat{\Phi}_{1,1} = \frac{2}{\pi} \int_0^1 \left( \frac{\hat{v}_{01}(\omega) (g^*)'(\omega)}{\omega (g^*(\omega))^2} + \frac{\hat{v}_{12}(\omega)}{\omega g^*(\omega)} \right) d\omega,
\]
(3.27) \[
\hat{\Phi}_{1,2,h} = \frac{2}{\pi} \hat{\nu}_{01}(1) - \frac{2}{\pi} \int_1^{1/h} \frac{1}{\omega^2} \hat{\nu}_{01}(\omega) \, d\omega.
\]

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

\[
\Xi_s(B_1, B_2) = \left\{ f : \int_{-\infty}^{\infty} \theta^{4M+4} f(\theta) \, d\theta \leq B_1, \sup_{\omega} \left| f^*(\omega) \right| \left( |\omega|^2 + 1 \right) \leq B_2 \right\}
\]

(3.28)

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

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

(3.29) \[
\sup_{|\omega| \leq 1} \left| \frac{1}{\omega} d_{2k+1} \left( \frac{1}{g^*(\omega)} \right) \right| \leq C_g M, \quad 0 \leq k \leq M.
\]

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

(3.30) \[
R_n^g = \begin{cases} 
C n^{-1}, & \text{if } 2\alpha - 4M < 3, \gamma = 0, \\
C n^{-1} \log n, & \text{if } 2\alpha - 4M = 3, \gamma = 0, \\
C n^{-\frac{4M+2\alpha+2}{2\alpha+2\gamma-1}}, & \text{if } 2\alpha - 4M > 3, \gamma = 0, \\
C (\log n)^{-\frac{4M+2\alpha+2}{2\alpha+2\gamma-1}}, & \text{if } \beta > 0, \gamma > 0,
\end{cases}
\]

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

(3.31) \[
R_n^b = \begin{cases} 
C n^{-1}, & \text{if } 2\alpha - 4M \leq 3, \beta = \gamma = 0, \\
C n^{-\frac{4M+2\alpha+2}{2\alpha+2\gamma-1}}, & \text{if } 2\alpha - 4M > 3, \beta = \gamma = 0, \\
C (\log n)^{-\frac{4M+2\alpha+2}{2\alpha+2\gamma-1}}, & \text{if } \beta > 0, \gamma > 0.
\end{cases}
\]

Here, \( R_n^g = R_n(\Phi_{2M+1,h_n}, \Xi_s(B_1, B_2)) = \sup_{f \in \Xi_s(B_1, B_2)} \mathbb{E}(\Phi_{2M+1,h_n} - \Phi_{2M+1})^2 \) and \( R_n^b = R_n(\Xi_s(B_1, B_2)) = \inf_{\Phi} \sup_{f \in \Xi_s(B_1, B_2)} \mathbb{E}(\tilde{\Phi} - \Phi_{2M+1})^2 \).

Note that the values of \( \tilde{h}_n \) are independent of unknown parameters \( s, B_1 \) and \( B_2 \) if \( 2\alpha - 4M \leq 3 \) 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 \( \tilde{h}_n \) using the Lepskii method similar to the regular case with the price of an extra \log n \) in the convergence rates.
Corollary 3. Let \( \theta_i, i = 1, \ldots, n \), 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

\[
\sup_{f \in \Xi_s(B)} \mathbb{E} \left( \tilde{\Phi}_{2M+1,n} \right) - \Phi_{2M+1,n})^2 \asymp R_n(\tilde{\Phi}_{2M+1,n}, \Xi_s(B_1, B_2)),
\]

\[
\inf_{\Phi_n} \sup_{f \in \Xi_s(B)} \mathbb{E} (\Phi_n - \Phi_{2M+1,n})^2 \asymp R_n(\Xi_s(B_1, B_2)),
\]

where \( R_n(\tilde{\Phi}_{2M+1,n}, \Xi_s(B_1, B_2)) \) 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 \( \Phi_n \) of the form (1.3) with \( \varphi(\theta) = |\theta|^{2M+1} \) based on data generated by model (1.2) where the errors \( \xi_i \) are i.i.d. \( \mathcal{N}(0, \sigma^2) \) and there are no probabilistic assumptions on vector \( \theta \). They showed that

\[
\inf_{\Phi_n} \sup_{f \in \Theta_n(M_0)} \mathbb{E} (\tilde{\Phi}_n - \Phi)^2 \asymp M_0^2 \left( \frac{\log \log n}{\log n} \right)^2, \quad \inf_{\Phi_n} \sup_{f \in \Theta_n(\infty)} \mathbb{E} (\tilde{\Phi}_n - \Phi)^2 \asymp \frac{1}{\log n},
\]

where \( \Theta_n(M_0) = \{ \theta : \|\theta\|_\infty \leq M_0 \} \). 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 \( \theta_i \) are generated independently from pdf \( f \), the problem reduces to estimation of \( \Phi_1 \) in (3.5), the first absolute moment of the mixing density. Using formulae (3.26) and (3.27), one can construct an estimator \( \tilde{\Phi}_1 \) of \( \Phi_1 \).

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 \( h = \tilde{h}_n = [\log n/\sigma^2]^{-1/2} \), and Corollary 3 implies that

\[
R_n(\tilde{\Phi}_{\tilde{h}_n}, \Phi_n, \Xi_s(B_1, B_2)) \asymp R_n(\Phi_n, \Xi_s(B_1, B_2)) \asymp (\log n)^{-s/2},
\]

where \( \Xi_s(B_1, B_2) \) is defined in (3.28). Since \( f \) is a pdf, it is absolutely integrable, so that \( s \geq 0 \). 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 \( \theta_i \), their estimator addresses this “worst-case scenario” but is unable to adapt to a more favorable situation where \( |f^s(\omega)| \to 0 \) 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 $\theta_i, i = 1, \ldots, n$, 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 $\theta$ and

$$f(x) = \mu_n f_0(x) + (1 - \mu_n) \delta(x),$$

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 $\Phi_\mu$ can be recovered as

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

Therefore, we estimate $\Phi_\mu$ by

$$\hat{\Phi}_{\mu,h} = \frac{\hat{\Phi}_h}{\mu_n} - \frac{1 - \mu_n}{\mu_n} \left[ \varphi(0) - \delta_h \right]$$

with $\delta_h = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi^*(\omega) \|(|\omega| > h^{-1}) \, d\omega,$

where $\hat{\Phi}_h$ is defined in (2.10) and the correction term $\delta_h$ 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, $\text{Var}(\hat{\Phi}_{\mu,h}) = \mu_n^{-2} \text{Var}(\hat{\Phi}_h)$ where $\text{Var}(\hat{\Phi}_h)$ is bounded by

$$\text{Var}(\hat{\Phi}_h) \leq \frac{\|g\|_\infty}{2\pi n \mu_n^2} \int_{-\infty}^{\infty} \frac{|\varphi^*(\omega)|^2}{|\varphi(\omega)|^2} \|(|\omega| \leq h^{-1}) \, d\omega.$$

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

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

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

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

Hence,

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

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 $\hat{\Phi}_{\mu,h}$ of $\Phi_{\mu}$ would coincide with the upper bounds for the risk of the estimator $\hat{\Phi}_h$ of $\Phi$ in the nonsparse case if the sample size $n$ were replaced by the effective sample size $n_{\mu}$. Denote

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

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 $R_{n,\mu n}^{\text{up}} \equiv R_{n,\mu n}(\hat{\Phi}_{\mu,h}, \Omega_s(B))$ and $\tilde{R}_{n,\mu n}^{\text{up}} \equiv R_{n,\mu n}(\hat{\Phi}_{\mu,hn}, \Omega_s(B))$ of, respectively, the estimator $\hat{\Phi}_{\mu,h}$ and of its adaptive version are provided in Table 2 where $U_3 = (2s + 2a + b - 1) - d(2a + 2s)/\gamma$, $\Delta U_3 = 2s(\gamma - d)/\gamma$, $U_4 = 2s + 2a + b - 1$, $\Delta U_4 = 2(s - s_0)_+$ where $s_0$ is defined in Theorem 3, $U_5 = 2a + 2s(1 - d/\gamma) - 2d\alpha/\gamma + 4\beta(s_0 + 1) - (U_\tau,s_0 + 1)d/\gamma - 1$, $U_6 = 4b(s_0 + 1) + 2a + 2s - 1$ and $U_{\tau,s_0} = \min(\beta(4s_0 + 3) - 2\tau s_0 - 1, \beta(2s_0 + 3) + 2s_0 - 1)$.

The corresponding values of $h$ are $h = h_n$ for the nonadaptive estimator and $h = \hat{h}_n$ 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; $h_n = n^{1 - 2\nu}$ and $\hat{h}_n = \hat{f}$ where $\hat{f}$ is defined in (2.14) with $C_{\Phi}$ given by (2.15) and $n$ replaced by $n^{2\nu - 1}$ for case 7; $h_n = H_n(t)$ and $\hat{h}_n = H_n(\hat{t})$ where $H_n(x) = [(2\nu - 1)[\log n - (t/\beta) \log \log n]/(2\gamma)]^{1/2}$, 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 $h_n = \hat{h}_n = [(2\nu - 1) \log n/(3\gamma)]^{1/2}$ for case 10.

4.2. The lower bounds for the risk. The upper bounds for the risk in formula (4.4) suggest that, for $0 \leq \nu \leq 1/2$, one has $n_{\mu} = n_{\mu n}^2 = n^{2\nu - 1} \leq 1$ 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 $n_{\mu}^{-1}$.

**Theorem 7.** Let $f(x)$ be given by (4.1) and $\Phi_{\mu}$ 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

$$
I_k = \int_{-\infty}^{\infty} g^{-1}(x) \left[ \int_{-\infty}^{\infty} g(x - \theta) f_k(\theta) d\theta \right]^2 dx \leq C_I < \infty, \quad k = 1, 2,
$$

then

$$
\lim_{n \to \infty} \sup_{f_0 \in \Omega_s(B)} \mathbb{E}(\hat{\Phi}_{\mu,h} - \Phi_{\mu})^2 = 0.
$$
\[ \int \phi(\theta) \left[ f_1(\theta) - f_2(\theta) \right] d\theta \neq 0. \]

then for some positive constants \( C_{01}, C_{02} \) and \( p_0 \) independent of \( n \), and any estimator \( \tilde{\Phi}_\mu \) of \( \Phi_\mu \) based on observations \( Y_1, \ldots, Y_n \) one has

\[ \inf_{\Phi_\mu \in [f_1, f_2]} \mathbb{P}\{(\tilde{\Phi}_\mu - \Phi_\mu)^2 \geq C_{01} \min(1, n^{-1})\} \geq p_0. \]

where \( n_\mu = n_{\mu n}^2 \). In particular, \( g(x) = N(x|0, \sigma^2) \) 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 \( \Phi_\mu \) is impossible for all pdfs \( g \) satisfying condition (4.6). Moreover, Theorem 7 does not require function \( \varphi \) to be integrable or square integrable, so

**Table 2**

Asymptotic expressions for the minimax lower bounds \( R_{n,\mu_n}^{\text{low}} \) and the risks of adaptive estimators \( \hat{R}_{n,\mu_n}^{\text{up}} \)

| Cases | Bounds for the risk |
|-------|---------------------|
| 1  \( b > \beta \) | \( R_{n,\mu_n}^{\text{low}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \asymp n^{-(2\nu-1)} \) |
| 2  \( b = \beta, d > \gamma > 0 \) | \( R_{n,\mu_n}^{\text{low}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \asymp n^{-(2\nu-1)} \) |
| 3  \( b = \beta, d = \gamma, a > \alpha + 1/2 \) | \( R_{n,\mu_n}^{\text{low}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \asymp n^{-(2\nu-1)} \) |
| 4  \( b = \beta > 0, d = \gamma > 0, a = \alpha + 1/2 \) | \( R_{n,\mu_n}^{\text{low}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \cdot \log \log n \) |
| 5  \( b = \beta > 0, d = \gamma = 0, a = \alpha + 1/2 \) | \( R_{n,\mu_n}^{\text{low}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \cdot \log n \) |
| 6  \( b = \beta > 0, d = \gamma > 0, a < \alpha + 1/2 \) | \( R_{n,\mu_n}^{\text{low}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \cdot \log n \) |
| 7  \( b = \beta = 0, d = \gamma = 0, a = \alpha + 1/2 \) | \( R_{n,\mu_n}^{\text{low}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \asymp \hat{R}_{n,\mu_n}^{\text{up}} \cdot \log n \) |
| 8  \( b = \beta > 0, \gamma > d > 0 \) | \( R_{n,\mu_n}^{\text{low}} \asymp (\log n)^{-\gamma} \cdot \frac{n^{2\nu-1}}{\nu \gamma} \) |
| 9  \( \beta > b > 0, d > 0, \gamma > 0 \) | \( R_{n,\mu_n}^{\text{low}} \asymp (\log n)^{-\gamma} \cdot \frac{n^{2\nu-1}}{\nu \gamma} \cdot \exp(-2d(\frac{y}{\beta})) \) |
| 10 \( b = d = 0, \beta > 0, \gamma > 0 \) | \( R_{n,\mu_n}^{\text{low}} \asymp R_{n,\mu_n}^{\text{up}} \asymp R_{n,\mu_n}^{\text{up}} \asymp (\log n)^{-\gamma} \cdot \frac{n^{2\nu-1}}{\nu \gamma} \) |
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 $|g(\theta)| \geq C_g(\theta^2 + 1)^{-\gamma}$. Let function $g^*$ be $\varsigma_0$ times continuously differentiable, where $\varsigma_0$ is the closest integer no less than $\varsigma$, and satisfy the following condition:

$$\frac{|d^l g^*(\omega)|}{d\omega^l} \leq C_g|g^*(\omega)|(1 + |\omega|)^{\tau_l},$$

(4.9)

$$l = 1, 2, \ldots, \varsigma_0,$$ where $\tau = 0$ if $\gamma = 0$.

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

Denote

$$R_{\mu, \mu_n}^\text{low} = R_{\mu, \mu_n}(\Omega_s(B)) = \inf_{\Phi, \Phi_0 \in \Omega_s(B)} \mathbb{E}(\tilde{\Phi}_\mu - \Phi_\mu)^2,$$

where $\Phi_\mu$ is any estimator of $\Phi_\mu$ based on observations $Y_1, \ldots, Y_n$ 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 $\mu_n = n^{v-1}$ with $v > 1/2$ and $\Phi$ 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 $R_{\mu, \mu_n}^\text{low}$ 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, $U_5 = 2a + 2s(1 - d/\gamma) - 2\alpha/\gamma + 4b(\varsigma_0 + 1) - (U_{\tau, \varsigma_0} + 1)d/\gamma - 1, U_6 = 4b(\varsigma_0 + 1) + 2a + 2s - 1$ 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 $\Phi_\mu$ defined in (1.5)

$$\Phi_{n\mu_n} = \frac{1}{n\mu_n} \sum_{i=1}^n \varphi(\theta_i) \mathbb{I}(\theta_i \neq 0), \quad \Phi_{kn} = \frac{1}{kn} \sum_{i=1}^n \varphi(\theta_i) \mathbb{I}(\theta_i \neq 0),$$

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

$$\mathbb{E}(\tilde{\Phi}_{\mu, h} - \Phi_{n\mu_n})^2 \leq \mathbb{E}(\tilde{\Phi}_{\mu, h} - \Phi_\mu)^2 + \frac{2n^{-(1-v)}}{n\mu_n^2} \int_{-\infty}^{\infty} \varphi^2(\theta) f_0(\theta) d\theta,$$

(4.12)

$$\mathbb{E}(\tilde{\Phi}_{\mu, h} - \Phi_{n\mu_n})^2 \geq \frac{\mathbb{E}(\tilde{\Phi}_{\mu, h} - \Phi_\mu)^2}{2} - \frac{n^{-(1-v)}}{n\mu_n^2} \int_{-\infty}^{\infty} \varphi^2(\theta) f_0(\theta) d\theta.$$
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 $\Phi_\mu$ and $\Phi_{n\mu_n}$ are of the same asymptotic order.

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

**COROLLARY 4.** Let $\theta_i, i = 1, \ldots, n$, in (1.2) be i.i.d. with pdf $f$ defined in (4.1) and $n\mu_n = n^\nu$. If $g, \varphi$ 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

\[ \inf \sup_{\tilde{\Phi}_k} \mathbb{P}\{(\tilde{\Phi}_k - \Phi_k)^2 \geq \tilde{C}_0 \min(1, (n\mu_n^2)^{-1})\} \geq \tilde{p}_0. \]

Here, $\tilde{\Phi}_k$ is any estimator of $\Phi_k$ based on $Y_1, \ldots, Y_n$ and $k_n$.

Recall that, in the case of Gaussian errors, Cai and Low proved that consistent estimation of $\Phi_k$ 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 $\theta_1, \ldots, \theta_n$ are unconstrained are higher than in the case where $\theta_i, i = 1, \ldots, n$, 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 $\theta_1, \ldots, \theta_n$ 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 $\varphi$ 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 $\Phi_n = \sum_{i=1}^n \varphi(\theta_i)$ with $\varphi(x) = x$ or $\varphi(x) = x^2$ where $\theta_i, i = 1, \ldots, n$, form an arbitrary sparse sequence, $Y_i = \theta_i + \sigma \xi_i$ and $\xi_i$ are i.i.d. standard normal variables. The results of Collier, Comminges and Tsybakov (2015) are nonasymptotic. They are stated for the functional $\tilde{\Phi}_k$ but after division by $k_n$ they obviously apply to $\Phi_k$. Therefore, they imply the same consequences for $\Phi_k$ 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 $\hat{\Phi}_{1,h} = \hat{\Phi}_{1,1} + \hat{\Phi}_{1,2,h}$ of the first absolute moment of the mixing density $\Phi_1$, where $\hat{\Phi}_{1,1}$ and $\hat{\Phi}_{1,2,h}$ are defined in (3.26) and (3.27), respectively, with the estimator $\hat{\Phi}_{CL}$ of Cai and Low (2011) which is based on approximation of the absolute value by combination of Chebyshev and Hermite
polynomials:

\[ \hat{\Phi}_{CL} = \sum_{k=1}^{K} G^*_k M_0^{1-2k} B_{2k}. \]

Here, \( M_0 \) is such that \( |\theta_i| \leq M_0, i = 1, \ldots, n \), \( G^*_k \) is the coefficient for \( \theta^{2k} \) 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 \( \mathcal{N}(m_\theta, \sigma^2_\theta) \), \( f \) is uniform on the interval \([a, b]\) and \( f \) is a is a combination of delta functions

\[ f(x) = \sum_{k=1}^{K} b_k \delta(x - a_k). \]

Note that, in the latter case, variable \( \theta \) 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 \( \hat{\Phi}_{\hat{h}_n} \) and \( \hat{\Phi}_{CL} \) 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 \( \hat{\Phi}_{CL} \) performs better when \( \theta \) is a discrete random variable while the set up where \( \theta \) is a continuous random variable benefits \( \hat{\Phi}_{\hat{h}_n} \). Also, though \( \hat{\Phi}_{\hat{h}_n} \) is designed to estimate \( \Phi \) while \( \hat{\Phi}_{CL} \) is intended to estimate \( \Phi_n \), the error \( \Delta_n \) turns out to be smaller for \( \hat{\Phi}_{\hat{h}_n} \) while \( \hat{\Phi}_{CL} \) is somewhat more accurate as an estimator of \( \Phi \). The latter confirms that, if \( \theta_i \) are i.i.d. random variables, the problems are equivalent up to a small additive error. However, the advantage of \( \hat{\Phi}_{\hat{h}_n} \) 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 \( \hat{\Phi}_{CL} \) is very sensitive to the choice of \( M_0 \) and works only in asymptotic setting. Indeed, the value \( \tilde{M}_0 = \sqrt{2 \log n} \) suggested in Cai and Low (2011) requires \( n \) to be large (e.g., \( n > 355,000 \) if \( M_0 = 5 \)). If \( \tilde{M}_0 < M_0 \), precision of the estimator deteriorates very significantly. The latter demonstrated the advantage of the estimator \( \hat{\Phi}_{\hat{h}_n} \) in comparison with \( \hat{\Phi}_{CL} \).

**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

\[ H(n, h) = h^{2A_1} \exp(-2dh^{-b}) \]

\[ + n^{-1} \int_{0}^{1/h} (\omega^2 + 1)^{A_2} \exp(2\gamma \omega^b - 2d\omega^b) d\omega. \]

The following statement (which is proved in Section 7.3 of the suplemental material [Pensky (2017)]) provides upper bounds for expression (A.1).
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 $h_n^*(t) = \left[\frac{1}{2\pi} \int_0^{\infty} \frac{\sin(at)}{x} \, dx = \text{sign}(a), \right.$
\begin{equation}
\left. \int_{\lambda_1}^{\infty} \frac{\sin(\lambda x)}{x} \, dx \right| \leq \pi,
\end{equation}
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{equation}
\int_{-\infty}^{\infty} \text{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_{\lambda_1 \to 0} \int_{-\infty}^{\infty} u(\theta) \int_{\lambda_1}^{\infty} \frac{\sin((\theta - t)\omega)}{\omega} \, d\omega \, d\theta.
\end{equation}
Since in the last integral $|\sin((\theta - t)\omega)/\omega| \leq \lambda_1^{-1}$, Fubini’s theorem yields
\[
\int_{-\infty}^{\infty} \text{sign}(\theta - t)u(\theta) d\theta = \frac{2}{\pi} \lim_{\lambda_2 \to \infty} \int_{\lambda_2}^{\lambda_1} \frac{1}{\omega} \int_{-\infty}^{\infty} \sin((\theta - t)\omega)u(\theta) d\theta d\omega,
\]
which completes the proof. □

**Proof of Theorem 4.** For derivation of (3.23), observe that $\mathbb{E}\tilde{\Phi}_1 = \Phi_1$ and, moreover, due to assumption A1 and conditions (3.21)–(3.16), one has $\text{Var}(\tilde{\Phi}_1) \leq Cn^{-1}$. For the same reason, the values of $A_m,j,k,l$, $l = 1, 2$, are uniformly bounded, so that $\mathbb{E}\tilde{F}_m(1) = F_m(1)$ and $\text{Var}[\tilde{F}_m(1)] \leq Cn^{-1}$. Observe that under assumption (3.22), one has
\[
\text{Var}[\tilde{\Phi}_{20h}] \asymp n^{-1} \int_{h^{-1}}^{h^{-1}} (\omega^2 + 1)^{-\alpha}^{(m+1)a_m} \exp(-2d\omega^b + 2\gamma\omega^\beta) d\omega
\]
\[
\mathbb{E}\tilde{\Phi}_{20h} - \Phi_1 \asymp \int_{h^{-1}}^{\infty} (\omega^2 + 1)^{-\alpha}^{(m+1)a_m} \exp(-2d\omega^b |f^*(\omega)| d\omega.
\]
In order to complete the proof, note that $\mathbb{E}\tilde{\Phi}_h - \Phi = \mathbb{E}\tilde{\Phi}_{20h} - \Phi_20$ and $\text{Var}[\tilde{\Phi}_h] \leq \text{Var}[\tilde{\Phi}_{20h}] + Cn^{-1}$ and use the Cauchy–Schwarz inequality if $f \in \Omega_s(B)$ or upper bounds for $|f^*|$ if $f \in \Xi_s(B)$. □

**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., GOLDSHLAGER, 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

LEPSKI, 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