Research Article

Self-Consistent Density Estimation in the Presence of Errors-in-Variables

Junhua Zhang,[1] Yuping Hu,[2,3] and Sanying Feng[2,3]

[1]*College of Mechanical Engineering, Beijing Information Science and Technology University, Beijing 100192, China*
[2]*College of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, China*
[3]*College of Applied Sciences, Beijing University of Technology, Beijing 100124, China*

Correspondence should be addressed to Sanying Feng; fsy5801@emails.bjut.edu.cn

This paper considers the estimation of the common probability density of independent and identically distributed variables observed with additive measurement errors. The self-consistent estimator of the density function is constructed when the error distribution is known, and a modification of the self-consistent estimation is proposed when the error distribution is unknown. The consistency properties of the proposed estimators and the upper bounds of the mean square error and mean integrated square error are investigated under some suitable conditions. Simulation studies are carried out to assess the performance of our proposed method and compare with the usual deconvolution kernel method. Two real datasets are analyzed for further illustration.

1. Introduction

The statistical methodology of density estimation has been widely studied in the literature (see [1–5]) and still is a hot issue in nonparametric statistics. For example, Park et al. [6] studied the kernel-based local likelihood density estimation. Jones and Henderson [7] proposed a Gaussian copula kernel estimator. Botev et al. [8] introduced an adaptive kernel density estimation method based on the smoothing properties of linear diffusion processes. The most commonly used nonparametric method is kernel density estimation. Usually, the choice of kernel function is not crucial, whereas the bandwidth parameter, which controls the degree of smoothing, must be chosen carefully. However, it is well known that the choice of bandwidth is difficult. Cross-validation techniques have been previously applied for this, but they are computationally expensive. Furthermore, it becomes difficult to estimate the density and choose the bandwidth when there exists the measurement error.

In this paper, we consider the density estimation in the presence of additive measurement error. Suppose we have n i.i.d. available observations $Y_1, \ldots, Y_n$ which have the same distribution as that of Y, to estimate the unknown density $f(x)$ of a random variable X, where

$$Y = X + \varepsilon, \tag{1}$$

with a measurement error ε, and ε is independent of X. The density function of ε is denoted as f_ε.

When the distribution of ε is known, the statistical methodology for estimating the unknown density $f(x)$ has been extensively discussed in the literature. For instance, Carroll and Hall [2] and Fan [9, 10] discussed the optimal rates of convergence over a class of functions whose derivatives are Lipschitz continuous using kernel method. Zhang [11] studied the deconvolution kernel density estimation of the mixing densities and distributions and derived the optimal rates of convergence. See also Koo [12], Pensky and Vidakovic [13], Fan and Koo [14], and Comte et al. [15], among others, for earlier contributions. Recently, Butucea [4] and Butucea and Tsybakov [5] evaluated the minimax rate of convergence of the pointwise risk using the kernel method and computed upper bounds for the L^2 risk of the estimator. However, most of

those papers deal only with theoretical aspects of the estimation, and very few focus attention on the yet important issue of choosing the bandwidth in practice. Delaigle and Gijbels [16] proposed a bandwidth selection procedure based on bootstrap techniques and proved that the mean integrated squared error (MISE) of the bootstrap estimator for the density function converges to the exact mean integrated squared error. However, the algorithm is complex and the quality of the estimator of the density depends strongly on the choice of the pilot bandwidth which must be chosen before the bootstrap procedure.

Indeed, in most practical applications, the distribution of ε cannot be perfectly known. Meister [17] studied the effect of the misspecification of the error density on the MISE of the deconvolution estimator. He pointed out that the limit of MISE can be infinite in some cases. Sometimes, this problem can be solved by repeated observations of the same variable of interest, each time with an independent error; see, for example, Li and Vuong [18], Delaigle et al. [19], or Neumann [20] and references therein. However, there are also many application fields where it is not possible to do repeated measurements of the same variable. In that case, the information of the error distribution can be drawn from an additional experiment; this means that knowledge of f_ε can be replaced by observations $\varepsilon_{-1}, \ldots, \varepsilon_{-M}$, which is a noise sample with distribution f_ε, and is independent of $Y_1, \ldots, Y_n$. Neumann [21] and Kerkyacharian et al. [22] replaced the characteristic function of the error by its empirical version. Johannes [23] studied the density deconvolution with unknown (but observed) noise and showed that the proposed estimators are asymptotically optimal in a minimax sense over a Sobolev space. The resulting estimators depend on two bandwidth-type parameters, but the data-driven selection of these bandwidths was not done. Recently, Comte and Lacour [24] proposed an adaptive estimator of the density based on a data-driven model selection strategy and discussed the convergence rates of the estimator. In addition, they studied the link between M and n if one wants to preserve the rate that is found in the case where the distribution of ε is known.

In the present paper, we study model (1), completed with a known or unknown distribution of error ε, and construct the self-consistent estimators of the density $f(x)$ by searching for the optimal kernel when the error distribution is ordinary smooth or super smooth. The method of constructing the estimator of density function by optimal kernel was first proposed by Watson and Leadbetter [25]. But the result can not be used directly, since the Fourier transform of the true density function is unknown. Bernacchia and Pigolotti [26] derived the exact expression of the density estimator based on optimal kernel and defined it as self-consistent estimate. They showed that the self-consistent method has better performance than existing methods for all examples that they studied. But the data in their paper did not include measurement errors. The self-consistent method shares some desired features: the choice of the bandwidth-type parameter is more convenient than that of kernel or spline method; thus the computing speed can be improved significantly; the resulting estimators are consistent; the proposed method is preferable for applications, especially for the large datasets. Thus, it is

of great significance to extend the self-consistent estimate method to more general case.

The paper is organized as follows. In Section 2, we propose the self-consistent method for density estimation with the known distribution of ε, or with the noise sample $\varepsilon_{-1}, \ldots, \varepsilon_{-M}$ when f_ε is unknown, and give the asymptotic properties. In Section 3, some simulations are carried out to illustrate the efficacy of the proposed method. Two real data examples are used for illustration in Section 4. The proofs of the main results are included in Appendix.

2. Methodology and the Main Results

2.1. Error Distribution Is Known. For the sake of descriptive convenience, we first introduce some notations. For two real numbers a and b, we denote $a \wedge b = \min(a, b)$ and $a_+ = \max(a, 0)$. Let z be a complex number, let $\overline{z}$ denote its conjugate, and let $|z|$ denote its modulus. Let $\|g\|$ be the L^2-norm of g; that is, $\|g\| = \int_{\mathbb{R}} |g(x)|^2 dx$. The Fourier transform of g is defined by

$$\phi_g(t) = \int \exp(\mathrm{i}tx) g(x) dx. \tag{2}$$

Similarly, we denote the characteristic function as $\phi_\varepsilon(t)$ for the known density function f_ε of ε.

The smoothness of f_ε is described by the following assumption. There exist some positive constants d_0, d_1, β, and b and constants $\beta_0 \in \mathbb{R}$ ($\beta_0 > 0$ if $\beta = 0$), such that the characteristic function $\phi_\varepsilon(t)$ of the error distribution satisfies

$$d_0 \left(t^2 + 1\right)^{-\beta_0/2} \exp\left(-b |t|^\beta\right)$$
$$\leq |\phi_\varepsilon(t)| \leq d_1 \left(t^2 + 1\right)^{-\beta_0/2} \exp\left(-b |t|^\beta\right) \quad \text{as } t \longrightarrow \infty. \tag{3}$$

When $\beta > 0$ we call the distribution of ε as super smooth of order β. For example, standard normal and Cauchy distributions are super smooth with $\beta = 2$ and 1, respectively. When $\beta = 0$ we call the distribution of ε as ordinary smooth of order β_0. For example, gamma distribution $Ga(\alpha, \lambda)$ and double exponential distribution are ordinary smooth with $\beta_0 = \alpha$ and 2, respectively.

Assume that the density function $f(x)$ of X belongs to the following type of smoothness spaces:

$$\mathcal{F}_{r,a,\delta}(m)$$
$$= \left\{ f \text{ density}: \int |\phi_f(x)|^2 \left(x^2 + 1\right)^\delta \exp\left(2a |x|^r\right) dx \leq m \right\}, \tag{4}$$

where $m > 0$, $r \geq 0$, $a > 0$, $\delta \in \mathbb{R}$, and $\delta > 1/2$ if $r = 0$. If $r > 0$, the density function $f(x)$ is called super smooth and ordinary smooth otherwise.

Let f_Y denote the common density of the Y_j's. By (1) and the independence assumption between X and ε, we have

$$f_Y = f * f_\varepsilon, \qquad \phi_Y = \phi_f \phi_\varepsilon. \tag{5}$$

where $*$ denotes the convolution and ϕ_Y denotes the characteristic function of Y. Let

$$\widehat{f}_Y(y) = \frac{1}{n}\sum_{j=1}^{n} K(y - Y_j) \tag{6}$$

be a kernel estimator of f_Y with the kernel function $K(\cdot)$. Then, the Fourier transform of $\widehat{f}_Y$ is defined by $\widehat{\phi}_Y(t) = \phi_K(t)\Delta(t)$, where $\Delta(t) = (1/n)\sum_{j=1}^{n}\exp(itY_j)$ is the empirical characteristic function of Y. Therefore, under the classical assumption $\phi_\varepsilon(t) \neq 0$, for any $t \in \mathbb{R}$, we can obtain an estimator of ϕ_f as follows:

$$\widehat{\phi}_f(t) = \frac{\phi_K(t)\Delta(t)}{\phi_\varepsilon(t)}. \tag{7}$$

By the inverse Fourier transform, we have

$$\widehat{f}(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp(-itx)\frac{\phi_K(t)\Delta(t)}{\phi_\varepsilon(t)}dt. \tag{8}$$

Note that (8) does not depend on any bandwidth h, compared to the deconvolution kernel density estimation methods in the literature. Instead of looking for an arbitrary shape for the kernel $K(\cdot)$ and choosing an optimal bandwidth (see [1]), we rather look for an optimal shape of the kernel $K(\cdot)$ such that the estimator $\widehat{f}(x)$ minimizes the mean integrated square error

$$\text{MISE} = E\left(\int_{-\infty}^{+\infty}\left(\widehat{f}(x) - f(x)\right)^2 dx\right). \tag{9}$$

By means of Parseval's theorem and (7), we have

$$\begin{aligned}
\text{MISE} &= \frac{1}{2\pi}E\left(\int_{-\infty}^{+\infty}\left|\widehat{\phi}_f(t) - \phi_f(t)\right|^2 dt\right) \\
&= \frac{1}{2\pi}E\left(\int_{-\infty}^{+\infty}\left|\frac{\phi_K(t)\Delta(t)}{\phi_\varepsilon(t)} - \phi_f(t)\right|^2 dt\right) \\
&= \frac{1}{2\pi}E\left(\int_{-\infty}^{+\infty}\frac{\left|\widehat{\phi}_Y(t) - \phi_Y(t)\right|^2}{\left|\phi_\varepsilon(t)\right|^2}dt\right).
\end{aligned} \tag{10}$$

Note that $E(\Delta(t)) = \phi_Y(t)$ and $E(|\Delta(t)|^2) = |\phi_Y(t)|^2 + n^{-1}(1 - |\phi_Y(t)|^2)$, and the MISE can be rewritten as

$$\begin{aligned}
&\text{MISE} \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\left(n^{-1}\left|\phi_K(t)\right|^2\left(1 - \left|\phi_Y(t)\right|^2\right) + \left|\phi_Y(t)\right|^2\right.\right. \\
&\qquad\qquad \left.\left.\times\left|1 - \phi_K(t)\right|^2\right)\times\left(\left|\phi_\varepsilon(t)\right|^2\right)^{-1}\right)dt
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{1 + (n-1)\left|\phi_Y(t)\right|^2}{n\left|\phi_\varepsilon(t)\right|^2} \\
&\qquad\times\left|\phi_K(t) - \frac{n\left|\phi_Y(t)\right|^2}{1 + (n-1)\left|\phi_Y(t)\right|^2}\right|^2 dt \\
&\quad + \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{\left|\phi_Y(t)\right|^2\left(1 - \left|\phi_Y(t)\right|^2\right)}{\left|\phi_\varepsilon(t)\right|^2\left(1 + (n-1)\left|\phi_Y(t)\right|^2\right)}dt.
\end{aligned} \tag{11}$$

Then we get the optimal kernel, which in Fourier space reads

$$\phi_{K_{\text{opt}}}(t) = \frac{n}{n - 1 + \left|\phi_Y(t)\right|^{-2}}. \tag{12}$$

Replacing $\phi_K(t)$ in (7) by the optimal kernel (12), we have

$$\begin{aligned}
\widehat{\phi}_f(t) &= \frac{\phi_{K_{\text{opt}}}(t)\Delta(t)}{\phi_\varepsilon(t)} \\
&= \frac{n\Delta(t)}{\left(n - 1 + \left|\phi_Y(t)\right|^{-2}\right)\phi_\varepsilon(t)} \\
&= \frac{n\Delta(t)}{\left(n - 1 + \left|\phi_f(t)\phi_\varepsilon(t)\right|^{-2}\right)\phi_\varepsilon(t)}.
\end{aligned} \tag{13}$$

Similar to the discussion of Bernacchia and Pigolotti [26], the self-consistent estimator of ϕ_f is equivalent to solving the equation

$$\widehat{\phi}_{f_{\text{sc}}}(t) = \frac{n\Delta(t)}{\left(n - 1 + \left|\widehat{\phi}_{f_{\text{sc}}}(t)\phi_\varepsilon(t)\right|^{-2}\right)\phi_\varepsilon(t)}. \tag{14}$$

Simple calculation yields

$$\widehat{\phi}_{f_{\text{sc}}}(t) = \frac{n\Delta(t)}{2(n-1)\phi_\varepsilon(t)}\left(1 + \sqrt{1 - \frac{4(n-1)}{n^2|\Delta(t)|^2}}\right)I_A(t), \tag{15}$$

where $I_A(\cdot)$ is the indicator function ($I_A(t) = 1$ if $t \in A$; $I_A(t) = 0$ if $t \notin A$) and $A \subseteq B$, $B = \{t : |\Delta(t)|^2 \geq 4(n-1)/n^2\}$, is a set of accepted frequencies (i.e., the frequencies giving a nonzero contribution to the estimate). In practical applications, when the sample $Y_1, \ldots, Y_n$ has been observed, we can choose a bounded interval

$$A = B \cap \left[-\widetilde{t}, \widetilde{t}\right], \tag{16}$$

where $\widetilde{t}$ is a truncation parameter and $\widetilde{t}$ can be chosen as follows.

Step 1. Give an initial value $\widetilde{t}_0$. In our simulation, we take the initial value $\widetilde{t}_0 = \pi/(Y_{(n)} - Y_{(1)})$, where $Y_{(1)}$ and $Y_{(n)}$ are the minimum and maximum sample quantiles of $\{Y_1, \ldots, Y_n\}$.

Step 2. Set $\widetilde{t} = \widetilde{t}_k$, if the inequality condition $|\Delta(\widetilde{t})|^2 \geq 4(n-1)/n^2$ holds in one half of the interval $[-\widetilde{t}_k, \widetilde{t}_k]$. Otherwise, set $\widetilde{t}_{k+1} = c\widetilde{t}_k$, where $c > 1$ is a positive constant, in our simulation we take $c = 1.2$.

Step 3. Repeat Step 2 until $\tilde{t}_k$ satisfies the inequality condition.

From the simulation studies in Section 3, we can see that the choice of $\tilde{t}$ is convenient and fast, and the density estimator is not sensitive to the choice of $\tilde{t}$.

By (15) and the inverse Fourier transform, the self-consistent estimate of $f(x)$ is defined by

$$\widehat{f}_{sc}(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itx)\, \widehat{\phi}_{f_{sc}}(t)\, dt. \tag{17}$$

Let

$$K^*(u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itu)\, \frac{n}{2(n-1)\,\phi_\varepsilon(t)} \times \left(1 + \sqrt{1 - \frac{4(n-1)}{n^2\,|\Delta(t)|^2}}\right) I_A(t)\, dt. \tag{18}$$

Then, $\widehat{f}_{sc}(x)$ can be rewritten in the kernel form

$$\widehat{f}_{sc}(x) = \frac{1}{n} \sum_{j=1}^{n} K^*\left(x - Y_j\right). \tag{19}$$

We state the asymptotic behavior of the self-consistent estimator in the following theorem.

Theorem 1. *In addition to the regularity conditions (A1)–(A3) listed in the Appendix, assume that either of the following conditions holds.*

(a) *The error distribution is super smooth and $\tilde{t} = o\{((1/2b)\log(n(\log(n))^{-\gamma}))^{1/\beta}\}$.*

(b) *The error distribution is ordinary smooth and $\tilde{t} = o(n^{1/2(\beta_0+1)})$,*

where $\gamma = (1 + 2\beta_0 - \beta)_+/\beta$.

Then the self-consistent density estimator $\widehat{f}_{sc}(x)$ defined by (17) is consistent; that is, $\widehat{f}_{sc}(x) \xrightarrow{P} f(x)$, as $n \to \infty$.

Remark 2. As many other methods, for example, higher-order kernel estimators, spline estimators, wavelet estimators, orthogonal expansion estimators, and so forth, the resulting self-consistent estimators are not nonnegative, but those can be corrected without any error cost by translating the estimate downwards until the positive part is normalized to 1 and setting to 0 the negative part. For example, the modified estimator is defined by $\check{f}_{sc}(x) = \max\{\widehat{f}_{sc}(x) - d, 0\}$, where d is chosen in such a way that $\int_{-\infty}^{+\infty} \check{f}_{sc}(x)dx = 1$; see more details in Efromovich [27], Glad et al. [28], and so on.

Theorem 3. *Assume that conditions (A1)–(A3) hold and $f \in \mathcal{F}_{r,a,\delta}(m)$.*

(a) *When the error distribution is super smooth and $\tilde{t} = O\{((2/5b)\log(n(\log(n))^{-\gamma}))^{1/\beta}\}$, one has*

$$E\left|\widehat{f}_{sc}(x) - f(x)\right|^2$$
$$\leq O\left(\tilde{t}^{1-r-2\delta} \exp\left(-2a\tilde{t}^r\right) + n^{-1/5}(\log(n))^{\gamma/5}\right),$$

$$E\left\|\widehat{f}_{sc}(x) - f(x)\right\|^2$$
$$\leq O\left(\left(\tilde{t}^2 + 1\right)^{-\delta} \exp\left(-2a\tilde{t}^r\right) + n^{-1/5}(\log(n))^{\gamma/5}\right). \tag{20}$$

(b) *When the error distribution is ordinary smooth and $\tilde{t} = O(n^{3/5(2\beta_0+1)})$, one has*

$$E\left|\widehat{f}_{sc}(x) - f(x)\right|^2$$
$$\leq O\left(\tilde{t}^{1-r-2\delta} \exp\left(-2a\tilde{t}^r\right) + n^{-2/5}\right),$$

$$E\left\|\widehat{f}_{sc}(x) - f(x)\right\|^2$$
$$\leq O\left(\left(\tilde{t}^2 + 1\right)^{-\delta} \exp\left(-2a\tilde{t}^r\right) + n^{-2/5}\right). \tag{21}$$

2.2. Error Distribution Is Unknown. Suppose that, in addition to a sample $Y_1, \ldots, Y_n$ from f_Y, we observe a noise sample $\varepsilon_{-1}, \ldots, \varepsilon_{-M}$ from the unknown distribution f_ε. We assume that information about f_ε is obtained by the preliminary calibration measures before the work or is provided by an independent experiment.

When the characteristic function ϕ_ε in (15) is unknown, we then cannot estimate the density function using (17) directly. Thus, we need to estimate the characteristic function ϕ_ε. Here, we can estimate ϕ_ε based on the preliminary noise sample as follows:

$$\widehat{\phi}_\varepsilon(t) = \frac{1}{M} \sum_{j=1}^{M} \exp\left(it\varepsilon_{-j}\right). \tag{22}$$

Similar to Neumann [21], we introduce the following truncated estimator:

$$\frac{1}{\widetilde{\phi}_\varepsilon(t)} = \frac{1_{\{|\widehat{\phi}_\varepsilon(t)| \geq M^{-1/2}\}}}{\widehat{\phi}_\varepsilon(t)}$$
$$= \begin{cases} \dfrac{1}{\widehat{\phi}_\varepsilon(t)}, & \text{if } \left|\widehat{\phi}_\varepsilon(t)\right| \geq M^{-1/2}, \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

Then we can obtain the self-consistent estimator of the density function

$$\widetilde{f}_{sc}(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itx)\, \widetilde{\phi}_{f_{sc}}(t)\, dt, \tag{24}$$

where $\widetilde{\phi}_{f_{sc}}(t) = (n\Delta(t)/2(n-1)\widetilde{\phi}_\varepsilon(t))(1 + \sqrt{1 - (4(n-1)/n^2|\Delta(t)|^2)})I_A(t)$.

Theorem 4. *In addition to the regularity conditions (A1)–(A3) listed in the Appendix, assume that either of the following conditions holds.*

(a) *The error distribution is super smooth, $\beta \geq r$, $M \geq n(\log(n))^{-(1+2(\beta_0 \wedge \delta)-\beta)_+/\beta}$, and $\tilde{t} = o\{((1/2b)\log(n(\log(n))^{-\gamma}))^{1/\beta}\}$.*

(b) *The error distribution is super smooth, $\beta < r$, $M \to \infty$ as $n \to \infty$, and $\tilde{t} = o\{((1/2b)\log(n(\log(n))^{-\gamma}))^{1/\beta}\}$.*

(c) *The error distribution is ordinary smooth, $\beta_0 > \delta$, $M \geq n^{2(\beta_0-\delta)/(2\beta_0+1)}$, and $\tilde{t} = o(n^{1/(2\beta_0+1)})$.*

(d) *The error distribution is ordinary smooth, $\beta_0 \leq \delta$, $M \to \infty$ as $n \to \infty$, and $\tilde{t} = o(n^{1/(2\beta_0+1)})$.*

Then the self-consistent density estimator $\widetilde{f}_{sc}(x)$ defined by (24) is consistent; that is, $\widetilde{f}_{sc}(x) \xrightarrow{P} f(x)$, as $n \to \infty$.

Remark 5. In the cases $0 < \beta < r$ or $\beta_0 \leq \delta$ and $\beta = 0$, from the proof of theorem in Appendix, one can check that the estimator defined by (24) is consistent as long as the noise sample size M tends to infinity as n increases, while the rate of M tending to infinity is not restricted. Therefore, in practice, the noise sample size M does not need to be very large in these cases.

Theorem 6. *Assume that conditions (A1)–(A3) hold and $f \in \mathscr{F}_{r,a,\delta}(m)$.*

(a) *When the error distribution is super smooth, $\beta \geq r$, $M \geq n(\log(n))^{-(1+2(\beta_0 \wedge \delta)-\beta)_+/\beta}$, and $\tilde{t} = O\{((2/5b)\log(n(\log(n))^{-\gamma}))^{1/\beta}\}$, then one has*

$$E\left\|\widetilde{f}_{sc}(x) - f(x)\right\|^2$$
$$\leq O\left(\left(\tilde{t}^2+1\right)^{-\delta}\exp\left(-2a\tilde{t}^r\right) + n^{-1/5}\left(\log(n)\right)^{\gamma/5}\right). \tag{25}$$

(b) *When the error distribution is super smooth, $\beta < r$, $M \geq n^{1/5}(\log(n))^{-\gamma/5}$, and $\tilde{t} = O\{((2/5b)\log(n(\log(n))^{-\gamma}))^{1/\beta}\}$, then one has*

$$E\left\|\widetilde{f}_{sc}(x) - f(x)\right\|^2$$
$$\leq O\left(\left(\tilde{t}^2+1\right)^{-\delta}\exp\left(-2a\tilde{t}^r\right) + n^{-1/5}\left(\log(n)\right)^{\gamma/5}\right). \tag{26}$$

(c) *The error distribution is ordinary smooth, $M \geq n^{(2/5)+(6(\beta_0-\delta)_+/5(2\beta_0+1))}$, and $\tilde{t} = O(n^{3/5(2\beta_0+1)})$; then one has*

$$E\left\|\widetilde{f}_{sc}(x) - f(x)\right\|^2$$
$$\leq O\left(\left(\tilde{t}^2+1\right)^{-\delta}\exp\left(-2a\tilde{t}^r\right) + n^{-2/5}\right). \tag{27}$$

3. Simulation Studies

In this section, we report on the simulation studies to illustrate the finite sample performances of the proposed self-consistent (SC) estimation method and compare it with the kernel (KN) method. For comparison, we compute the estimators for different signal densities and different types of noise. The following densities are considered:

(a) gamma distribution $f(x) = 10.125x^4 e^{-3x}$ with shape parameter $\alpha = 5$ and scale parameter $\lambda = 3$;

(b) standard normal distribution $f(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$.

Two kinds of error distributions are considered to study their effects on the MISE of the estimators.

(c) Double exponential error: the density of ε is given by

$$f_\varepsilon(x) = \left(\sqrt{2}\sigma_0\right)^{-1}\exp\left(-\frac{\sqrt{2}\,|x|}{\sigma_0}\right) \quad \text{for } \sigma_0^2 = \frac{1}{16}. \tag{28}$$

(d) Normal error: the density of ε is given by

$$f_\varepsilon(x) = \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(\frac{-x^2}{2\sigma_0^2}\right) \quad \text{for } \sigma_0^2 = \frac{1}{16}. \tag{29}$$

For kernel method, we choose the Gaussian kernel for double exponential error, and for normal error we suppose that the kernel has a Fourier transform $\phi_K(t) = (1-t^2)^2_+$. For the unknown error distributions, we suppose that the noise sample $\varepsilon_{-1}, \ldots, \varepsilon_{-M}$ comes from the above two error distributions. Throughout the simulations, we set $n = 100, 200, 400$ and $M = 50, 100$. The MISE of the estimators are computed with 1000 random samples, and the results are reported in Tables 1 and 2. When f_ε is known or unknown, the gamma density estimator through the double exponential noise and the normal density estimator through the normal noise are shown in Figures 1 and 2, respectively.

Tables 1 and 2 indicate the following simulation results.

(1) When the error distribution is unknown, we can see that the estimator of the characteristic function of the error does not spoil the procedure so much. It even happens that the estimation with unknown error distribution works better, which is probably due to the truncation (23).

(2) For the same observation sample, the self-consistent method performs slightly better than the kernel method. It is worth mentioning that the self-consistent method can save more time than the kernel method, because the truncation parameter $\tilde{t}$ can be chosen conveniently.

(3) The MISE obtained by self-consistent and kernel methods decrease as n increases whether the error distribution is known or unknown.

Figures 1 and 2 clearly show that the self-consistent estimators perform well, no matter whether the error distribution is known or unknown. Moreover, for $M = 10$, the

TABLE 1: MISE for the self-consistent estimator and the kernel estimator when the error distribution is double exponential.

n	Case	Gamma		Normal	
		SC	KN	SC	KN
100	f_ε known	0.0021	0.0036	0.0105	0.0102
	$M = 50$	0.0022	0.0041	0.0105	0.0106
	$M = 100$	0.0021	0.0039	0.0102	0.0103
200	f_ε known	0.0011	0.0019	0.0056	0.0065
	$M = 50$	0.0011	0.0025	0.0063	0.0068
	$M = 100$	0.0013	0.0021	0.0057	0.0067
400	f_ε known	0.0006	0.0011	0.0031	0.0037
	$M = 50$	0.0007	0.0012	0.0036	0.0035
	$M = 100$	0.0006	0.0012	0.0032	0.0033

TABLE 2: MISE for the self-consistent estimator and the kernel estimator when the error distribution is normal.

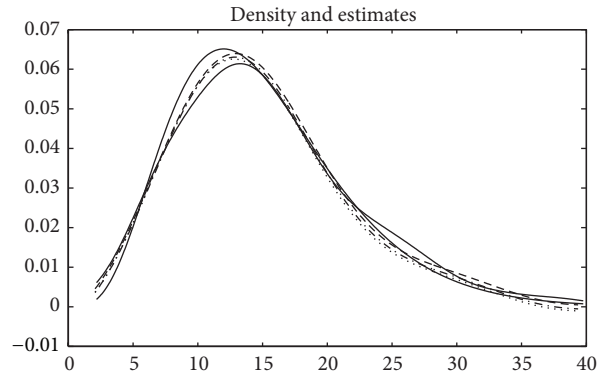| n	Case	Gamma		Normal	
		SC	KN	SC	KN
100	f_ε known	0.0028	0.0039	0.0102	0.0109
	$M = 50$	0.0030	0.0042	0.0106	0.0111
	$M = 100$	0.0029	0.0038	0.0106	0.0110
200	f_ε known	0.0012	0.0023	0.0061	0.0073
	$M = 50$	0.0015	0.0023	0.0069	0.0075
	$M = 100$	0.0012	0.0024	0.0059	0.0071
400	f_ε known	0.0007	0.0012	0.0035	0.0039
	$M = 50$	0.0007	0.0013	0.0038	0.0039
	$M = 100$	0.0007	0.0011	0.0036	0.0038

FIGURE 1: Estimates of gamma density through double exponential noise for $n = 400$. The thick solid line is the true density, the thin solid line and the dashed line are the kernel estimator and the SC estimator when f_ε is known, respectively, and the dot-dashed line and the dotted line are the SC estimators when the noise sample sizes $M = 50$ and $M = 10$, respectively.
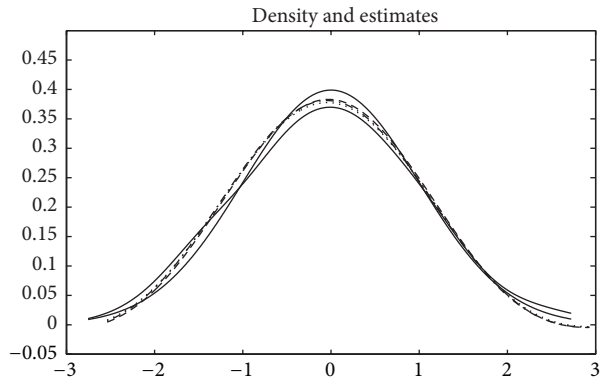

FIGURE 2: Estimates of normal density through normal noise for $n = 400$. The thick solid line is the true density, the thin solid line and the dashed line are the kernel estimator and the SC estimator when f_ε is known, respectively, and the dot-dashed line and the dotted line are the SC estimators when the noise sample sizes $M = 50$ and $M = 10$, respectively.

results show that our method is still very satisfactory for small noise sample size.

Next, we evaluate the sensitivity of the self-consistent density estimation procedure on the choice of truncation parameter $\tilde{t}$. In this simulation, we only consider the case when the error distribution is double exponential, and the value of $\tilde{t}$ is fixed at $\tilde{t} = \tilde{t}^*$, $\tilde{t}^*/1.25$, and $1.25\tilde{t}^*$, respectively, where $\tilde{t}^*$ is obtained by the iterative algorithm given in Section 2.1. The MISE of the self-consistent estimator is reported in Table 3. From Table 3 we can see that, for given n, the performance of the proposed method does not depend sensitively on the choice of the value of $\tilde{t}$.

4. Application to Real Data

We now use two real datasets to illustrate the proposed method. The first dataset is the density of direction chosen by an ant to an evenly illuminated black target. The experiment with 100 ants was described by Fisher [29], who concluded that the ants tend to run toward the target with some moderate variation. For this particular example, Y, a point where an ant intersects a circle, can be treated as indirect observation of the direction X chosen by an ant. The estimation method of (17) is used to analyze the dataset. In Figure 3, the solid line denotes the estimate based on the assumption that the data can be accurately measured (i.e., $Y = X$), and the

dashed and dot-dashed lines denote the case of $N(0, \sigma^2)$ measurement errors with $\sigma = 0.3$ and $\sigma = 0.5$, respectively. We compare our results with Efromovich [30] who already used the orthogonal series approach to estimate the density.

From Figure 3, we find that when the data are accurately measured ($\sigma = 0$), the corresponding density estimate is rather smooth and has a pronounced mode (direction of the target) with a relatively large deviation. When the standard deviation σ increases, it makes the density more picky interestingly and makes two additional local modes more significant. Therefore, the assumption about measurement error reveals the presence of three distinct groups (strata) of ants. A majority clearly tends to move toward the target with a rather small variation, and a minority tends to move in two other directions about the target also with small variation. These findings basically agree with those that were discovered by Efromovich [30]. But the two directions are not symmetric

TABLE 3: MISE for the self-consistent estimator with different values of $\widetilde{t}$ when the error distribution is double exponential.

| n	Case	Gamma			Normal		
		$\widetilde{t}^*$	$\widetilde{t}^*/1.25$	$1.25\widetilde{t}^*$	$\widetilde{t}^*$	$\widetilde{t}^*/1.25$	$1.25\widetilde{t}^*$
100	f_ε known	0.0020	0.0021	0.0020	0.0105	0.0105	0.0107
	$M = 50$	0.0022	0.0024	0.0020	0.0108	0.0106	0.0106
	$M = 100$	0.0021	0.0022	0.0022	0.0104	0.0103	0.0108
200	f_ε known	0.0011	0.0014	0.0012	0.0057	0.0054	0.0057
	$M = 50$	0.0013	0.0015	0.0013	0.0060	0.0059	0.0058
	$M = 100$	0.0012	0.0014	0.0012	0.0059	0.0058	0.0062
400	f_ε known	0.0006	0.0007	0.0007	0.0032	0.0034	0.0032
	$M = 50$	0.0007	0.0008	0.0007	0.0035	0.0034	0.0033
	$M = 100$	0.0006	0.0006	0.0007	0.0032	0.0035	0.0032
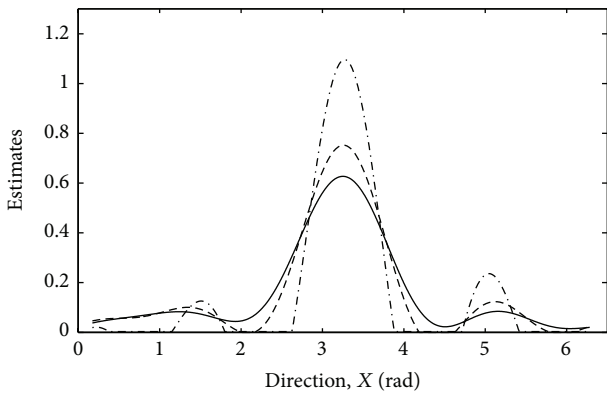
FIGURE 3: Estimates of density of movement of ants. The solid line is the estimate based on direct data, and the dashed and dot-dashed lines are the estimates under assumption on normal error with zero mean and standard deviations $\sigma = 0.3$ and $\sigma = 0.5$, respectively.

here, and the variation of the right side is slightly less than that of the left side.

The second example involves estimating density of the magnitudes of Alaskan earthquakes for the period from 1969 to 1978. The data comes from the National Oceanic and Atmospheric Administration's Hypocenter Data File [31]. In this dataset, Y denotes the logarithm of the seismogram amplitude of longitudinal body waves for 62 Alaskan earthquakes. There is a measurement error associated with the observations, which includes errors made in determining the amplitude of ground motion arising from such things as the orientation of a limited number of observation stations to the fault plane of the earthquake. In Figure 4, the solid line denotes the estimate based on assumption that the body waves are accurately measured (i.e., $Y = X$), and the dashed and dot-dashed lines denote the case of Laplace measurement errors with position parameter $\alpha = 0$ and scale parameters $\lambda = 0.3$ and $\lambda = 0.5$, respectively.

Figure 4 clearly shows the nature of the indirect data. Under the assumption that the data are accurately measured ($\lambda = 0$), the corresponding density estimate is smooth and
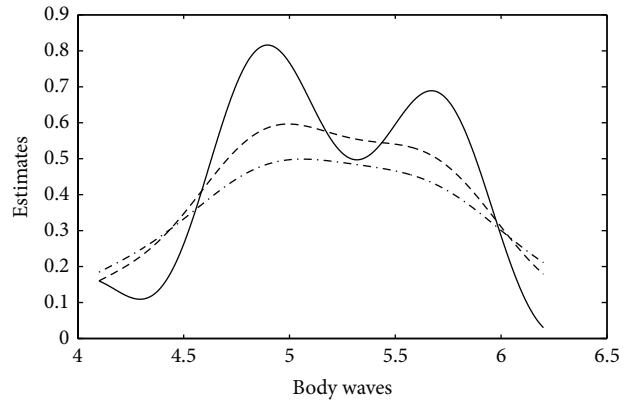

FIGURE 4: Estimates of density of body waves. The solid line is the estimate based on direct data, and the dashed and dot-dashed lines are the estimates under assumption on Laplace error with position parameter $\alpha = 0$ and scale parameters $\lambda = 0.3$ and $\lambda = 0.5$, respectively.

has two pronounced modes. For the case of $\lambda \neq 0$, the distinctions between the density estimates become more noticeable. It makes the data become a unimodal distribution with λ increasing. It is clear that the presence of the measurement error allows one to gain some insight into the data.

Appendix

Proofs of Theorems

We begin this section by listing the regularity conditions needed in the proofs of all the theorems:

(A1) $f(x)$ is square integrable and its Fourier transform is integrable;

(A2) $\phi_\varepsilon(t) \neq 0$ for any $t \in \mathbb{R}$;

(A3) $\widetilde{t} \to \infty$ as $n \to \infty$.

In the following text, we use C to represent any positive constant. The proofs of theorems rely on the following lemma.

Lemma A.1. Let $D(t) = (1/\widetilde{\phi}_\varepsilon(t)) - (1/\phi_\varepsilon(t))$; then there is a positive constant C such that

$$E\left\{|D(t)|^2\right\} \leq C \min\left\{\frac{1}{|\phi_\varepsilon(t)|^2}, \frac{M^{-1}}{|\phi_\varepsilon(t)|^4}\right\}. \tag{A.1}$$

Proof. Lemma A.1 is from Lemma 2.1 in Neumann [21]; the proof is omitted. □

Proof of Theorem 1. By (15), note that $\widehat{\phi}_{f_{sc}}(t) \widehat{=} \widehat{\phi}_Y(t)/\phi_\varepsilon(t)$ and $\widehat{\phi}_{f_{sc}}(t) = 0$ for $|t| > \widetilde{t}$. Then, we have

$$
\left| \widehat{f}_{sc}(x) - f(x) \right|
$$

$$
= \left| \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itx) \left\{ \widehat{\phi}_{f_{sc}}(t) - \phi_f(t) \right\} dt \right|
$$

$$
\leq \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left| \exp(-itx) \right| \left| \widehat{\phi}_{f_{sc}}(t) - \phi_f(t) \right| dt
$$

$$
= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\left| \widehat{\phi}_Y(t) - \phi_Y(t) \right|}{\left| \phi_\varepsilon(t) \right|} dt
$$

$$
= \frac{1}{2\pi} \int_{-\widetilde{t}}^{\widetilde{t}} \frac{\left| \widehat{\phi}_Y(t) - \phi_Y(t) \right|}{\left| \phi_\varepsilon(t) \right|} dt
$$

$$
+ \frac{1}{2\pi} \int_{|t|>\widetilde{t}} \left| \phi_f(t) \right| dt
$$

$$
\leq \frac{1}{2\pi} \int_{-\widetilde{t}}^{\widetilde{t}} \frac{\left| \widehat{\phi}_Y(t) - \Delta(t) \right|}{\left| \phi_\varepsilon(t) \right|} dt + \frac{1}{2\pi} \int_{-\widetilde{t}}^{\widetilde{t}} \frac{\left| \Delta(t) - \phi_Y(t) \right|}{\left| \phi_\varepsilon(t) \right|} dt
$$

$$
+ \frac{1}{2\pi} \int_{|t|>\widetilde{t}} \left| \phi_f(t) \right| dt
$$

$$
\widehat{=} J_1 + J_2 + J_3.
$$

$$(A.2)$$

By conditions (A1) and (A3), as $n \to \infty$, it is easy to show that $J_3 \to 0$. To finish the proof of Theorem 1, we need to show that $J_1 \xrightarrow{P} 0$ and $J_2 \xrightarrow{P} 0$.

We first consider the case that the error distribution is super smooth ($\beta > 0$). For J_2, by (3), we have

$$
J_2 \leq C \int_{-\widetilde{t}}^{\widetilde{t}} \left| \Delta(t) - \phi_Y(t) \right| \exp\left(b|t|^\beta \right) \left(t^2 + 1 \right)^{\beta_0/2} dt \widehat{=} CV_1.
$$

$$(A.3)$$

By Cauchy-Schwarz inequality and Jensen inequality, it is easy to verify that

$$
E\left(V_1^2 \right) \leq \int_{-\widetilde{t}}^{\widetilde{t}} E \left| \Delta(t) - \phi_Y(t) \right|^2 dt
$$

$$
\times \int_{-\widetilde{t}}^{\widetilde{t}} \exp\left(2b|t|^\beta \right) \left(t^2 + 1 \right)^{\beta_0} dt
$$

$$
\leq \frac{1}{n} \int_{-\widetilde{t}}^{\widetilde{t}} \exp\left(2b|t|^\beta \right) \left(t^2 + 1 \right)^{\beta_0} dt
$$

$$
= \frac{2}{n} \int_0^{\widetilde{t}} \exp\left(2bt^\beta \right) \left(t^2 + 1 \right)^{\beta_0} dt.
$$

$$(A.4)$$

By Lemma 2 in Butucea and Tsybakov [5] and a simple calculation, we have

$$
EV_1^2 \longrightarrow 0 \quad \text{if } \widetilde{t} = o\left\{ \left(\frac{1}{2b} \log\left(n\left(\log(n) \right)^{-\gamma} \right) \right)^{1/\beta} \right\}, \quad (A.5)
$$

where $\gamma = (1 + 2\beta_0 - \beta)_+/\beta$. Hence, we show that $J_2 \xrightarrow{P} 0$.

Now we consider J_1. Let $\Delta_+ = \{t : |\Delta(t)|^2 \geq 4(n-1)/n^2\}$ and $\Delta_- = \{t : |\Delta(t)|^2 < 4(n-1)/n^2\}$; we have

$$
J_1 = \frac{1}{2\pi} \int_{(-\widetilde{t},\widetilde{t})\cap\Delta_+} \frac{|\Delta(t)|}{|\phi_\varepsilon(t)|}
$$

$$
\times \left(1 - \frac{n}{2(n-1)} \left[1 + \sqrt{1 - \frac{4(n-1)}{n^2 |\Delta(t)|^2}} \right] \right) dt
$$

$$
+ \frac{1}{2\pi} \int_{(-\widetilde{t},\widetilde{t})\cap\Delta_-} \frac{|\Delta(t)|}{|\phi_\varepsilon(t)|} dt.
$$

$$(A.6)$$

By (3) and the inequality $\sqrt{1-x} \geq 1 - \sqrt{x}$ for $x \in (0,1)$, we have

$$
J_1 \leq \frac{1}{2\pi} \int_{(-\widetilde{t},\widetilde{t})\cap\Delta_+} \left(\frac{1}{\sqrt{n-1}} - \frac{|\Delta(t)|}{n-1} \right) \frac{1}{|\phi_\varepsilon(t)|} dt
$$

$$
+ \frac{C\sqrt{n-1}}{n} \int_{(-\widetilde{t},\widetilde{t})\cap\Delta_-} \frac{1}{|\phi_\varepsilon(t)|} dt
$$

$$
\leq C \left(\frac{1}{\sqrt{n-1}} + \frac{1}{n-1} \right) \int_{(-\widetilde{t},\widetilde{t})\cap\Delta_+} \frac{1}{|\phi_\varepsilon(t)|} dt
$$

$$
+ \frac{C\sqrt{n-1}}{n} \int_{(-\widetilde{t},\widetilde{t})\cap\Delta_-} \frac{1}{|\phi_\varepsilon(t)|} dt.
$$

$$(A.7)$$

As $\widetilde{t} = o\{((1/2b)\log(n(\log(n))^{-\gamma}))^{1/\beta}\}$, similar to the proof of (A.3), it is easy to show that $J_1 \xrightarrow{P} 0$. Thus, we finish the proof of super smooth case in Theorem 1.

Next we consider the case that the error distribution is ordinary smooth ($\beta = 0$, $\beta_0 > 0$). By (3), we can derive that

$$
J_2 \leq C \int_{-\widetilde{t}}^{\widetilde{t}} \left| \Delta(t) - \phi_Y(t) \right| \left(t^2 + 1 \right)^{\beta_0/2} dt \widehat{=} CV_2. \quad (A.8)
$$

Again using Cauchy-Schwarz inequality and Jensen inequality, we have

$$
E\left(V_2^2 \right) \leq \int_{-\widetilde{t}}^{\widetilde{t}} E \left| \Delta(t) - \phi_Y(t) \right|^2 dt \int_{-\widetilde{t}}^{\widetilde{t}} \left(t^2 + 1 \right)^{\beta_0} dt
$$

$$
\leq \frac{2}{n} \int_0^{\widetilde{t}} \left(t^2 + 1 \right)^{\beta_0} dt = \frac{C}{n} \widetilde{t}^{2\beta_0+1} \left(1 + o(1) \right),
$$

$$(A.9)$$

which implies that $E(V_2^2) \to 0$ if $n\widetilde{t}^{-(2\beta_0+1)} \to \infty$. Hence, we obtain that $J_2 \xrightarrow{P} 0$.

For J_1, using the same arguments, we can obtain that

$$
\begin{aligned}
J_1 \leq{}& C\left(\frac{1}{\sqrt{n-1}} + \frac{1}{n-1}\right) \int_{(-\tilde{t},\tilde{t})\cap\Delta_+} \frac{1}{|\phi_\varepsilon(t)|} dt \\
&+ \frac{C\sqrt{n-1}}{n} \int_{(-\tilde{t},\tilde{t})\cap\Delta_-} \frac{1}{|\phi_\varepsilon(t)|} dt \\
\leq{}& C\left(\frac{1}{\sqrt{n-1}} + \frac{1}{n-1}\right) \int_{(-\tilde{t},\tilde{t})\cap\Delta_+} \left(t^2+1\right)^{\beta_0/2} dt \\
&+ \frac{C\sqrt{n-1}}{n} \int_{(-\tilde{t},\tilde{t})\cap\Delta_-} \left(t^2+1\right)^{\beta_0/2} dt \\
\leq{}& C\left(\frac{1}{\sqrt{n-1}} + \frac{1}{n-1} + \frac{\sqrt{n-1}}{n}\right) \tilde{t}^{\beta_0+1} (1+o(1)).
\end{aligned}
$$
(A.10)

Similar to the proof of (A.8) we can show that $J_1 \to 0$ as $n\tilde{t}^{-(2\beta_0+1)} \to \infty$ and $n \to \infty$. This finishes the proof of Theorem 1. □

Proof of Theorem 3. From (19), by Lemma 2 in Butucea and Tsybakov [5] and Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
&\left|E\hat{f}_{\mathrm{sc}}(x) - f(x)\right| \\
&\leq \left|\frac{1}{2\pi}\frac{n}{n-1}\int_{-\tilde{t}}^{\tilde{t}} \exp(-itx)\phi_f(t)\,dt - f(x)\right| \\
&= \left|\frac{1}{n-1}f(x) + \frac{1}{2\pi}\frac{n}{n-1}\int_{|t|>\tilde{t}} \exp(-itx)\phi_f(t)\,dt\right| \\
&\leq \frac{1}{n-1}f(x) + \frac{1}{2\pi}\frac{n}{n-1}\int_{|t|>\tilde{t}} |\phi_f(t)|\,dt \\
&= \frac{1}{n-1}f(x) + \frac{1}{2\pi}\frac{n}{n-1} \\
&\quad \times \int_{|t|>\tilde{t}} |\phi_f(t)| \exp\left(a|t|^r\right)\left(t^2+1\right)^{\delta/2} \\
&\qquad \times \exp\left(-a|t|^r\right)\left(t^2+1\right)^{-\delta/2} dt \\
&\leq \frac{1}{n-1}f(x) + \frac{1}{2\pi}\frac{n\sqrt{m}}{n-1} \\
&\quad \times \left(\int_{|t|>\tilde{t}} \exp\left(-2a|t|^r\right)\left(t^2+1\right)^{-\delta} dt\right)^{1/2} \\
&= \frac{1}{n-1}f(x) + \frac{1}{2\pi}\frac{n\sqrt{m}}{n-1} \\
&\quad \times \left(\int_{|u|<\tilde{t}^{-1}} \exp\left(-2a|u|^{-r}\right)\left(u^{-2}+1\right)^{-\delta} u^{-2} du\right)^{1/2} \\
&\leq \frac{1}{n-1}f(x) \\
&\quad + \frac{1}{2\pi}\frac{n\sqrt{m}}{(n-1)\sqrt{ar}} \tilde{t}^{(1-r-2\delta)/2} \exp\left(-a\tilde{t}^r\right)(1+o(1)).
\end{aligned}
$$
(A.11)

Now we compute the variance of $\hat{f}_{\mathrm{sc}}(x)$ as follows:

$$
\begin{aligned}
\operatorname{Var}\left(\hat{f}_{\mathrm{sc}}(x)\right) &= \frac{1}{n}\operatorname{Var}\left(K^*(x-Y)\right) \\
&\leq \frac{1}{n}E\left(K^*(x-Y)\right)^2 \\
&\leq \frac{1}{2n\pi^2}E\int_{-\infty}^{+\infty} \frac{n^2}{(n-1)^2|\phi_\varepsilon|^2} I_A(t)\,dt \\
&\leq C\frac{n}{(n-1)^2}\int_0^{\tilde{t}} \frac{1}{|\phi_\varepsilon|^2}\,dt.
\end{aligned}
$$
(A.12)

When the error distribution is super smooth ($\beta > 0$), by (3), we have

$$
\operatorname{Var}\left(\hat{f}_{\mathrm{sc}}(x)\right) \leq C\frac{n}{(n-1)^2}\int_0^{\tilde{t}} \exp\left(2bt^\beta\right)\left(t^2+1\right)^{\beta_0}\,dt.
$$
(A.13)

If $\tilde{t} = O\{((2/5b)\log(n(\log(n))^{-\gamma}))^{1/\beta}\}$, we can obtain that

$$
\operatorname{Var}\left(\hat{f}_{\mathrm{sc}}(x)\right) \leq O\left(n^{-1/5}(\log(n))^{\gamma/5}\right).
$$
(A.14)

Therefore, $E|\hat{f}_{\mathrm{sc}}(x) - f(x)|^2 \leq O(\tilde{t}^{1-r-2\delta}\exp(-2a\tilde{t}^r) + n^{-1/5}(\log(n))^{\gamma/5})$.

When the error distribution is ordinary smooth ($\beta = 0$, $\beta_0 > 0$), by (3), we have

$$
\begin{aligned}
\operatorname{Var}\left(\hat{f}_{\mathrm{sc}}(x)\right) &\leq C\frac{n}{(n-1)^2}\int_0^{\tilde{t}} \left(t^2+1\right)^{\beta_0}\,dt \\
&= C\frac{n}{(n-1)^2}\tilde{t}^{2\beta_0+1}.
\end{aligned}
$$
(A.15)

If $\tilde{t} = O(n^{3/5(2\beta_0+1)})$, we can obtain that $\operatorname{Var}(\hat{f}_{\mathrm{sc}}(x)) \leq O(n^{-2/5})$. Therefore,

$$
E\left|\hat{f}_{\mathrm{sc}}(x) - f(x)\right|^2 \leq O\left(\tilde{t}^{1-r-2\delta}\exp\left(-2a\tilde{t}^r\right) + n^{-2/5}\right).
$$
(A.16)

Next, we consider the MISE of the estimator $\hat{f}_{\mathrm{sc}}(x)$. By Parseval theorem, we have

$$
\begin{aligned}
E\left\|\hat{f}_{\mathrm{sc}}(x) - f(x)\right\|^2 &= E\int_{-\infty}^{+\infty} \left[\tilde{f}_{\mathrm{sc}}(x) - f(x)\right]^2 dx \\
&= \frac{1}{2\pi}E\int_{-\infty}^{+\infty} \left|\hat{\phi}_{\mathrm{sc}}(t) - \phi_t\right|^2 dt \\
&= \frac{1}{2\pi}E\int_{-\infty}^{+\infty} \left|\frac{\hat{\phi}_Y(t) - \phi_Y(t)}{\phi_\varepsilon(t)}\right|^2 dt \\
&= \frac{1}{2\pi}\int_{-\tilde{t}}^{\tilde{t}} E\left|\frac{\hat{\phi}_Y(t) - \phi_Y(t)}{\phi_\varepsilon(t)}\right|^2 dt \\
&\quad + \frac{1}{2\pi}\int_{|t|>\tilde{t}} |\phi_f(t)|^2\,dt \\
&\cong J_4 + J_5.
\end{aligned}
$$
(A.17)

For J_5, by (4), we have

$$
\begin{aligned}
J_5 &= \frac{1}{2\pi} \int_{|t|>\tilde{t}} \left|\phi_f(t)\right|^2 \exp\left(2a\,|t|^r\right) \left(t^2+1\right)^{\delta} \\
&\quad \times \exp\left(-2a\,|t|^r\right) \left(t^2+1\right)^{-\delta} dt \\
&\leq \frac{m}{2\pi} \exp\left(-2a\,|\tilde{t}|^r\right) \left(\tilde{t}^2+1\right)^{-\delta}.
\end{aligned}
\tag{A.18}
$$

We now consider J_4. By some simple calculations, we can drive that

$$
\begin{aligned}
J_4 &\leq \frac{1}{\pi} \int_{-\tilde{t}}^{\tilde{t}} \frac{E\left|\widehat{\phi}_Y(t) - \Delta(t)\right|^2}{\left|\phi_\varepsilon(t)\right|^2} dt \\
&\quad + \frac{1}{\pi} \int_{-\tilde{t}}^{\tilde{t}} \frac{E\left|\Delta(t) - \phi_Y(t)\right|^2}{\left|\phi_\varepsilon(t)\right|^2} dt \\
&\leq \frac{1}{\pi} \int_{(-\tilde{t},\tilde{t})\cap\Delta_+} \frac{E\left|\widehat{\phi}_Y(t) - \Delta(t)\right|^2}{\left|\phi_\varepsilon(t)\right|^2} dt \\
&\quad + \frac{1}{\pi} \int_{(-\tilde{t},\tilde{t})\cap\Delta_-} \frac{E\left|\widehat{\phi}_Y(t) - \Delta(t)\right|^2}{\left|\phi_\varepsilon(t)\right|^2} dt \\
&\quad + \frac{1}{n\pi} \int_{-\tilde{t}}^{\tilde{t}} \frac{1}{\left|\phi_\varepsilon(t)\right|^2} dt \\
&\leq \frac{1}{\pi} \left(\frac{1}{n-1} + \frac{1}{(n-1)^2} + \frac{2}{(n-1)^{3/2}}\right) \int_{(-\tilde{t},\tilde{t})\cap\Delta_+} \frac{1}{\left|\phi_\varepsilon(t)\right|^2} dt \\
&\quad + \frac{1}{\pi} \frac{4(n-1)}{n^2} \int_{(-\tilde{t},\tilde{t})\cap\Delta_-} \frac{1}{\left|\phi_\varepsilon(t)\right|^2} dt \\
&\quad + \frac{1}{n\pi} \int_{-\tilde{t}}^{\tilde{t}} \frac{1}{\left|\phi_\varepsilon(t)\right|^2} dt \leq \frac{C}{n} \int_{-\tilde{t}}^{\tilde{t}} \frac{1}{\left|\phi_\varepsilon(t)\right|^2} dt.
\end{aligned}
\tag{A.19}
$$

Therefore, when the error distribution is super smooth ($\beta > 0$), we have

$$
J_4 \leq \frac{2C}{n} \int_0^{\tilde{t}} \exp\left(2bt^{\beta}\right) \left(t^2+1\right)^{\beta_0} dt.
\tag{A.20}
$$

Choosing $\tilde{t} = O\{((2/5b)\log(n(\log(n))^{-\gamma}))^{1/\beta}\}$, we have

$$
\begin{aligned}
&E\left\|\widehat{f}_{sc}(x) - f(x)\right\|^2 \\
&\leq O\left(\left(\tilde{t}^2+1\right)^{-\delta} \exp\left(-2a\tilde{t}^r\right) + n^{-1/5}\left(\log(n)\right)^{\gamma/5}\right).
\end{aligned}
\tag{A.21}
$$

When the error distribution is ordinary smooth ($\beta = 0$, $\beta_0 > 0$), we have

$$
J_4 \leq \frac{2C}{n} \int_0^{\tilde{t}} \left(t^2+1\right)^{\beta_0} dt = \frac{2C}{n} \tilde{t}^{2\beta_0+1}\left(1 + o(1)\right).
\tag{A.22}
$$

Choosing $\tilde{t} = O(n^{3/5(2\beta_0+1)})$, we have

$$
E\left\|\widehat{f}_{sc}(x) - f(x)\right\|^2 \leq O\left(\left(\tilde{t}^2+1\right)^{-\delta} \exp\left(-2a\tilde{t}^r\right) + n^{-2/5}\right).
\tag{A.23}
$$

The proof of Theorem 3 is completed. □

Proof of Theorem 4. The proof of Theorem 4 can easily be derived by Theorem 6, we hence omit the details, and we only need to prove Theorem 6. □

Proof of Theorem 6. By (24) and Parseval theorem, we have

$$
\begin{aligned}
&E \int_{-\infty}^{+\infty} \left[\widetilde{f}_{sc}(x) - f(x)\right]^2 dx \\
&= \frac{1}{2\pi} E \int_{-\infty}^{+\infty} \left|\frac{\widehat{\phi}_Y(t)}{\widetilde{\phi}_\varepsilon(t)} - \frac{\phi_Y(t)}{\phi_\varepsilon(t)}\right|^2 dt \\
&= \frac{1}{2\pi} E \int_{-\infty}^{+\infty} \left|\frac{\widehat{\phi}_Y(t)}{\widetilde{\phi}_\varepsilon(t)} - \frac{\widehat{\phi}_Y(t)}{\phi_\varepsilon(t)} + \frac{\widehat{\phi}_Y(t) - \phi_Y(t)}{\phi_\varepsilon(t)}\right|^2 dt \\
&\leq \frac{1}{\pi} \int_{-\infty}^{+\infty} E\left|\widehat{\phi}_Y(t)\right|^2 E\left|D(t)\right|^2 dt \\
&\quad + \frac{1}{\pi} \int_{-\infty}^{\infty} E\frac{\left|\widehat{\phi}_Y(t) - \phi_Y(t)\right|^2}{\left|\phi_\varepsilon(t)\right|^2} dt \\
&= \frac{1}{\pi} \int_{-\infty}^{+\infty} E\left|\widehat{\phi}_Y(t)\right|^2 E\left|D(t)\right|^2 dt \\
&\quad + \frac{1}{\pi} \int_{-\tilde{t}}^{\tilde{t}} E\frac{\left|\widehat{\phi}_Y(t) - \phi_Y(t)\right|^2}{\left|\phi_\varepsilon(t)\right|^2} dt \\
&\quad + \frac{1}{\pi} \int_{|t|>\tilde{t}} \left|\phi_f(t)\right|^2 dt \\
&\widehat{=} J_6 + 2J_4 + 2J_5.
\end{aligned}
\tag{A.24}
$$

Therefore, we only need to deal with J_6. Invoking Lemma A.1, it is easy to show that

$$
\begin{aligned}
J_6 &= \frac{1}{\pi} \int_{-\tilde{t}}^{\tilde{t}} E\left|\widehat{\phi}_Y(t)\right|^2 E\left|D(t)\right|^2 dt \\
&\leq \frac{2}{\pi} \int_{-\tilde{t}}^{\tilde{t}} E\left|\widehat{\phi}_Y(t) - \phi_Y(t)\right|^2 E\left|D(t)\right|^2 dt \\
&\quad + \frac{2}{\pi} \int_{-\tilde{t}}^{\tilde{t}} \left|\phi_Y(t)\right|^2 E\left|D(t)\right|^2 dt \\
&\leq \frac{2}{n\pi} \int_{-\tilde{t}}^{\tilde{t}} E\left|D(t)\right|^2 dt
\end{aligned}
$$

$$+ \frac{2}{\pi} \int_{-\tilde{t}}^{\tilde{t}} \left| \phi_f(t) \right|^2 \left| \phi_\varepsilon \right|^2 E \left| D(t) \right|^2 dt$$

$$\leq \frac{2}{n\pi} \int_{-\tilde{t}}^{\tilde{t}} \frac{1}{\left| \phi_\varepsilon(t) \right|^2} dt + \frac{2}{M\pi} \int_{-\tilde{t}}^{\tilde{t}} \frac{\left| \phi_f(t) \right|^2}{\left| \phi_\varepsilon(t) \right|^2} dt.$$

$$\text{(A.25)}$$

Then, when the error distribution is super smooth ($\beta > 0$), by (4), we have

$$J_6 \leq \frac{C}{n} \int_0^{\tilde{t}} \exp\left(2bt^\beta\right) \left(t^2 + 1\right)^{\beta_0} dt$$

$$+ \frac{mC}{M} \sup_{t \in [0,\tilde{t}]} \left(t^2 + 1\right)^{\beta_0 - \delta} \exp\left\{2bt^\beta - 2at^r\right\}.$$

$$\text{(A.26)}$$

If $\beta \geq r$, it is easy to show that

$$J_6 \leq \frac{C}{n} \exp\left(2b\tilde{t}^\beta\right) \tilde{t}^{1+2\beta_0-\beta} \left(1 + o(1)\right)$$

$$+ \frac{mC}{M} \tilde{t}^{2(\beta_0-\delta)_+} \exp\left(2b\tilde{t}^\beta\right) \left(1 + o(1)\right).$$

$$\text{(A.27)}$$

Hence, by choosing $M \geq n(\log(n))^{-(1+2(\beta_0 \wedge \delta)-\beta)_+/\beta}$ and $\tilde{t} = O\{((2/5b)\log(n(\log(n))^{-\gamma}))^{1/\beta}\}$, together with (A.18) and (A.20), we can drive that

$$E \left\| \tilde{f}_{\text{sc}}(x) - f(x) \right\|^2$$

$$\leq O\left(\left(\tilde{t}^2 + 1\right)^{-\delta} \exp\left(-2a\tilde{t}^r\right) + n^{-1/5} \left(\log(n)\right)^{\gamma/5} \right).$$

$$\text{(A.28)}$$

If $\beta < r$, we have

$$J_6 \leq \frac{C}{n} \exp\left(2b\tilde{t}^\beta\right) \tilde{t}^{1+2\beta_0-\beta} \left(1 + o(1)\right) + \frac{mC}{M}. \quad \text{(A.29)}$$

Hence, by choosing $M \geq n^{1/5}(\log(n))^{-\gamma/5}$ and $\tilde{t} = O\{((2/5b)\log(n(\log(n))^{-\gamma}))^{1/\beta}\}$, together with (A.18) and (A.20), we can drive that

$$E \left\| \tilde{f}_{\text{sc}}(x) - f(x) \right\|^2$$

$$\leq O\left(\left(\tilde{t}^2 + 1\right)^{-\delta} \exp\left(-2a\tilde{t}^r\right) + n^{-1/5} \left(\log(n)\right)^{\gamma/5} \right).$$

$$\text{(A.30)}$$

When the error distribution is ordinary smooth ($\beta = 0$, $\beta_0 > 0$), we have

$$J_6 \leq \frac{C}{n} \int_0^{\tilde{t}} \left(t^2 + 1\right)^{\beta_0} dt$$

$$+ \frac{mC}{M} \sup_{t \in [0,\tilde{t}]} \left(t^2 + 1\right)^{\beta_0 - \delta} \exp\left\{-2at^r\right\} \quad \text{(A.31)}$$

$$\leq \frac{C}{n} \tilde{t}^{2\beta_0+1} \left(1 + o(1)\right) + \frac{mC}{M} \tilde{t}^{2(\beta_0-\delta)_+}.$$

Then, by choosing $M \geq n^{(2/5)+(6(\beta_0-\delta)_+/5(2\beta_0+1))}$ and $\tilde{t} = O(n^{3/5(2\beta_0+1)})$, together with (A.18) and (A.22), we can drive that

$$E \left\| \tilde{f}_{\text{sc}}(x) - f(x) \right\|^2 \leq O\left(\left(\tilde{t}^2 + 1\right)^{-\delta} \exp\left(-2a\tilde{t}^r\right) + n^{-2/5} \right).$$

$$\text{(A.32)}$$

Thus, we finish the proof of Theorem 6. □

Conflict of Interests

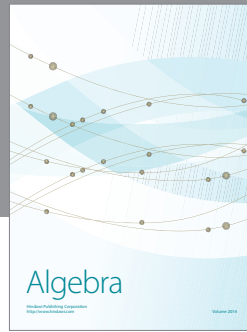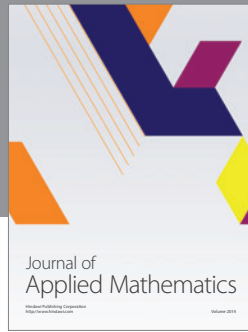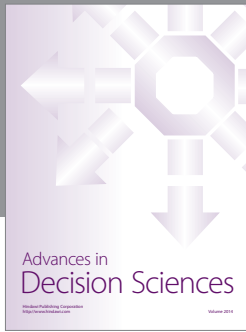The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, UK, 1986.

[2] R. J. Carroll and P. Hall, "Optimal rates of convergence for deconvolving a density," *Journal of the American Statistical Association*, vol. 83, no. 404, pp. 1184–1186, 1988.

[3] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman & Hall, London, UK, 1995.

[4] C. Butucea, "Deconvolution of supersmooth densities with smooth noise," *The Canadian Journal of Statistics*, vol. 32, no. 2, pp. 181–192, 2004.

[5] C. Butucea and A. B. Tsybakov, "Sharp optimality in density deconvolution with dominating bias: I," *Theory of Probability & Its Applications*, vol. 52, no. 1, pp. 111–128, 2007.

[6] B. U. Park, W. C. Kim, and M. C. Jones, "On local likelihood density estimation," *The Annals of Statistics*, vol. 30, no. 5, pp. 1480–1495, 2002.

[7] M. C. Jones and D. A. Henderson, "Kernel-type density estimation on the unit interval," *Biometrika*, vol. 94, no. 4, pp. 977–984, 2007.

[8] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *The Annals of Statistics*, vol. 38, no. 5, pp. 2916–2957, 2010.

[9] J. Fan, "On the optimal rates of convergence for nonparametric deconvolution problems," *The Annals of Statistics*, vol. 19, no. 3, pp. 1257–1272, 1991.

[10] J. Fan, "Deconvolution with supersmooth distributions," *The Canadian Journal of Statistics*, vol. 20, no. 2, pp. 155–169, 1992.

[11] C.-H. Zhang, "Fourier methods for estimating mixing densities and distributions," *The Annals of Statistics*, vol. 18, no. 2, pp. 806–831, 1990.

[12] J.-Y. Koo, "Logspline deconvolution in Besov space," *Scandinavian Journal of Statistics*, vol. 26, no. 1, pp. 73–86, 1999.

[13] M. Pensky and B. Vidakovic, "Adaptive wavelet estimator for nonparametric density deconvolution," *The Annals of Statistics*, vol. 27, no. 6, pp. 2033–2053, 1999.

[14] J. Fan and J. Y. Koo, "Wavelet deconvolution," *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 734–747, 2002.

[15] F. Comte, Y. Rozenholc, and M.-L. Taupin, "Penalized contrast estimator for adaptive density deconvolution," *The Canadian Journal of Statistics*, vol. 34, no. 3, pp. 431–452, 2006.

[16] A. Delaigle and I. Gijbels, "Bootstrap bandwidth selection in kernel density estimation from a contaminated sample," *Annals of the Institute of Statistical Mathematics*, vol. 56, no. 1, pp. 19–47, 2004.

[17] A. Meister, "On the effect of misspecifying the error density in a deconvolution problem," *The Canadian Journal of Statistics*, vol. 32, no. 4, pp. 439–449, 2004.

[18] T. Li and Q. Vuong, "Nonparametric estimation of the measurement error model using multiple indicators," *Journal of Multivariate Analysis*, vol. 65, no. 2, pp. 139–165, 1998.

[19] A. Delaigle, P. Hall, and A. Meister, "On deconvolution with repeated measurements," *The Annals of Statistics*, vol. 36, no. 2, pp. 665–685, 2008.

[20] M. H. Neumann, "Deconvolution from panel data with unknown error distribution," *Journal of Multivariate Analysis*, vol. 98, no. 10, pp. 1955–1968, 2007.

[21] M. H. Neumann, "On the effect of estimating the error density in nonparametric deconvolution," *Journal of Nonparametric Statistics*, vol. 7, no. 4, pp. 307–330, 1997.

[22] G. Kerkyacharian, T. M. Pham Ngoc, and D. Picard, "Localized spherical deconvolution," *The Annals of Statistics*, vol. 39, no. 2, pp. 1042–1068, 2011.

[23] J. Johannes, "Deconvolution with unknown error distribution," *The Annals of Statistics*, vol. 37, no. 5A, pp. 2301–2323, 2009.

[24] F. Comte and C. Lacour, "Data-driven density estimation in the presence of additive noise with unknown distribution," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, vol. 73, no. 4, pp. 601–627, 2011.

[25] G. S. Watson and M. R. Leadbetter, "On the estimation of the probability density, I," *The Annals of Mathematical Statistics*, vol. 34, no. 2, pp. 480–491, 1963.

[26] A. Bernacchia and S. Pigolotti, "Self-consistent method for density estimation," *Journal of the Royal Statistical Society Series B*, vol. 73, no. 3, pp. 407–422, 2011.

[27] S. Efromovich, "Adaptive orthogonal series density estimation for small samples," *Computational Statistics & Data Analysis*, vol. 22, no. 6, pp. 599–617, 1996.

[28] I. K. Glad, N. L. Hjort, and N. G. Ushakov, "Correction of density estimators that are not densities," *Scandinavian Journal of Statistics*, vol. 30, no. 2, pp. 415–427, 2003.

[29] N. I. Fisher, *Statistical Analysis of Circular Data*, Cambridge University Press, New York, NY, USA, 1993.

[30] S. Efromovich, "Density estimation for the case of supersmooth measurement error," *Journal of the American Statistical Association*, vol. 92, no. 438, pp. 526–535, 1997.

[31] H. Meyers and C. A. von Hake, *Earthquake Data File Summary*, National Geophysical and Solar-Terrestrial Data Center, Boulder, Colo, USA, 1976.