

Density deconvolution in a two-level heteroscedastic model with unknown error density

Alexander Meister

Institut für Mathematik

Universität Rostock

D-18057 Rostock, Germany

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

Ulrich Stadtmüller and Christian Wagner

Institut für Zahlentheorie und Wahrscheinlichkeitstheorie

Universität Ulm

D-89081 Ulm, Germany

e-mail: ulrich.stadtmueller@uni-ulm.de; christian.wagner@uni-ulm.de

Abstract: We consider a statistical experiment where two types of contaminated data are observed. Therein, both data sets are affected by additive measurement errors but the scaling factors of the error density may be different and/or the observations have been averaged over different numbers of independent replicates. That kind of heteroscedasticity of the data allows us to identify the target density although the error density is unknown and we can allow that the characteristic function of the error variables may have zeros. We introduce a novel nonparametric procedure which estimates the target density with nearly optimal convergence rates. The main goal in this paper is to derive the upper and lower bounds for the convergence rates. A small simulation study addresses the finite sample properties of the procedure.

AMS 2000 subject classifications: Primary 62G07.

Keywords and phrases: Hermite polynomials, measurement errors, min-max convergence rates, nonparametric statistics, statistical inverse problems.

Received July 2009.

1. Introduction, model and applications

We consider nonparametric density estimation based on data which are affected by additive independent measurement error and hence a deconvolution problem occurs. The basic approaches to tackle this situation go back to the papers of [Stefanski and Carroll \(1990\)](#), [Carroll and Hall \(1988\)](#), [Fan \(1991\)](#) and others. In the recent years, this field of statistics has attracted an increasing number of researchers. See e.g. the book of [Meister \(2009\)](#) for a comprehensive and recent review on nonparametric deconvolution problems.

A major drawback of the usual deconvolution techniques is the fact that the error density is required to be known. Otherwise, one faces heavy problems of non-identifiability, which is rather intuitive: Suppose we would like to estimate a density function f_X while any empirical access is restricted to some independent observations drawn from the density $f_X * f_\varepsilon$, the symmetry of convolution makes it impossible to identify f_X in the general case. Moreover, Meister (2004) shows that misspecification of the error density in the standard deconvolution kernel estimator as introduced in Stefanski and Carroll (1990) may have fatal consequences on the asymptotic behavior of the estimator.

The fact that it is unrealistic to assume to know exactly the error density in many applications has well been recognized in the statistical community. Therefore, many authors have tried to relax that condition by various modifications of the model. In order to have a situation where the measurement system can somehow be calibrated one may assume the availability of additional independent direct data from the error density, see Diggle and Hall (1993), Neumann (1997), Efromovich (1997), Johannes (2009) for that framework. Another popular model where the error density also is not required to be perfectly known in advance, uses replicated measurements of the same incorrupted random variable with the density f_X , see e.g. Horowitz and Markatou (1996), Li and Vuong (1998), Hall and Yao (2003), Neumann (2007), Delaigle et al. (2008). In the related topic of errors-in-variables regression Schennach (2004a,b) consider repeated observation and instrumental variables models.

Other concepts making the error distribution accessible require more restrictive conditions on the target density f_X . E.g. consider the standard experiment where one observes the data Y_1, \dots, Y_n generated by

$$Y_j = X_j + \sigma\varepsilon_j, \quad j = 1, \dots, n, \quad (1.1)$$

where all $X_1, \varepsilon_1, \dots, X_n, \varepsilon_n$ are independent; the X_j have the density f_X to be estimated; the error variables ε_j have the density f_ε and $\sigma > 0$ denotes a scaling parameter. In the papers of Butucea and Matias (2005), Butucea et al. (2008) and Meister (2006) this model is considered under the assumption that the Fourier transform of f_X , denoted by f_X^{ft} , has a specific known positive lower bound. Then, σ can be estimated consistently, however, the density f_ε is still assumed to be known so that f_X is finally identifiable. Those models use semiparametric approaches to the generic problem of “blind deconvolution”, i.e. deconvolution with unknown or partly known noise density. Meister (2007) establishes consistency in a model where f_X is compactly supported, σ may be put equal to one and $f_\varepsilon^{ft}(t)$ has to be known on some bounded interval $t \in [-T, T]$, $T > 0$ only.

In the present paper, we consider an observation scheme which allows for a specific type of heteroscedasticity in the data. Some recent contributions address the problem that the data may be contaminated by different error densities, e.g. when the observations are drawn from two experiments with different measurement systems; see Delaigle and Meister (2008) and Staudenmayer et al. (2008). In our model, we are given the data $Y_1, \dots, Y_n, Y'_1, \dots, Y'_n$, ($n \in \mathbb{N}$) where for

some fixed $m \in \mathbb{N}$

$$\begin{aligned} Y_j &= X_j + \varepsilon_j, \\ Y'_j &= X'_j + \sigma \sum_{k=1}^m \varepsilon'_{j,k}. \end{aligned} \quad (1.2)$$

All the random variables $X, X', \varepsilon, \varepsilon, \dots$ are independent; the X_j and X'_j have the density f_X ; the ε_j and the $\varepsilon_{j,k}$ have the error density f_ε . We allow for both f_X and f_ε to be unknown; only the integer $m > 0$ and the scaling parameter $\sigma \in (0, 1)$ if $m = 1$ and $\sigma \in (0, 1/m]$ if $m \geq 2$ are supposed to be known. For instance, in the replicated data model, the assumption of known σ located in this given interval can be justified as $\sigma = 1/m$. We will focus on that model below. Otherwise, if σ is unknown and e.g. $m = 1$, then the statistical model becomes non-identifiable, in general. Note that e.g. $X_j \sim N(0, 3), \varepsilon_j \sim N(0, 1), \sigma^2 = 1/2$ on the one hand and $X_j \sim N(0, 2), \varepsilon_j \sim N(0, 2), \sigma^2 = 3/4$ on the other hand lead to the same distribution of the observed data set; here $N(\mu, \sigma^2)$ denotes the normal density with the mean μ and the variance σ^2 . So we have to assume prior knowledge on σ .

In addition some decay constraints will be imposed on f_X , details will be deferred to Section 3 due to their technical nature. Note that another scaling parameter σ' could be added in the first line of (1.2); however, this may be absorbed by the error variable ε_j . Also, we mention that model (1.2) and our estimation procedure as introduced in the following section are fully extendable to the statistical experiment where the data

$$\begin{aligned} Y_j &= X_j + \sum_{k=1}^{\tilde{m}} \varepsilon_{j,k}, \\ Y'_j &= X'_j + \sigma \sum_{k=1}^m \varepsilon_{j,k}, \end{aligned}$$

are observed where $\tilde{m} < m$, i.e. both data sets have cumulative noise structure. Our method is applicable when putting $\varepsilon_j = \sum_{k=1}^{\tilde{m}} \varepsilon_{j,k}$ and the m in the estimator equal to the ratio of m and \tilde{m} in the above model.

In the sequel, we will discuss two examples to which model (1.2) is applicable. First, we consider the case of integer $m > 1$, $\sigma = 1/m$. As referred to above, we have repeated measurements for the data in many real life applications. However, in many cases only the averages of those replicates are reported; that problem is mentioned in e.g. [Linton and Whang \(2002\)](#). A direct application of model (1.2) occurs when for some observed individuals only one contaminated measurement is reported while each of the other represent the average of a known number of measurements of the same individual. Also see [Morris et al. \(1977\)](#) and [Thamerus \(1996\)](#) for related data sets in the field of medical statistics. [Linton and Whang \(2002\)](#) consider a grouped data model with an additional error component. They mention that when at least two data aggregates with

different size are available the error distribution is identifiable. The idea is similar to the special case $\sigma = 1/m$ of our model. A least square estimate of the cumulant of the error is roughly suggested, which is completely different from our methods. However, the properties of that estimator are not studied.

As a more specific example for the case of $m = 1$, $\sigma \in (0, 1)$, we mention the Consumer Expenditure Survey from the United States Department of Labor, which is referred to in Nelson (1994) and analyzed by e.g. Schennach (2004a). The goal is to estimate the log income of an individual while the empirical access is restricted to the log expenditure of an individual. This latter quantity can be modelled as the sum of the log income and the logarithm of a certain independent random quantity δ_j , which may be interpreted as the readiness of the individual to spend his earnings. Then, we have the standard additive measurement error model while the original model (before applying the logarithm) follows an independent multiplicative error scheme. We modify the model by assuming that the readiness of spending money depends on the total wealth of an individual rather than his income. The wealth is modelled by the income plus an independent component. Then the observed expenditure of a person is equal to his wealth multiplied by the independent random variable δ_j . Furthermore, at some point in time the individuals' income may be affected e.g. by some change in the tax rate. For data surveyed after that change the random variables representing the income are multiplied with the known factor $\sigma \in (0, 1)$. Thus, the data can be split into two independent but non-identically distributed data sets. If the distribution of δ_j is known or estimable by replicates one can generate independent data from the estimated densities and hence obtain two pseudo-data sets which approximately follow model (1.2).

The paper is organized as follows: In Section 2, we will describe our estimation procedure; its asymptotic properties are investigated in Section 3. Numerical simulations are given in Section 4, the proofs are deferred to Section 5.

2. Methodology

We begin with the definition of the classes for the target density f_X and the error density f_ε in model (1.2).

With respect to the target density f_X , we assume that f_X is β -fold continuously differentiable for some integer $\beta \geq 1$ and its derivatives satisfy

$$|f_X^{(j)}(x)| \leq C_1 \exp(-C_2 x^2), \quad C_2 > 1/4, \quad \text{for all } x \in \mathbb{R}, \quad j = 0, \dots, \beta. \quad (2.1)$$

Furthermore, we impose that

$$\int x f_X(x) dx = 0.$$

Those densities satisfying the conditions above are collected in the class $\mathcal{F} = \mathcal{F}_{\beta, C_1, C_2}$. Note that all symmetric, sufficiently smooth and compactly supported densities satisfy these conditions. Also when the error density is centered one

could think of centering the data X_j in practice. On the other hand, normal densities and some appropriate normal mixtures are also included in these constraints.

With respect to the error density f_ε , we assume that its Fourier transform satisfies $|f_\varepsilon^{ft}(t) - 1| \leq T_1|t|$ for all $|t| \leq T_0$ with $T_0, T_1 > 0$. Those densities f_ε are summarized in the class $\mathcal{G} = \mathcal{G}_{T_0, T_1}$. Thus, our conditions on f_ε are rather mild; Note that any density with finite first moment is included in \mathcal{G} for T_1 and T_0 sufficiently large and small, respectively. But the Cauchy density, for example, is also included in \mathcal{G} . In particular, no restrictions about the set of all zeros of f_ε^{ft} are required in contrary to the standard situation in deconvolution problems.

Now we address the question how to reconstruct $f_X^{ft}(t)$ on a compact interval. By the conditions contained in $f_X \in \mathcal{F}$ and the Taylor expansion of f_X^{ft} , we may establish that

$$|f_X^{ft}(t) - 1| \leq \frac{1}{2}t^2 \int |x|^2 f_X(x) dx \leq \frac{C_1 \sqrt{\pi}}{4C_2^{3/2}} t^2, \quad (2.2)$$

for all $t \in \mathbb{R}$. This implies also the existence of a uniform positive lower bound of $|f_X^{ft}|$ and $|f_\varepsilon^{ft}|$ on any compact interval $[-T, T]$ with

$$T < \min\{T_0, 2C_2^{3/4} C_1^{-1/2} \pi^{-1/4}\}. \quad (2.3)$$

That lower bound is essential as terms involving both $|f_X^{ft}|$ and $|f_\varepsilon^{ft}|$ will occur in our estimators. Writing

$$\tilde{\Phi}(t) = \prod_{k=0}^{\bar{K}} [f_X^{ft}(\sigma^k t)]^{m^k}, \quad \Phi(t) = \prod_{k=0}^{\bar{K}} [f_X^{ft}(\sigma^{k+1} t)]^{m^{k+1}}, \quad t \in \mathbb{R},$$

we deduce that

$$f_X^{ft}(t) / [f_X^{ft}(\sigma^{\bar{K}+1} t)]^{(m^{\bar{K}+1})} = \tilde{\Phi}(t) / \Phi(t), \quad (2.4)$$

holds true for all integer $\bar{K} > 0$ as the right side of (2.4) may be viewed as a telescopic product. The integer parameter \bar{K} remains to be selected. A similar technique has been used in [Belomestny \(2003\)](#) in the field of time series analysis; while the author considers data from an autoregression model, assumes that the target density is ordinary smooth and uses kernel regularisation.

Combining the condition $\sigma \leq 1/m$ and (2.2), we may conclude that $\lim_{k \rightarrow \infty} [f_X^{ft}(\sigma^k t)]^{(m^k)} = 1$, which will be made more precise in the proof of [Theorem 3.1](#). On the other hand, we have

$$\begin{aligned} \tilde{\Phi}(t) / \Phi(t) &= \prod_{k=0}^{\bar{K}} \frac{[f_X^{ft}(\sigma^k t) (f_\varepsilon^{ft}(\sigma^{k+1} t))^m]^{m^k}}{[f_X^{ft}(\sigma^{k+1} t) f_\varepsilon^{ft}(\sigma^{k+1} t)]^{m^{k+1}}} \\ &= \prod_{k=0}^{\bar{K}} \left([E \exp(i\sigma^k t Y_j')] / [E \exp(it\sigma^{k+1} Y_j)]^m \right)^{m^k}, \end{aligned} \quad (2.5)$$

so that the function $\tilde{\Phi}/\Phi$ is uniquely determined by the characteristic functions of Y_j' and Y_j . Combining that with (2.4), we notice that $f_X^{ft}(t)$ is indeed identifiable from the data in model (1.2).

We employ the empirical characteristic functions

$$\hat{\tilde{\Psi}}(t) = \frac{1}{n} \sum_{j=1}^n \exp(itY_j'), \quad \hat{\Psi}(t) = \frac{1}{n} \sum_{j=1}^n \exp(itY_j),$$

where their true counterparts are denoted by $\tilde{\Psi}(t)$ and $\Psi(t)$, respectively, and we propose

$$\hat{\tilde{\Phi}}(t) = \prod_{k=0}^{\bar{K}} [\hat{\tilde{\Psi}}(\sigma^k t)]^{m^k}, \quad \hat{\Phi}(t) = \prod_{k=0}^{\bar{K}} [\hat{\Psi}(\sigma^{k+1} t)]^{m^{k+1}},$$

as the estimators for $\tilde{\Phi}(t) h_\varepsilon(t)$ and $\Phi(t) h_\varepsilon(t)$, respectively, with the common factor $h_\varepsilon(t) = \prod_{k=1}^{\bar{K}} (f_\varepsilon^{ft}(\sigma^{k+1} t))^{m^{k+1}}$. Motivated by (2.4) and (2.5) we define, utilizing some fixed $0 < \rho < 1$, the following estimator for $f_X^{ft}(t)$,

$$\hat{\Psi}_X(t) = \frac{\hat{\tilde{\Phi}}(t) \overline{\hat{\Phi}(t)}}{\max\{|\hat{\Phi}(t)|^2, \rho\}} = \frac{\prod_{k=0}^{\bar{K}} [\hat{\tilde{\Psi}}(\sigma^k t)]^{m^k} \prod_{k=0}^{\bar{K}} \overline{[\hat{\Psi}(\sigma^{k+1} t)]^{m^{k+1}}}}{\max\left\{\prod_{k=0}^{\bar{K}} |\hat{\Psi}(\sigma^{k+1} t)|^{2m^{k+1}}, \rho\right\}}. \quad (2.6)$$

Therein, a ridge parameter $\rho \in (0, 1)$ is introduced in order to prevent the denominator from being too close to zero due to some stochastic deviation. Actually, the ridge parameter has only minor influence on the estimator as the estimators use some shrinking interval on which $\Phi(t)$ tends to one.

In deconvolution problems it is common to assume that $f_\varepsilon^{ft}(t) \neq 0$, for all $t \in \mathbb{R}$. In order to derive convergence rates, even specific lower positive bounds on $|f_\varepsilon^{ft}|$ are required (see e.g. Fan (1991)). However, such conditions are often introduced for mathematical convenience; they are satisfied for normal or Laplace densities but they are not valid for other important densities such as the convolution of uniform densities with any other distribution. The case where the error density has some periodic isolated zeros has been studied in the papers of Hall and Meister (2007) and Meister (2008). By contrast, Meister (2007) introduces a consistent estimator f_X provided that $f_X^{ft}(t)$ is compactly supported so that f_ε^{ft} is permitted to vanish on an open non-void interval. Here, we do not assume that f_X is compactly supported but use a decay constraint which is contained in the assertion $f_X \in \mathcal{F}$. The estimation procedure in Meister (2007) is based on global polynomial extension in the Fourier domain and kernel regularization. While that method achieves consistency its finite sample performance suffers from Gibbs phenomenon. We introduce a novel procedure based on an orthogonal series approach, which seems to be more appropriate. We consider the orthogonal expansion in $L_2(\mathbb{R}, \exp(-\cdot^2/2))$ of $g(x) = f(x) \exp(x^2/2)$ with respect to the normalized Hermite polynomials $H_k(x)/\sqrt{\sqrt{2\pi}k!}$, $k \geq 0$, i.e.,

$$g(x) = \sum_{k=0}^{\infty} c_k \frac{H_k(x)}{\sqrt{\sqrt{2\pi}k!}} \quad \text{with} \quad c_k = \int g(x) \frac{H_k(x)}{\sqrt{\sqrt{2\pi}k!}} \exp(-x^2/2) dx$$

which is equivalent to

$$f_X(x) = \sum_{k=0}^{\infty} \frac{1}{\sqrt{2\pi k!}} \int f_X(y) H_k(y) dy \cdot H_k(x) \exp(-x^2/2), \quad (2.7)$$

where the infinite sum converges in the $L_2(\mathbb{R})$ -sense and pointwise. Therein the space $L_2(\mathbb{R}, \exp(-\cdot^2/2))$ denotes the Hilbert space consisting of all squared-integrable functions with respect to the weight function $\exp(-\cdot^2/2)$. Note that $g \in L_2(\mathbb{R}, \exp(-\cdot^2/2))$ is guaranteed by $f \in \mathcal{F}$.

We realize that the coefficients of this expansion are a linear composition of the moments of the X_1 , which are accessible by the restriction of the Fourier transform f_X^{ft} to a bounded domain. This will be made precise in the following paragraph. This is a great advantage of the Hermite polynomial basis in the current setting. Besides the fact that f_ε^{ft} may have non-isolated zeros another aspect encourages us to apply an approach which uses an estimate of f_X^{ft} only on a bounded interval. For large $|t|$, the impact of the ridge regularization of the estimator $\hat{\Psi}_X(t)$ in (2.6) will increase, what however adds non-negligible bias. Here that phenomenon is even more critical than for usual deconvolution as the denominator does not only depend on f_ε^{ft} but also on f_X^{ft} , yielding faster decay of the unregularized denominator.

Furthermore, we have

$$\left. \frac{d^l}{dt^l} f_X^{ft}(t) \right|_{t=0} = i^l \int x^l f_X(x) dx,$$

for all integer $l \geq 0$. Next for some $K \in \mathbb{N}$ we define the quantities $\mathbf{c} = (c_0, c_1, \dots, c_K)^T$,

$$\mathbf{d} = \left((-i)^0 \left. \frac{d^0}{dt^0} f_X^{ft}(t) \right|_{t=0}, \dots, (-i)^K \left. \frac{d^K}{dt^K} f_X^{ft}(t) \right|_{t=0} \right)^T,$$

and

$$(\mathbf{W})_{l,k} = \begin{cases} \frac{1}{\sqrt{2\pi k!}} \int x^l H_k(x) \exp(-x^2/2) dx, & \text{for } l \geq k, \\ 0, & \text{otherwise,} \end{cases}$$

for $l, k = 0, \dots, K$ (note $(\mathbf{W})_{l,k}$ denotes the k th coefficient of the function \cdot^l with respect to the orthogonal basis generated by the Hermite polynomials).

We derive that with $\mathbf{e}_j = (\delta_{j,k})_{k=0, \dots, K}$,

$$\int H_j(x) f_X(x) dx = \mathbf{e}_j^T \mathbf{W}^{-1} \mathbf{d} \quad \text{for } j = 0, \dots, K. \quad (2.8)$$

Next, the vector \mathbf{d} can be estimated via the characteristic function since by integration by parts, we have for $0 \leq l \leq K$

$$\begin{aligned} \left. \frac{d^l}{dt^l} f_X^{ft}(t) \right|_{t=0} &= \lim_{b \downarrow 0} \int b^{-1} \varphi_K(t/b) \frac{d^l}{dt^l} f_X^{ft}(t) dt \\ &= (-1)^l \lim_{b \downarrow 0} b^{-l-1} \int \varphi_K^{(l)}(t/b) f_X^{ft}(t) dt, \end{aligned}$$

where the function

$$\varphi_K(x) = 1_{[-1,1]}(x) \cdot (1-x^2)^{K+2} / \left(\int_{-1}^1 (1-y^2)^{K+2} dy \right) = \frac{1_{[-1,1]}(x) \cdot (1-x^2)^{K+2}}{B(k+3, 1/2)}$$

is $(K+1)$ -fold continuously differentiable on the whole real line and supported on $[-1, 1]$ and $B(.,.)$ denotes the Beta-function. Thus we define

$$\hat{D}_l = i^l b^{-l-1} \int \varphi_K^{(l)}(t/b) \hat{\Psi}_X(t) dt$$

with $\hat{\Psi}_X$ as in (2.6); and the vector

$$\hat{\mathbf{d}} := (\hat{D}_0, \dots, \hat{D}_K)^T.$$

Therefore, combining the equations (2.7) and (2.8) motivates the final density estimator

$$\hat{f}_X(x) = \sum_{k=0}^K \frac{1}{\sqrt{2\pi k!}} (\mathbf{e}_k^T \mathbf{W}^{-1} \hat{\mathbf{d}}) \cdot H_k(x) \exp(-x^2/2), \quad (2.9)$$

where the smoothing and ridge parameters K, b, ρ, \bar{K} are still to be selected.

3. Asymptotic properties

3.1. Convergence rates – upper bounds

Throughout this paper, we write $\|f\|_p$, $p > 0$, for the $L_p(\mathbb{R}^d)$ -norm of a function $f \in L_p(\mathbb{R}^d)$ where d may be any integer larger or equal to 1. As the dimension d of the domain of f is clear for any involved function we feel that there is no need to include d in the notation of $\|\cdot\|_p$.

Theorem 3.1. *For $n \in \mathbb{N}$ we consider data drawn from the statistical model (1.2) where $\sigma \in (0, 1) \cap (0, 1/m]$. We select $\rho \in (0, 1)$, $\bar{K} = C_{\bar{K}} \log n$, $K = C_K (\log n) (\log \log n)^{-1}$, $b = (C_b \log n)^{-\gamma}$ with some finite constants $C_{\bar{K}} \in (0, 1/(2 \log m))$, $C_b > 0$, $\gamma > \max\{1, (1+\beta)/4\}$, $C_K \in (0, \min\{-C_{\bar{K}} \log \sigma / (\gamma+1), (1-2C_{\bar{K}} \log m)/(2\gamma+2)\})$. Then, for $\beta \geq 1$, the estimator (2.9) satisfies*

$$\sup_{f_X \in \mathcal{F}, f_\varepsilon \in \mathcal{G}} E \|\hat{f}_X - f_X\|_2^2 = O((\log n)^{-\beta} (\log \log n)^\beta).$$

Thus, we give the convergence rates from a double-uniform asymptotic view, i.e. the statistical risk is considered uniformly with respect to both $f_X \in \mathcal{F}$ and $f_\varepsilon \in \mathcal{G}$. The logarithmic rates established for our estimator in Theorem 3.1 coincide with those derived for density deconvolution under known normal error distributions (see Fan (1991)) up to some iterated logarithmic loss. As in supersmooth deconvolution, we realize that the exact smoothness degree β of the target density is not required to be known in advance in order to obtain those rates by parameter selection.

3.2. Convergence rates – lower bounds

As normal densities are included in our condition $f_\varepsilon \in \mathcal{G}$, it is apparently impossible to significantly improve the result in Theorem 3.1 by getting faster rates such as algebraic rates, for instance. At most, the iterated logarithmic factor could be removed. Nevertheless, in the case of $m = 1$ we will establish a logarithmic lower bound on the convergence rates with respect to an arbitrary estimator based on the given data structure (1.2); although we restrict our consideration to a subclass $\mathcal{G}' \subseteq \mathcal{G}$ of admitted error densities whose Fourier transform have a polynomial lower bound. Thus, the densities contained in \mathcal{G}' are ordinary smooth in the terminology of Fan (1991). Note that the smaller the class \mathcal{G}' the stronger the results as we are considering lower bounds. Therefore, if the error density $f_\varepsilon \in \mathcal{G}'$ was known in model (1.2), then algebraic rates could be achieved by e.g. the standard deconvolution kernel density estimator as defined in Stefanski and Carroll (1990). This highlights the fact that the slow convergence rates in Theorem 3.1 are not due to some fast decay of the Fourier transform of the error density; but allowing for the error density to be unknown really causes that remarkable deterioration of the rates from algebraic to logarithmic under the given smoothness constraints on the target density f_X . Hence, the following Theorem 3.2 will establish nearly optimality of our estimator (2.9) in the given setting.

On the other hand, assuming polynomial lower bounds on both f_ε^{ft} and on f_X^{ft} , algebraic rates are attainable by a different estimation procedure, whose applicability is restricted to the cases where the existence of those lower bounds can be justified, see Wagner (2009).

Theorem 3.2. *Consider the model (1.2) with $m = 1$. Assume that C_1 and C_2 are sufficiently large and small, respectively, and $\beta \geq 1$. We put \mathcal{G}' equal to the set of all $f_\varepsilon \in \mathcal{G}$, which satisfy $|f_\varepsilon^{ft}(t)| \geq C(1 + |t|^\alpha)^{-1}$ for all t and some fixed $C > 0, \alpha > 1/2$. Any estimator \hat{f}_X of f_X based on the data Y_1, \dots, Y_n and Y'_1, \dots, Y'_n satisfies*

$$\sup_{f_X \in \mathcal{F}, f_\varepsilon \in \mathcal{G}'} E \|\hat{f}_X - f_X\|_2^2 \geq \text{const.} \cdot (\log n)^{-\beta}, \quad \text{for } n \text{ sufficiently large.}$$

Combining the Theorem 3.1 and 3.2, we conclude that our estimator (2.9) attains nearly optimal convergence rates when $m = 1$; i.e. optimal up to some iterated logarithmic factor.

4. Numerical simulations

The finite sample size performance of our estimator (2.9) is studied based on some numerical simulations. We applied the estimator to data generated from four different target densities f_X : the normal density $N(0, 1/4)$ (Figure 1, 2), the bimodal normal mixture density $0.6 \cdot N(1, 1/4) + 0.4 \cdot N(-2, 1/4)$ (Figure 3, 4), the triangle density on $[-2, 2]$ (Figure 5, 6) and the triangle density on $[-0.5, 2.5]$ (Figure 7, 8). In the Figures 1, 3, 5, 7 we consider the case of small sample size

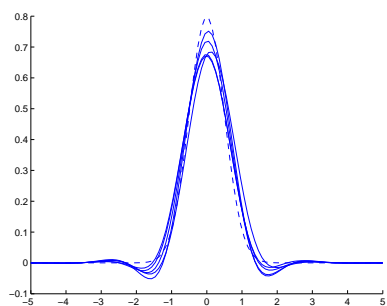


FIG 1. $(N(0, 1/4), n = 100)$.

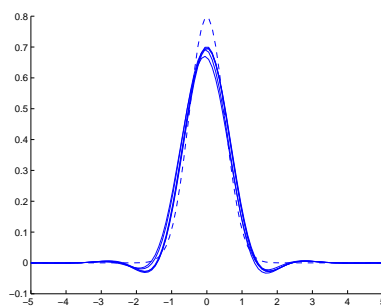


FIG 2. $(N(0, 1/4), n = 1000)$.

TABLE 1
(unimodal density)

	$n = 100$	$n = 1000$
average	0.0221	0.0116
maximum	0.0723	0.0178
minimum	0.0077	0.0082

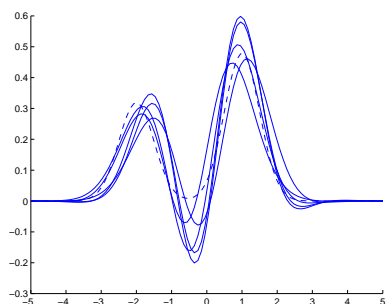


FIG 3. $(.6N(1, 1/4) + .4N(-2, 1/4), n = 100)$.

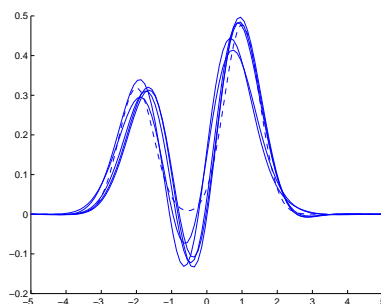


FIG 4. $(.6N(1, 1/4) + .4N(-2, 1/4), n = 1000)$.

TABLE 2
(bimodal density)

	$n = 100$	$n = 1000$
average	0.0996	0.0239
maximum	0.6806	0.0751
minimum	0.0132	0.0124

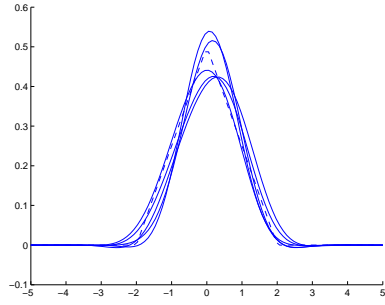


FIG 5. $(\Delta(-2, 2), n = 100)$.

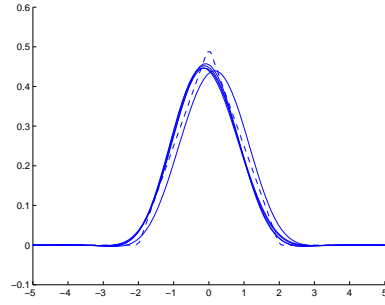


FIG 6. $(\Delta(-2, 2), n = 1000)$.

TABLE 3
(triangle density)

	$n = 100$	$n = 1000$
average	0.0146	0.0023
maximum	0.0604	0.0073
minimum	0.0013	0.0011

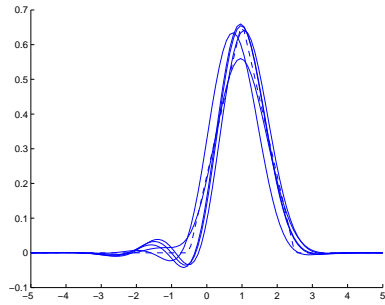


FIG 7. $(\Delta(-.5, 2.5), n = 100)$.

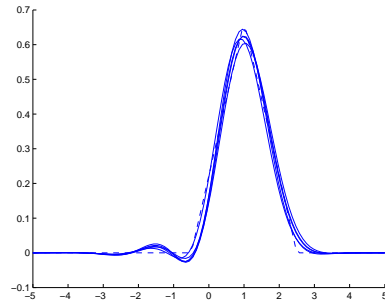


FIG 8. $(\Delta(-.5, 2.5), n = 1000)$.

TABLE 4
(shifted triangle density)

	$n = 100$	$n = 1000$
average	0.0175	0.0048
maximum	0.0809	0.0148
minimum	0.0023	0.0023

($n = 100$) while in the Figures 2, 4, 6, 8 the simulations are based on $n = 1000$ data. The error density f_ε is $N(0, 1)$, $\sigma = 0.5$ and $m = 1$. We have run 100 independent replicates in each setting. The figures show five arbitrarily chosen representatives of those replicates along with the true density, which is plotted as a dotted curve. Furthermore, we have computed the integrated squared errors (ISE) of the estimates; the average, the maximum and the minimum of the ISE taken over these 100 simulations are provided in the tabulars below the corresponding figures.

We choose the parameters $b = 0.08$, $K = 6$, $\rho = 0.2$ and $\bar{K} = 6$ in any simulation. Further numerical inspection indicates that the specific parameter selection of \bar{K} and ρ does not affect the outcome very much. Indeed in our examples the protection of the denominator by the threshold parameter ρ was redundant, i.e. putting $\rho = 0$ leads to the same result. The selection of the parameters b and K is more critical. However, the simulations indicate that the current choice seems reasonable for realistic sample sizes. In particular, one should avoid to select K too large in order to avoid an explosion of the variance, which is also supported by the theoretical aspect of the logarithmic convergence rates. Also, a slight modification of the estimator (2.9) was carried out. The function φ_K is appropriate for the theoretical purposes of this work; however, its derivatives show heavy oscillations near the support boundaries. Thus, the numerical integration of these functions is highly unstable. Therefore, we have replaced this function by a normal density, which is not compactly supported. Still, it performs well due to its fast decay in the tails. As a disadvantage of Hermite polynomials, the flexibility of the choice of the smoothing parameter is limited as only integers are admitted. Nevertheless, kernel techniques as studied in Wagner (2009) do not provide better simulation results than the current Hermite polynomial approach in the considered setting.

Of course, we realize improvement with respect to the variance when increasing the number n of the simulated data; in particular for the bimodal target densities where some structural outliers cannot be avoided for $n = 100$. Nevertheless, the results indicate that the basic structure of the target densities are well detected by our estimator; although we certainly face a very difficult estimation problem which does not admit estimation at algebraic convergence rates.

5. Proofs

Proof of Theorem 3.1. Applying the decomposition (2.7) to f_X , we obtain that

$$\begin{aligned} & E \|\hat{f}_X - f_X\|_2^2 \\ &= E \int \left| \sum_{k=0}^{\infty} \frac{1}{\sqrt{2\pi k!}} \left(\mathbf{e}_k^T \mathbf{W}^{-1} \hat{\mathbf{d}} \cdot \chi_{[0, K]}(k) \right. \right. \\ &\quad \left. \left. - \int f_X(y) H_k(y) dy \right) H_k(x) \right|^2 \exp(-x^2) dx \end{aligned}$$

$$\leq \sum_{k=0}^K \frac{1}{\sqrt{2\pi k!}} E |\mathbf{e}_k^T \mathbf{W}^{-1} [\hat{\mathbf{d}} - \mathbf{d}]|^2 + \sum_{k=K+1}^{\infty} \frac{1}{\sqrt{2\pi k!}} \left| \int f_X(y) H_k(y) dy \right|^2, \quad (5.1)$$

where we have used the orthogonality of the Hermite polynomials with respect to the inner product involving the weight function $\exp(-x^2/2)$ and the inequality $\exp(-x^2) \leq \exp(-x^2/2)$ for all $x \in \mathbb{R}$.

The smoothness conditions contained in $f_X \in \mathcal{F}$ must be exploited in order to bound the last term in (5.1). Note that the Hermite polynomials based on the weight $\exp(-x^2/2)$ satisfy the relation $H'_k(x) = kH_{k-1}(x)$. It follows by integration by parts that

$$\int f'_X(x) H_k(x) dx = - \int f_X(x) H'_k(x) dx = -k \int f_X(x) H_{k-1}(x) dx,$$

and, hence, by induction that

$$\int f_X^{(j)}(x) H_k(x) dx = (-1)^j \frac{k!}{(k-j)!} \int f_X(x) H_{k-j}(x) dx,$$

for all $j = 0, \dots, \beta$. Since $f_X \in \mathcal{F}$, the functions $g_j(x) = f_X^{(j)}(x) \exp(x^2/2)$ are contained in $L_2(\mathbb{R}, \exp(-x^2/2))$ and the $L_2(\mathbb{R}, \exp(-x^2/2))$ -norm of g_j is uniformly bounded with respect to $f_X \in \mathcal{F}$. By Parseval's identity with respect to the Hermite polynomial basis and by assumption (2.1), we derive that

$$\begin{aligned} C_3 &:= \sup_{f_X \in \mathcal{F}} \sum_{k=0}^{\infty} \frac{1}{\sqrt{2\pi k!}} \left| \int g_j(x) H_k(x) \exp(-x^2/2) dx \right|^2 \\ &= \sup_{f_X \in \mathcal{F}} \sum_{k=0}^{\infty} \frac{1}{\sqrt{2\pi k!}} \left| \int f_X^{(j)}(x) H_k(x) dx \right|^2 < \infty. \end{aligned}$$

Setting $j = \beta$ we find for $K \geq 4(\beta - 1)$ and any $f_X \in \mathcal{F}$

$$\begin{aligned} &\sum_{k=K+1}^{\infty} \frac{1}{\sqrt{2\pi k!}} \left| \int f_X(y) H_k(y) dy \right|^2 \\ &= \sum_{k=K+1+\beta}^{\infty} \frac{1}{\sqrt{2\pi(k-\beta)!}} \left| \int f_X(x) H_{k-\beta}(x) dx \right|^2 \\ &= \sum_{k=K+1-\beta}^{\infty} \frac{(k-\beta)!}{\sqrt{2\pi k!^2}} \left| \int f_X^{(\beta)}(x) H_k(x) dx \right|^2 \\ &\leq C_3 2^\beta (K^{-\beta}). \end{aligned} \quad (5.2)$$

Thus, we have derived an upper bound on the bias term.

Now we focus on the first term in (5.1). As φ_K is compactly supported and $(K+1)$ -fold continuously differentiable, its derivatives up to order $K+1$ and,

hence, their Fourier transforms $\cdot^l \varphi_K^{ft}$ lie in $L_2(\mathbb{R})$ for any $l = 0, \dots, K + 1$; and we have

$$\varphi_K^{(l)}(t) = \frac{1}{2\pi} \int \exp(-itx) (-ix)^l \varphi_K^{ft}(x) dx,$$

for almost all $t \in \mathbb{R}$ by simple Fourier inversion so that

$$\hat{D}_j = \frac{1}{2\pi} \iint \exp(-itx) x^j \varphi_K^{ft}(bx) dx \hat{\Psi}_X(t) dt.$$

Representing the functions \cdot^l , $l = 0, \dots, K$ in the orthogonal basis formed by the Hermite polynomials leads to

$$\mathbf{W}^{-1}(x^0, \dots, x^K)^T = (H_0(x), \dots, H_K(x))^T, \quad \forall x \in \mathbb{R},$$

from what follows that

$$\begin{aligned} & \mathbf{e}_k^T \mathbf{W}^{-1} [\hat{\mathbf{d}} - \mathbf{d}] \\ &= \frac{1}{2\pi} \iint \exp(-itx) H_k(x) \varphi_K^{ft}(bx) dx \hat{\Psi}_X(t) dt - \int f_X(x) H_k(x) dx \\ &= \frac{1}{2\pi} \iint \exp(-itx) H_k(x) \varphi_K^{ft}(bx) dx [\hat{\Psi}_X(t) - f_X^{ft}(t)] dt \\ & \quad + \int H_k(x) [\varphi_K^{ft}(bx) - 1] f_X(x) dx, \end{aligned}$$

where the Plancherel isometry has been used yielding

$$\int h(t) \overline{g(t)} dt = \frac{1}{2\pi} \int h^{ft}(t) \overline{g^{ft}(t)} dt,$$

for all $g, h \in L_2(\mathbb{R})$. In the sequel, we write const. for a generic constant which depends on neither f_X nor f_ε . As $\int \exp(-i \cdot x) H_k(x) \varphi_K^{ft}(bx) dx$ is a linear combination of finitely many derivatives of the function $\varphi_K(\cdot/b)$ which is supported on $[-b, b]$, we may conclude by the Cauchy-Schwarz inequality and Parseval's identity that

$$\begin{aligned} E |\mathbf{e}_k^T \mathbf{W}^{-1} [\hat{\mathbf{d}} - \mathbf{d}]|^2 &\leq \frac{1}{\pi} \left(\int_{-b}^b E |\hat{\Psi}_X(t) - f_X^{ft}(t)|^2 dt \right) \int H_k^2(x) |\varphi_K^{ft}(bx)|^2 dx \\ & \quad + \frac{1}{2} b^4 \cdot \left(\int |H_k(x)| x^2 f_X(x) dx \right)^2 \\ &\leq \text{const.} \cdot \left(\int_{-b}^b E |\hat{\Psi}_X(t) - f_X^{ft}(t)|^2 dt \int H_k^2(x) |\varphi_K^{ft}(bx)|^2 dx + k! b^4 \right), \quad (5.3) \end{aligned}$$

where we have used a second order Taylor approximation of φ_K^{ft} taking into account that $\varphi_K^{ft}(0) = 1$ as φ_K is a density, $[\varphi_K^{ft}]'(0) = 0$ by the symmetry of φ_K and

$$\|[\varphi_K^{ft}]''\|_\infty \leq \int x^2 \varphi_K(x) dx \leq 1.$$

yielding $|\varphi_K^{ft}(bx) - 1| \leq b^2 x^2$ and then

$$\begin{aligned} \left(\int |H_k(x)| x^2 f_X(x) dx \right)^2 &\leq \int H_k^2(x) \exp(-x^2/2) dx \int x^4 \exp(x^2/2) f_X^2(x) dx \\ &\leq \int H_k^2(x) \exp(-x^2/2) dx \int x^4 \exp(-x^2(4C_2 - 1)/2) dx \leq \text{const.} \cdot k!. \end{aligned}$$

The recursion relation for Hermite polynomials,

$$H_{k+1}(x) = xH_k(x) - kH_{k-1}(x), \quad \text{for all } x \in \mathbb{R},$$

implies that

$$|H_k(x)| \leq \begin{cases} k! \exp(1 + |x|), & \text{for } |x| \leq k - 1, \\ (1 + |x|)^k, & \text{otherwise,} \end{cases}$$

by induction and an estimate by the exponential series. These upper bounds yield that

$$\begin{aligned} &\int H_k^2(x) |\varphi_K^{ft}(bx)|^2 dx \\ &\leq \text{const.} \cdot \left((k!)^2 \exp(\text{const.} \cdot k) + \int x^{2k} |\varphi_K^{ft}(x)|^2 dx b^{-2k-1} \right) \\ &\leq \text{const.} \cdot \left(\exp(2k[\log k + \text{const.}]) + \int |\varphi_K^{(k)}(x)|^2 dx b^{-2k-1} \right) \\ &\leq \text{const.} \cdot \left(\exp(2k[\log k + \text{const.}]) + \text{const.}^K (K!)^2 \cdot b^{-2k-1} \right), \end{aligned}$$

by the Fourier representation of the Sobolev norm. Moreover, we have used that

$$|\varphi_K^{(k)}(x)| \leq \text{const.}^K \sum_{j=\lceil k/2 \rceil}^{K+2} \binom{K+2}{j} \binom{2j}{k} k! \leq \text{const.}^K K!,$$

for any $k = 0, \dots, K$ and $|x| \leq 1$. Inserting that inequality along with (5.2) into (5.3) and then into (2.7), we obtain that

$$\begin{aligned} &E \|\hat{f}_X - f_X\|_2^2 \\ &\leq \text{const.} \cdot b^{-1} \exp(2K[\log K - \log b + \text{const.}]) \int_{-b}^b E |\hat{\Psi}_X(t) - f_X^{ft}(t)|^2 dt \\ &\quad + \text{const.} \cdot Kb^4 + O(K^{-\beta}), \end{aligned} \quad (5.4)$$

Focussing on the term $E |\hat{\Psi}_X(t) - f_X^{ft}(t)|^2$, we write Ψ and $\tilde{\Psi}$ for the expectation of $\hat{\Psi}$ and $\hat{\tilde{\Psi}}$, respectively. Moreover, we introduce the notation

$$\Phi_D(t) = \prod_{k=0}^{\bar{K}} [\Psi(\sigma^k t)]^{m^k}, \quad \tilde{\Phi}_D(t) = \prod_{k=0}^{\bar{K}} [\tilde{\Psi}(\sigma^k t)]^{m^k}.$$

We find that

$$\begin{aligned} & E|\hat{\Psi}_X(t) - f_X^{ft}(t)|^2 \\ & \leq \text{const.} \left(E \left| \hat{\Psi}_X(t) - \frac{\tilde{\Phi}_D(t)\overline{\Phi_D(t)}}{|\Phi_D(t)|^2 \vee \rho} \right|^2 + \left| \frac{\tilde{\Phi}_D(t)\overline{\Phi_D(t)}}{|\Phi_D(t)|^2 \vee \rho} - f_X^{ft}(t) \right|^2 \right). \end{aligned} \quad (5.5)$$

For the first term we have

$$\begin{aligned} & E \left| \frac{\hat{\Phi}(t)(|\Phi_D(t)|^2 \vee \rho)\overline{\hat{\Phi}(t)} - \tilde{\Phi}_D(t)(|\hat{\Phi}(t)|^2 \vee \rho)\overline{\Phi_D(t)}}{(|\hat{\Phi}(t)|^2 \vee \rho)(|\Phi_D(t)|^2 \vee \rho)} \right|^2 \\ & \leq \rho^{-4} \cdot \text{const.} \cdot E \left(|\hat{\Phi}(t) - \tilde{\Phi}_D(t)|^2 + (|\hat{\Phi}(t)|^2 \vee \rho) \right. \\ & \quad \left. - (|\Phi_D(t)|^2 \vee \rho) \right)^2 + |\hat{\Phi}(t) - \Phi_D(t)|^2. \end{aligned}$$

As $\|\hat{\Phi}\|_\infty, \|\Phi_D\|_\infty \leq 1$ a.s., we have

$$|(|\hat{\Phi}(t)|^2 \vee \rho) - (|\Phi_D(t)|^2 \vee \rho)|^2 \leq 4|\hat{\Phi}(t) - \Phi_D(t)|^2.$$

Hence, it suffices to bound $E|\hat{\Phi}(t) - \tilde{\Phi}_D(t)|^2$ and $E|\hat{\Phi}(t) - \Phi_D(t)|^2$. We restrict our consideration to the first term as the second one can be treated analogously,

$$\begin{aligned} E|\hat{\Phi}(t) - \tilde{\Phi}_D(t)|^2 &= E \left| \prod_{k=0}^{\bar{K}} [\hat{\Psi}(\sigma^k t)]^{m^k} - \prod_{k=0}^{\bar{K}} [\tilde{\Psi}(\sigma^k t)]^{m^k} \right|^2 \\ &\leq E \left(\sum_{k=0}^{\bar{K}} |[\hat{\Psi}(\sigma^k t)]^{m^k} - [\tilde{\Psi}(\sigma^k t)]^{m^k}| \right)^2 \leq \bar{K} \sum_{k=0}^{\bar{K}} E |[\hat{\Psi}(\sigma^k t)]^{m^k} - [\tilde{\Psi}(\sigma^k t)]^{m^k}|^2 \\ &\leq \bar{K} m^{2\bar{K}} \sum_{k=0}^{\bar{K}} E |\hat{\Psi}(\sigma^k t) - \tilde{\Psi}(\sigma^k t)|^2 = O(\bar{K}^2 m^{2\bar{K}} / n), \end{aligned}$$

where we have applied the Cauchy-Schwarz inequality along with the inequality $a^p - b^p = (a - b) \sum_{j=0}^{p-1} a^j b^{p-1-j}$ and $\|\hat{\Psi}\|_\infty, \|\tilde{\Psi}\|_\infty \leq 1$ a.s.. For the second deterministic term in (5.5) we consider that, for $|t| \leq b$,

$$\begin{aligned} |\Phi_D(t)| &= \left| \prod_{k=0}^{\bar{K}} [\Psi(\sigma^{k+1} t)]^{m^{k+1}} \right| = \exp \left(\sum_{k=0}^{\bar{K}} m^{k+1} \log |\Psi(\sigma^{k+1} t)| \right) \\ &= \exp \left(\sum_{k=0}^{\bar{K}} m^{k+1} \frac{\log |\Psi(\sigma^{k+1} t)|}{|\Psi(\sigma^{k+1} t)| - 1} (|\Psi(\sigma^{k+1} t)| - 1) \right) \\ &\geq \exp \left(-\text{const.} \sum_{k=0}^{\bar{K}} m^{k+1} |\Psi(\sigma^{k+1} t) - 1| \right) \geq \exp \left(-\text{const.} \cdot b\bar{K} \right) \rightarrow 1 \end{aligned}$$

as $b \rightarrow 0$ when using (2.2) combined with the assumption $f_\varepsilon \in \mathcal{G}$ and the condition $\sigma \leq 1/m$. Therein, note that $\Psi(t) = f_X^{ft}(t) f_\varepsilon^{ft}(t)$. As an important

result, the convergence of $\Phi_D(t)$ to one takes place uniformly with respect to $|t| \leq b$, $f_X \in \mathcal{F}$ and $f_\varepsilon \in \mathcal{G}$.

Thus, for n sufficiently large, we have uniformly in $|t| \leq b$ that

$$\frac{\tilde{\Phi}_D(t)\overline{\Phi_D(t)}}{|\Phi_D(t)|^2 \vee \rho} = \frac{\tilde{\Phi}_D(t)}{\Phi_D(t)} = \frac{\tilde{\Phi}(t)}{\Phi(t)} = \frac{f_X^{ft}(t)}{[f_X^{ft}(\sigma^{\bar{K}+1}t)]^{m^{\bar{K}+1}}}.$$

Applying this formula to evaluate the second term in (5.5) we find that

$$\begin{aligned} \left| \frac{\tilde{\Phi}_D(t)\overline{\Phi_D(t)}}{|\Phi_D(t)|^2 \vee \rho} - f_X^{ft}(t) \right|^2 &= \frac{|f_X^{ft}(t)|^2}{|f_X^{ft}(\sigma^{\bar{K}+1}t)|^{2m^{\bar{K}+1}}} |1 - [f_X^{ft}(\sigma^{\bar{K}+1}t)]^{m^{\bar{K}+1}}|^2 \\ &\leq \text{const.} \cdot m^{2\bar{K}} |1 - f_X^{ft}(\sigma^{\bar{K}+1}t)|^2 = O(b^4 \sigma^{2(\bar{K}+1)}), \end{aligned}$$

where, again, we have used (2.2) and the inequality $\sigma \leq 1/m$. Again, $O(\dots)$ does not depend on f_X or f_ε . Summarizing, we conclude that

$$E|\hat{\Psi}_X(t) - f_X^{ft}(t)|^2 = O(\bar{K}^2 m^{2\bar{K}} n^{-1} + b^4 \sigma^{2\bar{K}}),$$

uniformly over all $|t| \leq b$. Applying that result to (5.4), we obtain that

$$\begin{aligned} E\|\hat{f}_X - f_X\|_2^2 \\ \leq \text{const.} \cdot \left\{ \exp(2K[\log K - \log b]) (\bar{K}^2 m^{2\bar{K}} n^{-1} + b^4 \sigma^{2\bar{K}}) + Kb^4 + K^{-\beta} \right\}. \end{aligned}$$

Inserting the parameter choice as requested in the theorem, we realize that all terms except the last one are asymptotically negligible. Thus, we have a bias-dominated problem; the order of the last term determines the convergence rate stated in the theorem. \square

Proof of Theorem 3.2. As two density sequences competing to be the true density of the X_j , we define

$$f_{X,d}(x) = [f_N * V](x) + a_n d \cos((K_n + 1/2)\pi x) [f_N * V](x),$$

where f_N denotes the standard normal density; V denotes the uniform distribution on the set $\{-1, 1\}$; $(a_n)_n \downarrow 0$; $(K_n)_n \uparrow \infty$ is integer-valued; and $d = d_0, d_1$ with two constants $d_1 > d_0 > 0$. The specific choice of $(a_n)_n$ and $(K_n)_n$ is deferred. We derive that

$$\begin{aligned} f_{X,d}^{ft}(t) &= f_N^{ft}(t) \cos(t) + \frac{a_n}{2} d (-1)^{K_n} \sin(t) [f_N^{ft}(t - (K_n + 1/2)\pi) \\ &\quad - f_N^{ft}(t + (K_n + 1/2)\pi)]. \end{aligned}$$

Under the selection $a_n = c_a K_n^{-\beta}$ with an appropriate constant $c_a > 0$, we can verify that $f_{X,d} \in \mathcal{F}$ for n sufficiently large. Also we are guaranteed that $f_{X,d}$, $d = d_0, d_1$, are density function as their nonnegativity and their membership in $L_1(\mathbb{R})$ follow from the definition of $f_{X,d}$ (note that $1 + a_n d \cos((K_n + 1/2)\pi x) \geq 0$

for all x and n large enough). Furthermore, these functions integrate to one as $f_{X,d}^{ft}(0) = 1$.

With respect to the error density, we will also specify the two competing density sequences $f_{\varepsilon,d}$, $d = d_0, d_1$. We define $f_{\varepsilon,d}$ via its Fourier transform; we put

$$f_{\varepsilon,d_1}^{ft}(t) = d_1(|t| + d_1^{1/\alpha'})^{-\alpha'}.$$

Therein, we choose $\alpha' \in (1/2, \alpha)$. The Fourier transform of f_{ε,d_0} coincides with f_{ε,d_1}^{ft} on the interval $[-c_{t,0}K_n, c_{t,0}K_n]$ and is equal to the function $d_0(|\cdot| + d_1^{1/\alpha'})^{-\alpha'}$ on $\mathbb{R} \setminus [-c_{t,1}K_n, c_{t,1}K_n]$. By appropriate selection of the terms $c_{t,1} > c_{t,0} > 0$, the function f_{ε,d_0}^{ft} may be continued onto $[-c_{t,1}K_n, -c_{t,0}K_n]$ and $[c_{t,0}K_n, c_{t,1}K_n]$ by the tangent lines of f_{ε,d_1}^{ft} taken at the points $-c_{t,0}K_n$ and $c_{t,0}K_n$, respectively, so that f_{ε,d_0}^{ft} is continuous on the whole real line. Therein, we allow for $c_{t,0}$ and $c_{t,1}$ to depend on n but $c_{t,0}$ is bounded away from zero and $c_{t,1}$ is bounded from above by a positive number smaller than σ . Those conditions may be fixed by choosing $c_{t,0}$ sufficiently small since those tangent lines take the value 0 at $t_n = \pm(1 + 1/\alpha')c_{t,0}K_n \pm d_1^{1/\alpha'}/\alpha'$; and we have $c_{t,1}K_n \leq |t_n|$. We realize that the competing error densities have the following properties: f_{ε,d_0}^{ft} and f_{ε,d_1}^{ft} are symmetric; $f_{\varepsilon,d_0}^{ft}(0) = f_{\varepsilon,d_1}^{ft}(0) = 1$; they are continuously differentiable on $(0, \infty)$ with finite right side derivatives at 0; continuous on the whole of \mathbb{R} ; monotonously decreasing and convex on $[0, \infty]$; $\lim_{t \rightarrow +\infty} f_{\varepsilon,d_0}^{ft}(t) = \lim_{t \rightarrow +\infty} f_{\varepsilon,d_0}^{ft}(t) = 0$. Hence, we learn from Polya's criterion (see e.g. [Lukacs \(1970\)](#), p.83, Theorem 4.3.1) that $f_{\varepsilon,d}$, $d = d_0, d_1$, are both probability densities and contained in $L_2(\mathbb{R})$ where the latter properties follows from the integrability of $|f_{\varepsilon,d}^{ft}|^2$, $d = d_0, d_1$, by Parseval's identity. One can easily recognize that the polynomial lower bound on $f_{\varepsilon,d}^{ft}$ is satisfied; as is the Lipschitz condition on $f_{\varepsilon,d}^{ft}$ contained in the definition of \mathcal{G} . Therefore, we have verified that $f_{\varepsilon,d} \in \mathcal{G}'$ for $d = d_0, d_1$.

After constructing those densities, we write \hat{f}_X for an arbitrary estimator of f_X based on the given observation scheme. We use the following arguments from statistical decision theory.

$$\begin{aligned} & \sup_{f_X \in \mathcal{F}, f_\varepsilon \in \mathcal{G}'} E \|\hat{f}_X - f_X\|_2^2 \\ & \geq \frac{1}{2} \left(\int \|\hat{f}_X(\mathbf{y}) - f_{X,d_0}(\mathbf{y})\|_2^2 h_{d_0,d_1}(\mathbf{y}) d\mathbf{y} + \int \|\hat{f}_X(\mathbf{y}) - f_{X,d_1}(\mathbf{y})\|_2^2 h_{d_1,d_0}(\mathbf{y}) d\mathbf{y} \right) \\ & \geq \frac{1}{4} \|f_{X,d_0} - f_{X,d_1}\|_2^2 \int \min \{h_{d_0,d_1}(\mathbf{y}), h_{d_1,d_0}(\mathbf{y})\} d\mathbf{y} \\ & = \frac{1}{4} \|f_{X,d_0} - f_{X,d_1}\|_2^2 \left(1 - \frac{1}{2} \|h_{d_0,d_1} - h_{d_1,d_0}\|_1 \right), \end{aligned} \tag{5.6}$$

where $h_{d_\theta, d_{1-\theta}}$, $\theta \in \{0, 1\}$, denotes the densities of the data set $(Y_1, Y'_1, \dots, Y_n, Y'_n)$ when f_{X,d_θ} and $f_{\varepsilon, d_{1-\theta}}$ are the true densities of X_1 and ε_1 , respectively. By a

telescopic sum argument, we derive that

$$\begin{aligned} \|h_{d_0, d_1} - h_{d_1, d_0}\|_1 &\leq n(\|f_{X, d_0} * f_{\varepsilon, d_1} - f_{X, d_1} * f_{\varepsilon, d_0}\|_1 \\ &\quad + \|f_{X, d_0} * \sigma^{-1} f_{\varepsilon, d_1}(\sigma^{-1} \cdot) - f_{X, d_1} * \sigma^{-1} f_{\varepsilon, d_0}(\sigma^{-1} \cdot)\|_1), \end{aligned} \quad (5.7)$$

Now we consider the $L_1(\mathbb{R})$ -distance

$$\begin{aligned} &\|f_{X, d_0} * \sigma^{-1} f_{\varepsilon, d_1}(\sigma^{-1} \cdot) - f_{X, d_1} * \sigma^{-1} f_{\varepsilon, d_0}(\sigma^{-1} \cdot)\|_1 \\ &\leq \text{const.} \cdot \left(\int (1+x^2) | [f_{X, d_0} * \sigma^{-1} f_{\varepsilon, d_1}(\sigma^{-1} \cdot)](x) \right. \\ &\quad \left. - [f_{X, d_1} * \sigma^{-1} f_{\varepsilon, d_0}(\sigma^{-1} \cdot)](x) |^2 dx \right)^{1/2} \\ &\leq \text{const.} \cdot \left(\int |f_{X, d_0}^{ft}(t) f_{\varepsilon, d_1}^{ft}(\sigma t) - f_{X, d_1}^{ft}(t) f_{\varepsilon, d_0}^{ft}(\sigma t)|^2 dt \right. \\ &\quad + \int |(f_{X, d_0}^{ft})'(t) f_{\varepsilon, d_1}^{ft}(\sigma t) - (f_{X, d_1}^{ft})'(t) f_{\varepsilon, d_0}^{ft}(\sigma t)|^2 dt \\ &\quad \left. + \sigma^2 \int |f_{X, d_0}^{ft}(t) (f_{\varepsilon, d_1}^{ft})'(\sigma t) - f_{X, d_1}^{ft}(t) (f_{\varepsilon, d_0}^{ft})'(\sigma t)|^2 dt \right)^{1/2}, \end{aligned} \quad (5.8)$$

by the Cauchy-Schwarz inequality and Parseval's identity. The analogous upper bound holds true for the first $L_1(\mathbb{R})$ -distance occurring on the right side of (5.7) when setting $\sigma = 1$. Note that the derivatives of $f_{\varepsilon, d_0}^{ft}$ and $f_{\varepsilon, d_1}^{ft}$ are to be understood in the weak Sobolev sense. For $|t| \geq \sigma^{-1} c_{t,1} K_n$, we derive that

$$\begin{aligned} [f_{\varepsilon, d_0}^{ft}]^{(l)}(\sigma^k t) &= (d_0/d_1) [f_{\varepsilon, d_1}^{ft}]^{(l)}(\sigma^k t), \\ [f_{X, d_1}^{ft}]^{(l')}(t) &= (d_1/d_0) [f_{X, d_0}^{ft}]^{(l')}(t) + [1 - d_1/d_0] \{ [f_N^{ft}]^{(l')}(t) \cos(t) \\ &\quad - \chi_{\{1\}}(l') f_N^{ft}(t) \sin(t) \}, \end{aligned}$$

for $k, l, l' = 0, 1$. Thus, we obtain that

$$\begin{aligned} &\int_{|t| \geq \sigma^{-1} c_{t,1} K_n} |[f_{\varepsilon, d_0}^{ft}]^{(l)}(\sigma^k t) [f_{X, d_1}^{ft}]^{(l')}(t) - [f_{\varepsilon, d_1}^{ft}]^{(l)}(\sigma^k t) [f_{X, d_0}^{ft}]^{(l')}(t)|^2 dt \\ &\leq \text{const.} \cdot \int_{|t| \geq \sigma^{-1} c_{t,1} K_n} |[f_{\varepsilon, d_1}^{ft}]^{(l)}(\sigma^k t)|^2 [|[f_N^{ft}]'(t)| + |f_N^{ft}(t)|]^2 dt. \end{aligned} \quad (5.9)$$

From the definition of $f_{\varepsilon, d_1}^{ft}$, we easily derive that its first weak derivative is bounded (with a jump discontinuity at $t = 0$) while the function itself is bounded by 1. Due to the specific exponential tails of $f_N^{ft}(t)$ and its derivative, we obtain that (5.9) has the upper bound $O(K_n^3 \exp(-\text{const.} K_n^2))$. This bound affects all partial integrals taken over $|t| \geq \sigma^{-1} c_{t,1} K_n$ in (5.8).

Now we consider those partial integrals taken over $|t| \leq c_{t,0} K_n$. Note that, on this domain, we have coincidence of $f_{\varepsilon, d_0}^{ft}$ and $f_{\varepsilon, d_1}^{ft}$ on the one hand and $f_{\varepsilon, d_0}^{ft}(\sigma \cdot)$

and $f_{\varepsilon, d_1}^{ft}(\sigma)$ on the other hand and, hence, their weak derivatives coincides on that interval, too. As $f_{\varepsilon, d}^{ft}$, $d = d_0, d_1$ and their weak derivatives are bounded, those partial integrals have the upper bound

$$\begin{aligned} & \text{const.} \cdot \max_{l=0,1} \int_{|t| \leq c_{t,0} K_n} |[f_{X, d_0}^{ft}]^{(l)}(t) - [f_{X, d_1}^{ft}]^{(l)}(t)|^2 dt \\ & = O(a_n^2) \int_{|t| \leq c_{t,0} K_n} (|[f_N^{ft}]'(t - (K_n + 1/2)\pi)|^2 + |f_N^{ft}(t - (K_n + 1/2)\pi)|^2) dt. \end{aligned}$$

Again, we obtain the upper bound $O(K_n^3 \exp(-\text{const.} K_n^2))$.

The partial integrals in (5.8) taken over $\sigma^{-1} c_{t,1} K_n \geq |t| \geq c_{t,0} K_n$ remain to be studied. These integrals are bounded from above by

$$\text{const.} \cdot \max_{d=d_0, d_1} \max_{l=0,1} \int_{\sigma^{-1} c_{t,1} K_n \geq |t| \geq c_{t,0} K_n} |[f_{X, d}^{ft}]^{(l)}(t)|^2 dt.$$

As $c_{t,1} < \sigma$ and $c_{t,0}$ is bounded away from zero the explicit expression for $f_{X, d}^{ft}$ and its derivative provides the upper bound $O(K_n^3 \exp(-\text{const.} K_n^2))$ in this case, too.

Summarizing, (5.8) is bounded above by $O(K_n^{3/2} \exp(-\text{const.} K_n^2))$. Selecting $K_n = C_K (\log n)^{1/2}$ with a constant $C_K > 0$ sufficiently large, the right side of (5.7) converges to zero. Then, we conclude from (5.6) and Parseval's identity that

$$\begin{aligned} & \sup_{f_X \in \mathcal{F}, f_\varepsilon \in \mathcal{G}'} E \|\hat{f}_X - f_X\|_2^2 \geq \text{const.} \cdot \|f_{X, d_0}^{ft} - f_{X, d_1}^{ft}\|_2^2 \\ & \geq \text{const.} \cdot a_n^2 \int \cos^2(t) |f_N^{ft}(t)|^2 dt \asymp a_n^2 \asymp K_n^{-2\beta} \asymp (\log n)^{-\beta}, \end{aligned}$$

by using simple trigonometric relations. Thus, we have established the desired asymptotic lower bound on the risk. \square

References

- BELOMESTNY, D. (2003). Rates of convergence for constrained deconvolution problem. *Preprint* arXiv [math.ST/0306237 v1](#).
- BUTUCEA, C. AND MATIAS, C. (2005). Minimax estimation of the noise level and of the signal density in a semiparametric convolution model. *Bernoulli* **11**, 309–340. [MR2132729](#)
- BUTUCEA, C., MATIAS, C. AND POUET, C. (2008). Adaptivity issues in convolution models with known or partially known noise distribution. *Electr. J. Stat.* **2**, 897–915. [MR2447344](#)
- CARROLL, R.J. AND HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184–1186. [MR0997599](#)
- DELAIGLE, A., HALL, P. AND MEISTER, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.* **36**, 665–685. [MR2396811](#)

- DELAIGLE, A. AND MEISTER, A. (2008). Density estimation with heteroscedastic error. *Bernoulli* **14**, 562–579. [MR2544102](#)
- DIGGLE, P. AND HALL, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser. B* **55**, 523–531. [MR1224414](#)
- EFROMOVICH, S. (1997). Density estimation for the case of supersmooth measurement error. *J. Amer. Statist. Assoc.* **92**, 526–535. [MR1467846](#)
- FAN, J. (1991). On the optimal rates of convergence for non-parametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272. [MR1126324](#)
- HALL, P. AND MEISTER, A. (2007). A ridge-parameter approach to deconvolution. *Ann. Statist.* **35**, 1535–1558. [MR2351096](#)
- HALL, P. AND YAO, Q. (2003). Inference in components of variance models with low replication. *Ann. Statist.* **31**, 414–441. [MR1983536](#)
- HOROWITZ, J.L. AND MARKATOU, M. (1996). Semiparametric estimation of regression models for panel data. *Rev. Econom. Stud.* **63**, 145–168. [MR1372250](#)
- JOHANNES, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist.*, to appear. [MR2543693](#)
- LI, T. AND VUONG, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivar. Anal.* **65**, 139–165. [MR1625869](#)
- LINTON, O. AND WHANG, Y.L. (2002). Nonparametric estimation with aggregated data. *Econometric Theo.* **18**, 420–468. [MR1891830](#)
- LUKACS, E. *Characteristic Functions*. 2nd ed., 1970, Griffin, London. [MR0346874](#)
- MEISTER, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Canad. J. Statist.* **32**, 439–449. [MR2125855](#)
- MEISTER, A. (2006). Density estimation with normal measurement error with unknown variance. *Statist. Sinica* **16**, 195–211. [MR2256087](#)
- MEISTER, A. (2007). Deconvolving compactly supported densities. *Math. Meth. Stat.* **16**, 63–76. [MR2319471](#)
- MEISTER, A. (2008). Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Problems* **24**, 015003 (14 pages). [MR2384762](#)
- MEISTER, A. *Deconvolution problems in nonparametric statistics*. Lecture Notes in Statistics **193**, 2009, Springer, Heidelberg.
- MORRIS, J.N., MARR, J.W. AND CLAYTON, D.G. (1977). Diet and heart: a postscript. *British Med. J.* **2**, 1307–1314.
- NELSON, J.A. (1994). Consumer Expenditure Surveys 1980-1989: Interview Surveys, for Household-Level Analysis, Computer file, United States Department of Labor, Bureau of Labor Statistics, Washington, DC. Distributed by: Inter-University Consortium for Political and Social Research, Ann Arbor, MI.
- NEUMANN, M.H. (1997). On the effect of estimating the error density in non-parametric deconvolution. *J. Nonparam. Stat.* **7**, 307–330. [MR1460203](#)
- NEUMANN, M.H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivar. Anal.* **98**, 1955–1968. [MR2396948](#)

- SCHENNACH, S.M. (2004a). Estimation of nonlinear models with measurement error. *Econometrica* **72**, 33–75. [MR2031013](#)
- SCHENNACH, S.M. (2004b). Nonparametric regression in the presence of measurement error. *Econometric Theory* **20**, 1046–1093. [MR2101951](#)
- STAUDENMAYER, J., RUPPERT, D. AND BUONACCORSI, J. (2008). Density estimation in the presence of heteroskedastic measurement error. *J. Amer. Statist. Assoc.* **103**, 726–736.
- STEFANSKI, L.A. AND CARROLL, R.J. (1990). Deconvoluting kernel density estimators. *Statistics* **21**, 169–184. [MR1054861](#)
- THAMERUS, M. (1996). Fitting a finite mixture distribution to a variable subject to heteroscedastic measurement error. *Preprint*, <http://epub.ub.uni-muenchen.de/>
- WAGNER, C. (2009). *Deconvolution problems in density estimation*. Dissertation, University of Ulm, Germany.