Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2020, Vol. 56, No. 1, 36–67
https://doi.org/10.1214/18-AIHP954

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Nonparametric density estimation from observations with multiplicative measurement errors

Denis Belomestny[a,c] and Alexander Goldenshluger[b,c]

[a] *Faculty of Mathematics, Duisburg-Essen University, D-45127 Essen, Germany. E-mail: denis.belomestny@uni-due.de*
[b] *Department of Statistics, University of Haifa, Haifa 31905, Israel. E-mail: goldensh@stat.haifa.ac.il*
[c] *National Research University Higher School of Economics, Moscow*

Received 25 October 2017; revised 28 November 2018; accepted 2 December 2018

Abstract. In this paper we study the problem of pointwise density estimation from observations with multiplicative measurement errors. We elucidate the main feature of this problem: the influence of the estimation point on the estimation accuracy. In particular, we show that, depending on whether this point is separated away from zero or not, there are two different regimes in terms of the rates of convergence of the minimax risk. In both regimes we develop kernel-type density estimators and prove upper bounds on their maximal risk over suitable nonparametric classes of densities. We show that the proposed estimators are rate-optimal by establishing matching lower bounds on the minimax risk. Finally we test our estimation procedures on simulated data.

Résumé. Dans cet article, nous étudions le problème de l'estimation de densité ponctuelle à partir d'observations avec erreurs multiplicatives. Nous clarifions l'élément essentiel de ce problème: l'influence du point d'estimation sur la précision de l'estimation. En particulier, nous montrons que, selon que le point est éloigné de zéro ou pas, il y a deux régimes différents qui s'expriment en termes de la vitesse de convergence d'un risque minimax. Dans les deux régimes, nous développons des estimateurs de type noyau et prouvons des bornes supérieures sur leur risque maximal, ceci sur une classe convenable non paramétrique de densités. Nous montrons que les estimateurs proposés sont d'ordres optimaux en établissant des bornes inférieures correspondantes sur le risque minimax. Enfin, nous testons notre procédé d'estimation sur des données simulées.

MSC: 60G05; 60G20

Keywords: Multiplicative measurement errors; Scale mixtures; The Mellin transform; Multiplicative censoring; Density estimation

1. Introduction

Problem formulation and background. In this paper we study the problem of nonparametric density estimation from observations with multiplicative measurements errors. In particular, assume that we observe a sample $Y_1, \dots, Y_n$ generated by the model

$$Y_i = X_i \eta_i, \quad i = 1, \dots, n, \tag{1.1}$$

where $X_1, \dots, X_n$ are independent identically distributed (i.i.d) random variables with density f_X, and $\eta_1, \dots, \eta_n$ are i.i.d. random variables, independent of $X_1, \dots, X_n$, with known density g. Our goal is to estimate the value of f_X at a single given point x_0 from observations $Y_1, \dots, Y_n$. If f_Y stands for the density of $Y = X\eta$, then

$$\begin{aligned} f_Y(y) = [f_X \star g](y) &:= \int_{-\infty}^{\infty} \frac{1}{x} f_X(y/x) g(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} \frac{1}{x} g(y/x) f_X(x)\,\mathrm{d}x, \quad y \in \mathbb{R}. \end{aligned} \tag{1.2}$$

Thus f_Y is a scale mixture of g, and estimation of f_X from observations $Y_1, \dots, Y_n$ can be viewed as the problem of demixing of a scale mixture.

The outlined estimation problem appears in the literature in various contexts. First, the model (1.1) with normal errors η_i and positive random variables X_i represents a *stochastic volatility model* without drift. In this context estimation of the volatility density f_X from observations $Y_1, \dots, Y_n$ was studied by Van Es et al. [19], Van Es & Speij [18] and Belomestny & Shoenmakers [6].

Second, if (η_i) are uniformly distributed on [0, 1] then the corresponding model (1.1) is referred to as the *multiplicative censoring model*. In this setting Vardi [20] studied the problem of estimating the distribution function of X under the assumption that two samples $Y_1, \dots, Y_n$ and $X_{n+1}, \dots, X_{n+m}$ are available. The aforementioned paper develops a nonparametric maximum likelihood estimator; large sample properties of this estimator are studied in Vardi & Zhang [21]. The problem of density estimation in the multiplicative censoring model was considered in Andersen & Hansen [1] and Comte & Dion [8], where estimators based on orthogonal series have been developed. Kernel density estimators were studied in Asgharian et al. [4] and Brunel et al. [7]. We also refer the reader to the recent work by Belomestny et al. [5] where a generalized multiplicative censoring model with (η_i) being beta-distributed random variables was introduced and studied; see also references therein.

Third, as mentioned above, the outlined problem can be viewed as the problem of demixing of a scale mixture. Closely related problems of estimating mixing densities were considered by Zhang [23,24] and Loh & Zhang [15]. In particular, the paper [23] develops Fourier techniques for estimating mixing densities in location models, while [24] and [15] focus on estimating mixing densities in discrete exponential family models. However we are not aware of works on estimating mixing densities in the context of scale models. Finally, we also mention related results on estimating regression functions with multiplicative errors-in-variables that are reported in Iturria et al. [13].

A naive approach to the problem of density estimation in the model with multiplicative errors is based on reduction to the additive measurement error model. In particular, assuming that X_i's and η_i's are positive random variables and taking logarithms of the both sides of (1.1), we come to the additive model $Y_i' = X_i' + \eta_i'$, where $Y_i' = \ln Y_i$, $X_i' = \ln X_i$ and $\eta_i' = \ln \eta_i$. In this model, the density $f_{X'}$ of X' can be estimated using the well developed methodology for additive deconvolution problems (see, e.g., [23] and [10]), and then an estimator for f_X can be obtained using the inverse transformation $f_X(x) = (1/x) f_{X'}(\ln x)$. This idea has been utilized in Van Es & Spreij [18] and Van Es et al. [19]. However, several questions about applicability of this approach arise. First, it can be used only if X and η are nonnegative random variables. Second, it does not provide an estimator of f_X at the origin $x = 0$ since the inverse transformation is not well-defined there. Third, even if this approach is applicable, it is not clear whether the resulting estimator possesses the desired optimality properties.

In contrast to voluminous literature on density deconvolution in the model with additive measurement errors, the problem of density estimation from observations with multiplicative errors was studied to a much lesser extent. In fact, it was considered only for specific distributions of errors (η_i) such as normal, uniform or beta, and the estimators proposed in the literature are tailored to a specific form of the error density g. In this context the following natural questions arise. How to estimate f_X under general assumptions on the error density g? Which properties of the error density g do affect the estimation accuracy, and what is the achievable accuracy in estimating f_X? What can be said about properties of the deconvolution estimators based on the logarithmic transformation of the data?

The main goal of the present paper is to develop optimal estimators of f_X in a principled way under general assumptions on the error density g and to provide answers to the questions raised above. Our approach makes use of the Mellin transform which, in view of its properties, is an appropriate tool for constructing estimators in this setting.

We adopt minimax framework for measuring estimation accuracy. Specifically, accuracy of an estimator $\hat{f}_X(x_0)$ of $f_X(x_0)$ is measured by the maximal risk

$$\mathcal{R}_n[\hat{f}_X; \Sigma] := \sup_{f_X \in \Sigma} \big[\mathbb{E}_{f_X} \big|\hat{f}_X(x_0) - f_X(x_0)\big|^2\big]^{1/2},$$

where Σ is a class of densities. Here and in what follows, $\mathbb{E}_{f_X}$ denotes the expectation with respect to the distribution of the observations $Y_1, \dots, Y_n$ when the unknown density of X is f_X. The minimax risk is defined by

$$\mathcal{R}_n^*[\Sigma] := \inf_{\hat{f}_X} \mathcal{R}_n[\hat{f}_X; \Sigma] = \inf_{\hat{f}_X} \sup_{f_X \in \Sigma} \big[\mathbb{E}_{f_X} \big|\hat{f}_X(x_0) - f_X(x_0)\big|^2\big]^{1/2},$$

where inf is taken over all possible estimators. Our goal is to develop an estimator $\hat{f}_X(x_0)$ which is *rate-optimal*, i.e.,

$$\mathcal{R}_n[\hat{f}_X; \Sigma] \le C_n \mathcal{R}_n^*[\Sigma], \qquad \sup_n C_n < \infty.$$

Main contributions. The main contributions of this work are as follows.

We elucidate the main feature of the multiplicative measurement errors setting: the influence of the estimation point x_0 on the achievable estimation accuracy. In particular, assuming that unknown density f_X belongs to a local Hölder functional class in a vicinity of x_0, we show that, depending on the value of x_0, there are two different regimes in terms of the rates of convergence of the minimax risk. We develop a general method for estimating $f_X(x_0)$ in these two regimes.

The first regime corresponds to the situation when the value of x_0 is separated away from zero. Here the achievable rate of convergence is primarily determined by the value of x_0, by the local smoothness of f_X, and by the ill-posedness of the integral transform in (1.2). The latter is characterized in terms of the rate at which the Mellin transform of g decreases at infinity on a line parallel to the imaginary axis in the complex plane. It is worth noting that this characteristic is global in the sense that it is determined by the global behavior of the error density g on its support. We construct a kernel-type estimator of $f_X(x_0)$ and prove that it is rate-optimal in terms of dependence on the sample size n, parameters of the considered functional class Σ and x_0. It turns out that the deconvolution estimator based on the logarithmic transformation of the data is a special case of the proposed estimation procedure. As a by-product of our general results, we demonstrate that if x_0 is separated away from zero, the random variables X and η are nonnegative, and f_X belongs to a local Hölder class in a vicinity of x_0, then under certain conditions on g the deconvolution estimator is rate-optimal. However, if f_X satisfies some additional constraints, e.g., a moment condition, then the accuracy of the deconvolution estimator can be improved.

In the second regime, where $x_0 = 0$, completely different phenomena are observed. It turns out that in this case the achievable accuracy in estimating $f_X(0)$ is determined by smoothness of f_X and by local behavior of g in vicinity of the origin. Thus, in contrast to the first regime, the minimax rate depends only on local characteristics of g and is not affected by the ill-posedness of the integral transform in (1.2). In particular, our results imply that if g is bounded and does not vanish in a vicinity of the origin, then the minimax rate of convergence is only by a $\ln n$-factor worse than the one achievable in the problem of density estimation from direct observations. We also construct a rate-optimal estimator of $f_X(0)$ and prove a matching lower bound on the minimax risk.

Organization of the paper. The rest of the paper is organized as follows. In Section 2 we introduce notation, discuss some properties of the Mellin transform that are used throughout the paper and present an identifiability result. Section 3 deals with the setting when x_0 is separated away from zero; we construct estimators under different assumptions on the error density g and present results on their accuracy over suitable classes of densities. Section 4 is devoted to the problem of estimating $f_X(0)$. A simulation study of the proposed estimators is presented in Section 5. Finally, proofs of main results are presented in Section 6 while proofs of auxiliary statements are given in Section 7.

2. Preliminaries

In this section we introduce notation and discuss basic properties of the Mellin transform that will be extensively used throughout the paper. This material can be found, e.g., in [16] and [22]. In addition, we present a result on identifiability of the distribution of X in the model (1.1).

The Mellin transform. For a generic locally integrable function u on $(0,\infty)$ the Mellin transform of u is defined by

$$\widetilde{u}(z) = \mathcal{M}[u; z] := \int_0^\infty x^{z-1} u(x)\,\mathrm{d}x \tag{2.1}$$

for all $z \in \mathbb{C}$ such that the integral on the right hand side is absolutely convergent. The region of convergence Ω_u is an infinite vertical strip in the complex plane $\mathbb{C}$,

$$\Omega_u = \{z \in \mathbb{C} : a < \mathrm{Re}(z) < b\}, \quad a < b,$$

or a vertical line $\Omega_u = \{z : \mathrm{Re}(z) = c\}$ if $u(x)x^{c-1} \in \mathbb{L}_1(\mathbb{R}_+)$ for one $c \in \mathbb{R}$. For example, if $u(x) = O(x^{-a+\epsilon})$ as $x \to 0+$ and $u(x) = O(x^{-b-\epsilon})$ as $x \to \infty$ for some $\epsilon > 0$, then the integral in (2.1) converges absolutely and defines an analytic function $\widetilde{u}(z)$ on $\Omega_u = \{z : a < \mathrm{Re}(z) < b\}$.

The inversion formula for the Mellin transform is

$$u(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-z}\widetilde{u}(z)\,\mathrm{d}z, \quad c \in \Omega_u \cap (-\infty,\infty).$$

Let $u(x)$ and $v(x)$ be functions such that the integral $I = \int_0^\infty u(x)v(x)\,\mathrm{d}x$ exists. Assume also that the Mellin transforms $\widetilde{u}(1-z) = \mathcal{M}[u; 1-z]$ and $\widetilde{v}(z) = \mathcal{M}[v; z]$ have a common strip of analyticity, which will be the case when I is absolutely convergent. Then for any line $\{z : \mathrm{Re}(z) = c\}$ in this common strip the Parseval formula is valid:

$$\int_0^\infty u(x)v(x)\,\mathrm{d}x = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \widetilde{u}(1-z)\widetilde{v}(z)\,\mathrm{d}z.$$

In particular, we get for $u = v$ and $c = \frac{1}{2}$,

$$\int_0^\infty u^2(x)\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\widetilde{u}\left(\frac{1}{2}+i\omega\right)\right|^2 \mathrm{d}\omega.$$

It also holds

$$\int_0^\infty u^2(x)x^{2s-1}\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\widetilde{u}(s+i\omega)\right|^2 \mathrm{d}\omega. \tag{2.2}$$

Let us mention the relation of the Mellin transform to a multiplicative convolution integral (1.2); this property is central in subsequent developments. Let u and v be defined on $[0,\infty)$, and let

$$[u\star v](y) := \int_0^\infty \frac{1}{x}u(x)v(y/x)\,\mathrm{d}x;$$

then

$$\widetilde{[u\star v]}(z) = \mathcal{M}[u\star v; z] = \mathcal{M}[u;z]\mathcal{M}[v;z] = \widetilde{u}(z)\widetilde{v}(z).$$

We shall use the Mellin transform techniques for functions defined on the whole real line. To this end, for a function u on $(-\infty,\infty)$ we set

$$u^+(x) := \begin{cases} u(x), & x\ge 0,\\ 0, & x<0\end{cases} \quad\text{and}\quad u^-(x) := \begin{cases} u(-x), & x> 0,\\ 0, & x\le 0.\end{cases} \tag{2.3}$$

It is evident that with this notation $u(x) = u^+(x)$ for $x \ge 0$ and $u(x) = u^-(-x)$ for $x < 0$. The one-sided Mellin transforms of function u defined on $(-\infty,\infty)$ are given by

$$\widetilde{u}^+(z) = \int_0^\infty x^{z-1}u^+(x)\,\mathrm{d}x = \int_0^\infty x^{z-1}u(x)\,\mathrm{d}x,$$
$$\widetilde{u}^-(z) = \int_0^\infty x^{z-1}u^-(x)\,\mathrm{d}x = \int_{-\infty}^0 (-x)^{z-1}u(x)\,\mathrm{d}x.$$

The Laplace and Fourier transforms. The bilateral Laplace transform of function u on $(-\infty,\infty)$ is defined as

$$\check{u}(z) = \mathcal{L}[u;z] := \int_{-\infty}^{\infty} u(x)e^{-zx}\,\mathrm{d}x,$$

and if the integral absolutely converges on a line $\{z : \mathrm{Re}(z) = c\}$, then the inverse Laplace transform is given by

$$u(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\check{u}(z)e^{zx}\,\mathrm{d}z.$$

The Fourier transform of u is $\widehat{u}(\omega) = \mathcal{F}[u;\omega] := \mathcal{L}[u;i\omega] = \check{u}(i\omega)$.

Identifiability. In the model (1.1) we do not assume that the random variables X and η are nonnegative. This fact raises the question whether the distribution of X is identifiable from the distribution of Y. The next statement provides a necessary and sufficient condition for the identifiability.

Lemma 1. *The probability density f_X is identifiable from f_Y if and only if $g(x) \ne g(-x)$ on a set of positive Lebesgue measure.*

The proof of Lemma 1 is given in Section 7. It shows that the identifiability condition is equivalent to the requirement that $|[\widetilde{g}^+(z)]^2 - [\widetilde{g}^-(z)]^2|$ is not zero for almost all z in the common strip of analyticity of $\widetilde{g}^+$ and $\widetilde{g}^-$. Finally, we note that if one of the variables X or η is nonnegative, then the condition of identifiability is trivially fulfilled.

3. Estimation at a point separated away from zero

In this section we consider the problem of estimation of f_X at a point x_0 separated away from zero.

3.1. *Construction of estimator*

We adopt the linear functional strategy for constructing our estimators. This strategy has been frequently used for solving ill-posed inverse problems (see, e.g., [12] and [2]). In our context, the main idea of this method is to find a pair of kernels, say, $K(x,y)$ and $L(x,y)$ such that:

(i) $\int_{-\infty}^{\infty} K(x,y) f_X(y)\,\mathrm{d}y$ approximates "well" the value $f_X(x)$ to be recovered;
(ii) kernel $L(x,y)$ is related to $K(x,y)$ via the equation

$$\int_{-\infty}^{\infty} K(x,y) f_X(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} L(x,y) f_Y(y)\,\mathrm{d}y. \tag{3.1}$$

Then under (i) and (ii), the empirical estimator of the integral on the right hand side of (3.1) provides a sensible estimator for $f_X(x)$.

Kernel construction. Let $K:\mathbb{R}\to\mathbb{R}$ be a kernel function and for any positive real number h define

$$K_h(x,y) = \begin{cases} \frac{1}{xh} K(\frac{\ln(y/x)}{h}), & y/x > 0, \\ 0, & y/x < 0. \end{cases} \tag{3.2}$$

Let $\widetilde{g}^+(z) = \mathcal{M}[g^+; z]$ and $\widetilde{g}^-(z) = \mathcal{M}[g^-; z]$ be the one-sided Mellin transforms of g, and let

$$\Omega_{g^+} \cap \Omega_{g^-} =: \{z \in \mathbb{C} : a < \mathrm{Re}(z) < b\} \tag{3.3}$$

be the common strip of their analyticity. Since g is a probability density, we always have $a < 1 < b$; hence $\Omega_{g^+}\cap\Omega_{g^-}$ is non-empty – it always contains the line $\{z\in\mathbb{C}:\mathrm{Re}(z)=1\}$. We note that Ω_{g^+} and/or Ω_{g^-} can degenerate to this line. In this case, by convention, we put $a=1$, $b=1$, and corresponding open interval should be replaced by a singleton.

For $s\in(1-b,1-a)$ define

$$L_{s,h}(x,y) := \begin{cases} \frac{1}{2\pi i x}\int_{s-i\infty}^{s+i\infty} |\frac{x}{y}|^z \frac{\check{K}(-zh)\widetilde{g}^+(1-z)}{[\widetilde{g}^+(1-z)]^2-[\widetilde{g}^-(1-z)]^2}\,\mathrm{d}z, & y/x>0, \\ -\frac{1}{2\pi i x}\int_{s-i\infty}^{s+i\infty} |\frac{x}{y}|^z \frac{\check{K}(-zh)\widetilde{g}^-(1-z)}{[\widetilde{g}^+(1-z)]^2-[\widetilde{g}^-(1-z)]^2}\,\mathrm{d}z, & y/x<0. \end{cases} \tag{3.4}$$

For the time being, we suppose that the kernel K and the error density g are such that the function $L_{s,h}$ is well defined; the corresponding conditions on K and g will be formulated later. Several remarks on this definition are in order.

Remark 1. (i) We can assume that the Laplace transform $\check{K}(\cdot)$ of kernel K is an entire function. This does not restrict generality since K can be always chosen to satisfy this assumption.

(ii) If $[\widetilde{g}^+(z)]^2-[\widetilde{g}^-(z)]^2\neq 0$ for all $z\in\Omega_{g^+}\cap\Omega_{g^-}$ then the integrands in (3.4) are analytic functions in $\{z\in\mathbb{C}: 1-b<\mathrm{Re}(z)<1-a\}$. In this case the integrals in (3.4) do not depend on the integration path, and $L_{s,h}(x,y)$ does not depend on $s\in(1-b,1-a)$. If function $[\widetilde{g}^+(z)]^2-[\widetilde{g}^-(z)]^2$ has zeros in $\Omega_{g^+}\cap\Omega_{g^-}$ then the functions under the integral sign in (3.4) are meromorphic, and $L_{s,h}(x,y)$ depends on parameter s.

The relationship between kernels $L_{s,h}(x,y)$ and $K_h(x,y)$ in (3.4) and (3.2) is revealed in the following statement.

Lemma 2. *Let $K_h(x,y)$ be given by* (3.2). *Let $s\in(1-b,1-a)$ where a and b are given in* (3.3), *and suppose that the integrals on the right hand side of* (3.4) *are absolutely convergent. Then it holds that*

$$\int_{-\infty}^{\infty} L_{s,h}(x,y) f_Y(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} K_h(x,t) f_X(t)\,\mathrm{d}t. \tag{3.5}$$

The proof of Lemma 2 is given in Section 7. We note that relationship (3.5) is in full accordance with the linear functional strategy [cf. (3.1)]. Because $a<1<b$, it holds that $0\in(1-b,1-a)$; hence one can always choose $s=0$ in (3.4). This choice yields

$$L_{0,h}(x,y) = \begin{cases} \frac{1}{2\pi x}\int_{-\infty}^{\infty} |\frac{x}{y}|^{i\omega} \frac{\widehat{K}(-\omega h)\widetilde{g}^+(1-i\omega)}{[\widetilde{g}^+(1-i\omega)]^2-[\widetilde{g}^-(1-i\omega)]^2}\,\mathrm{d}\omega, & y/x>0, \\ -\frac{1}{2\pi x}\int_{-\infty}^{\infty} |\frac{x}{y}|^{i\omega} \frac{\widehat{K}(-\omega h)\widetilde{g}^-(1-i\omega)}{[\widetilde{g}^+(1-i\omega)]^2-[\widetilde{g}^-(1-i\omega)]^2}\,\mathrm{d}\omega, & y/x<0. \end{cases}$$

If g is supported on $[0,\infty)$, then $\widetilde{g}^- = 0$, $\widetilde{g}^+ = \widetilde{g}$; in this case

$$L_{s,h}(x,y) = \frac{1}{2\pi i x}\int_{s-i\infty}^{s+i\infty}\left|\frac{x}{y}\right|^z \frac{\check{K}(-zh)}{\widetilde{g}(1-z)}\,\mathrm{d}z,\quad y/x>0, \tag{3.6}$$

and $L_{s,h}(x,y)=0$ whenever $x/y<0$. In particular, for $s=0$ we have

$$L_h(x,y) := L_{0,h}(x,y) = \frac{1}{2\pi x}\int_{-\infty}^{\infty}\left|\frac{x}{y}\right|^{i\omega}\frac{\widehat{K}(-\omega h)}{\widetilde{g}(1-i\omega)}\,\mathrm{d}\omega,\quad y/x>0. \tag{3.7}$$

Estimator. For $|x_0|>0$ we define the estimator of $f_X(x_0)$ by

$$\hat{f}_{s,h}(x_0) = \frac{1}{n}\sum_{j=1}^n L_{s,h}(x_0, Y_j), \tag{3.8}$$

where $L_{s,h}$ is given in (3.4), $h>0$ and $s\in(1-b,1-a)$ are two tuning parameters to be specified. In what follows with a slight abuse of notation we shall write $\hat{f}_h(x_0) := \hat{f}_{0,h}(x_0)$ and $L_h(x,y) := L_{0,h}(x,y)$.

Note also that (3.5) implies

$$\mathbb{E}_{f_X}\big[\hat{f}_{s,h}(x_0)\big] = \int_{-\infty}^{\infty} K_h(x_0,t) f_X(t)\,\mathrm{d}t.$$

The latter formula is crucial for the analysis of the bias of $\hat{f}_{s,h}(x_0)$.

3.2. *Relation to the additive deconvolution problem*

There is close connection between the kernel $L_h(x,y) = L_{0,h}(x,y)$ defined in (3.7) and kernels used in the additive deconvolution problems. Specifically, suppose that X and η are positive random variables, and let $\eta' = \ln\eta$. If g is the density of η, and $\widehat{g}$ is the corresponding characteristic function, then $g_{\eta'}(x) = e^x g(e^x)$ is the density of η', and the characteristic function of η' is $\widehat{g}_{\eta'}(\omega) = \mathcal{F}[g_{\eta'};\omega] = \mathcal{M}[g;1-i\omega] = \widetilde{g}(1-i\omega)$. Therefore the expression for $L_h(x,y)$ in (3.7) can be rewritten as

$$L_h(x,y) = \frac{1}{2\pi x}\int_{-\infty}^{\infty}\frac{\widehat{K}(-\omega h)}{\widehat{g}_{\eta'}(\omega)} e^{-i\omega(\ln y - \ln x)}\,\mathrm{d}\omega,\quad x>0,\ y>0,$$

and the corresponding estimator of $f_X(x_0)$ [cf. (3.8)] is

$$\hat{f}_X(x_0) = \frac{1}{n}\sum_{j=1}^n L_h(x_0,Y_j) = \frac{1}{2\pi x_0 n}\sum_{j=1}^n\int_{-\infty}^{\infty}\frac{\widehat{K}(-\omega h)}{\widehat{g}_{\eta'}(\omega)} e^{-i\omega(\ln Y_j - \ln x_0)}\,\mathrm{d}\omega. \tag{3.9}$$

On the other hand, consider the additive deconvolution model for the logarithms, $Y' = X' + \eta'$, where $Y' = \ln Y$, $X' = \ln X$ and $\eta' = \ln\eta$. Then the standard deconvolution estimator of $f_{X'}(t_0)$ is of the form

$$\hat{f}_{X'}(t_0) = \frac{1}{2\pi n}\sum_{j=1}^n\int_{-\infty}^{\infty}\frac{\widehat{K}(-\omega h)}{\widehat{g}_{\eta'}(\omega)} e^{-i\omega(Y_j' - t_0)}\,\mathrm{d}\omega.$$

Since $f_{X'}(t_0) = e^{t_0} f_X(e^{t_0})$, we can estimate $f_X(x_0) = \frac{1}{x_0} f_{X'}(\ln x_0)$ by

$$\begin{aligned}\hat{f}_X(x_0) &= \frac{1}{x_0}\hat{f}_{X'}(\ln x_0)\\ &= \frac{1}{2\pi x_0 n}\sum_{j=1}^n\int_{-\infty}^{\infty}\frac{\widehat{K}(-\omega h)}{\widehat{g}_{\eta'}(\omega)} e^{-i\omega(Y_j' - \ln x_0)}\,\mathrm{d}\omega,\end{aligned} \tag{3.10}$$

which coincides with (3.9).

We conclude that if random variables X and η are positive, and the parameter s of the estimator $\hat{f}_{s,h}(x_0)$ in (3.8) is set to zero, then both approaches lead to the same estimator. Thus, the estimator (3.10) is a particular case of our estimator $\hat{f}_{s,h}(x_0)$ defined in (3.8). We note however that tuning parameter s adds some flexibility, and its proper choice can improve accuracy of $\hat{f}_{s,h}(x_0)$ under suitable assumptions (see, e.g., Theorem 3 below).

3.3. *Convergence analysis*

We proceed with convergence analysis of the risk of the proposed estimator $\hat{f}_{s,h}(x_0)$. In order to avoid unnecessary technicalities, from now on we will assume that X and η are nonnegative random variables, i.e.,

$$\operatorname{supp}(g) \subseteq [0, \infty), \qquad \Omega_g = \{z \in \mathbb{C} : a < \operatorname{Re}(z) < b\}, \qquad \operatorname{supp}(f_X) \subseteq [0, \infty) \tag{3.11}$$

for some $a > 0$ and $b > a$. Under these conditions the kernel $L_{s,h}(x, y)$ is given by (3.6).

Assumption (3.11) streamlines the presentation and, in fact, does not lead to loss of generality. In particular, the ensuing analysis of the risk of $\hat{f}_{s,h}(x_0)$ remains valid for general random variables X and η, provided that the conditions imposed in the sequel on the Mellin transform $\widetilde{g}$ of g are replaced by the corresponding conditions on $([\widetilde{g}^+]^2 - [\widetilde{g}^-]^2)/\widetilde{g}^+$ and $([\widetilde{g}^+]^2 - [\widetilde{g}^-]^2)/\widetilde{g}^-$ [cf. (3.4)].

The risk of $\hat{f}_{s,h}(x_0)$ will be analyzed under a local smoothness assumption on f_X and two different sets of assumptions on the error density g.

Definition 1. Let $\beta > 0$, $A > 0$, $x_0 > 0$ and $r > 1$. We say that $f \in \mathscr{H}_{x_0,r}(A, \beta)$ if f is a probability density, that is, $\ell = \lfloor \beta \rfloor := \max\{k \in \mathbb{N}_0 : k < \beta\}$ times continuously differentiable, and $\max_{k=1,\dots,\ell} |f^{(k)}(x)| \le A$,

$$\left|f^{(\ell)}(x) - f^{(\ell)}(x')\right| \le A\left|x - x'\right|^{\beta-\ell}, \quad \forall x, x' \in \left[r^{-1}x_0, rx_0\right].$$

As for the conditions on the error density g, some assumptions characterizing the rate of decay of the Mellin transform $\widetilde{g}(\sigma + i\omega)$ as $|\omega| \to \infty$ for a fixed $\sigma \in \Omega_g$ will be considered. Depending on the tail behavior of $\widetilde{g}$, we distinguish between the following two cases:

- *smooth error densities*, when the tails of $\widetilde{g}$ are polynomial, i.e.,

$$\widetilde{g}(\sigma + i\omega) \asymp |\omega|^{-\gamma}, \quad |\omega| \to \infty, \sigma \in \Omega_g$$

- *super-smooth error densities*, when the tails of $\widetilde{g}$ are exponential, i.e.,

$$\widetilde{g}(\sigma + i\omega) \asymp \exp\{-\gamma|\omega|\}, \quad |\omega| \to \infty, \sigma \in \Omega_g.$$

Our terminology here is similar to that used in the additive deconvolution problem, even though the words *smooth* and *super-smooth* should not be understood literally.

3.3.1. *Smooth error densities*

The class of smooth error densities is determined by the following assumption.

[G1] For some $\sigma \in (a, b)$, there exist real numbers $\omega_0 > 0$, $c_0 > 0$, $B_2 > B_1 > 0$ and $\gamma > 0$ such that

$$\begin{aligned} &\min_{|\omega| \le \omega_0} \left|\widetilde{g}(\sigma + i\omega)\right| \ge c_0 > 0, \\ &B_1|\omega|^{-\gamma} \le \left|\widetilde{g}(\sigma + i\omega)\right| \le B_2|\omega|^{-\gamma}, \quad \forall |\omega| \ge \omega_0. \end{aligned} \tag{3.12}$$

We will require Assumption [G1] for a particular choice of $\sigma \in (a, b)$, and parameters c_0, ω_0, B_1, B_2 and γ may depend on σ. Assumption [G1] stipulates the rate of decay of $\widetilde{g}$ on the line $\{z : \operatorname{Re}(z) = \sigma\}$ as $|\operatorname{Im}(z)| \to \infty$ and implies that $\tilde{g}$ does not have zeros on this line. This requirement is similar to standard assumptions in the additive deconvolution problem on the rate of decay of the error characteristic function. The following examples show that [G1] holds for many well-known distributions.

Example 1 (a Beta distribution). Let $g(x) = (\nu + 1)x^\nu/\theta^{\nu+1}$, $0 < x < \theta$ with $\nu > -1$; then

$$\widetilde{g}(z) = (\nu + 1)\theta^{z-1}/(\nu + z), \qquad \operatorname{Re}(z) > -\nu,$$

$a = -\nu$, $b = \infty$, and

$$\left|\widetilde{g}(\sigma + i\omega)\right| = \theta^{\sigma-1}(\nu + 1)\left[(\nu + \sigma)^2 + \omega^2\right]^{-1/2}, \quad \sigma > -\nu.$$

Then Assumption [G1] is verified for any $\sigma > -\nu$ with $\gamma = 1$, $\omega_0 = 2(\sigma + \nu)$, $c_0 = (1/5)^{1/2}\theta^{\sigma-1}(\nu + 1)/(\nu + \sigma)$ and $B_1 = (4/5)^{1/2}\theta^{\sigma-1}(\nu + 1)$, $B_2 = \theta^{\sigma-1}(\nu + 1)$. The case $\nu = 0$, $\theta = 1$ corresponds to the uniform distribution with $\widetilde{g}(z) = 1/z$ and $|\widetilde{g}(\sigma + i\omega)| = (\sigma^2 + \omega^2)^{-1/2}$ for $\sigma > 0$.

Example 2 (Pareto's distribution). Let $g(x) = (\nu - 1)\theta^{\nu-1}/x^{\nu}$, $x > \theta$ with $\theta > 0$ and $\nu > 1$. Then

$$\widetilde{g}(z) = (\nu - 1)\theta^{z-1}/(\nu - z), \qquad \mathrm{Re}(z) < \nu,$$

$a = -\infty$, $b = \nu$, and

$$\left|\widetilde{g}(\sigma + i\omega)\right| = (\nu - 1)\theta^{\sigma-1}\left[(\nu - \sigma)^2 + \omega^2\right]^{-1/2}, \quad \sigma < \nu.$$

Hence Assumption [G1] is verified for any $\sigma < \nu$ with $\gamma = 1$, $\omega_0 = 2(\nu - \sigma)$, $c_0 = (1/5)^{1/2}(\nu - 1)\theta^{\sigma-1}/(\nu - \sigma)$, $B_1 = (4/5)^{1/2}(\nu - 1)\theta^{\sigma-1}$, $B_2 = (\nu - 1)\theta^{\sigma-1}$.

Example 3. Natural examples of random variables whose distributions satisfy Assumption [G1] with $\gamma > 1$ can be obtained by multiplication of independent random variables with densities as in Examples 1 and 2. For instance, the probability density of a random variable which is a product of two independent random variables uniformly distributed on $[0, 1]$ is $g(x) = \ln(1/x)$, $0 \le x \le 1$. For this density $\widetilde{g}(z) = 1/z^2$ and $|\widetilde{g}(\sigma + i\omega)| = (\sigma^2 + \omega^2)^{-1}$, so that Assumption [G1] holds with $\gamma = 2$.

Bounds on the risk. We begin with establishing an upper bound on the risk of the estimator $\hat{f}_{s,h}(x_0)$ under Assumption [G1].

In this case the kernel K is chosen to satisfy the following conditions. Assume that $K : \mathbb{R} \to \mathbb{R}$ is a bounded function that vanishes outside $[-1, 1]$ and satisfies

(i) for a positive integer number m,

$$\int_{-1}^{1} K(t)\,\mathrm{d}t = 1, \qquad \int_{-1}^{1} t^k K(t)\,\mathrm{d}t = 0, \quad k = 1, \dots, m; \tag{3.13}$$

(ii) for a positive integer number q, function K is q times continuously differentiable on $\mathbb{R}$ and for $j = 0, 1, \dots, q$

$$\max_{x\in[-1,1]}\left|K^{(j)}(x)\right| \le C_K < \infty. \tag{3.14}$$

Theorem 1. *Fix some $\beta > 0$, $r > 0$, $A > 0$, $x_0 > 0$ and consider the class $\mathscr{H}_{x_0,r}(A, \beta)$. Suppose that Assumption* [G1] *holds with $\sigma = 1$ and some $\gamma > 1$. Let $\hat{f}_{h_*}(x_0) = \hat{f}_{0,h_*}(x_0)$ be the estimator defined in* (3.7)–(3.8) *and associated with a kernel K satisfying* (3.13)–(3.14) *with parameters $m \ge \lfloor\beta\rfloor + 1$, $q > \gamma + 1$, and*

$$h = h_* := \left[A^2 x_0^2\left(x_0^{\beta} + 1\right)^2 n\right]^{-\frac{1}{2\beta+2\gamma+1}}. \tag{3.15}$$

Then for $h_* < \min\{\ln r, 1\}$ it holds that

$$\mathcal{R}_n\left[\hat{f}_{0,h_*}; \mathscr{H}_{x_0,r}(A, \beta)\right] \le C_1\left[A\left(x_0^{\beta} + 1\right)\right]^{\frac{2\gamma+1}{2\beta+2\gamma+1}}\left(x_0^2 n\right)^{-\frac{\beta}{2\beta+2\gamma+1}}, \tag{3.16}$$

where C_1 depends on β only.

Several remarks on the result of Theorem 1 are in order.

Remark 2. (i) If $\gamma \le 1$, then the result of Theorem 1 holds for a slightly smaller set of functions than $\mathscr{H}_{x_0,r}(A, \beta)$. In particular, if

$$f_X \in \mathscr{H}_{x_0,r}(A, \beta) \cap \left\{f_X : \int_{-\infty}^{\infty}\frac{|\widetilde{f}_X(1 + i\omega)|}{(1 + |\omega|)^{\gamma}}\,\mathrm{d}\omega \le c < \infty\right\}, \tag{3.17}$$

for some $c > 0$, then $\widetilde{f}_Y(1 + i\omega)$ is integrable, and the statement of Theorem 1 is still valid. Note that this additional condition on $\widetilde{f}_X$ is very mild: by the Riemann–Lebesgue lemma $\widetilde{f}_X(1 + i\omega) \to 0$ as $|\omega| \to \infty$.

(ii) The above upper bound critically depends on the value of x_0. If x_0 is separated away from zero by a constant, then for large enough n the bound takes the form

$$\mathcal{R}_n\left[\hat{f}_{h_*}; \mathscr{H}_{x_0,r}(A, \beta)\right] \le C_2 A^{\frac{2\gamma+1}{2\beta+2\gamma+1}}\left(x_0^{2\gamma-1}n^{-1}\right)^{\frac{\beta}{2\beta+2\gamma+1}}. \tag{3.18}$$

In particular, this shows that estimation accuracy gets worse for larger values of x_0.

Now we establish a lower bound on the minimax risk under Assumption [G1]. We require the following additional condition on the error density g.

[G1′] For $\sigma \in (a, b)$ the first derivative of $\widetilde{g}$ satisfies

$$|\widetilde{g}'(\sigma + i\omega)\big| \le B|\omega|^{-\gamma}, \quad \forall|\omega| \ge \omega_0.$$

Assumption [G1′] is similar to standard conditions on derivatives of the characteristic function of the measurement error distribution in the proofs of lower bounds for density deconvolution; cf., e.g., Theorem 5 in [10].

Theorem 2. *Let $x_0 \ge C_3 > 0$ for some constant C_3, and suppose that Assumptions* [G1] *and* [G1′] *hold with $\sigma = 1$ and $\gamma > 1/2$. Then*

$$\liminf_{n\to\infty}\big\{\phi_n^{-1}\mathcal{R}_n^*\big[\mathscr{H}_{x_0,r}(A,\beta)\big]\big\} \ge C_4,$$

where

$$\phi_n := A^{\frac{2\gamma+1}{2\beta+2\gamma+1}}\big(x_0^{2\gamma-1}n^{-1}\big)^{\frac{\beta}{2\beta+2\gamma+1}},$$

and C_4 depends on β and r only.

Remark 3. (i) Note that the lower bound of Theorem 2 coincides with the upper bound (3.18) in terms of its dependence on n, x_0 and A. This implies that for x_0 separated away from zero, the estimator $\hat{f}_{h_*}(x_0)$ is rate-optimal, and dependence of the risk on x_0 over the functional class $\mathscr{H}_{x_0,r}(A,\beta)$ cannot be improved.

(ii) In view of the interpretation of $\hat{f}_{h_*}(x_0)$ given in Section 3.2, Theorems 1 and 2 assert rate-optimality of the standard deconvolution estimator in the additive measurement error model based on the log-transformed data, provided that the bandwidth parameter h_* is selected as in (3.15). Note however that the standard choice of h in additive deconvolution does not involve x_0.

(iii) The proof of the lower bound in Theorem 2 is based on the reduction to a two-point hypotheses testing problem when under the null hypothesis

$$f_X(x) = f_X^{(0)}(x) := \frac{1}{\pi x(1 + \ln^2(x/x_0))}, \quad x > 0.$$

The convergence region of the Mellin transform $\widetilde{f}_X^{(0)}(z)$ of $f_X^{(0)}(x)$ is the line $\{z : \mathrm{Re}(z) = 1\}$, and this fact is essential for the result of Theorem 2. If the Mellin transform is analytic in a non-degenerating strip around $\{z : \mathrm{Re}(z) = 1\}$ then, under certain assumptions on measurement error density g, the estimation accuracy can be improved in terms of dependence on x_0. This issue is a subject of the next paragraph.

Choice of parameter s and improvements. It is important to realize the interplay between conditions on g and f_X that lead to the results of Theorems 1 and 2. In particular, the following two facts are essential for the stated results.

(a) Since f_X is a probability density, the Mellin transform $\widetilde{f}_X(z)$ always exists on the vertical line $\{z : \mathrm{Re}(z) = 1\}$. Note however that the local smoothness assumption $f_X \in \mathscr{H}_{x_0,r}(A,\beta)$ is not sufficient in order to guarantee the existence of $\widetilde{f}_X(z)$ outside this line in the complex plane.
(b) The premise of Theorems 1 and 2 stipulates behavior of $\widetilde{g}$ on the line $\{z : \mathrm{Re}(z) = 1\}$ only; in particular, $\widetilde{g}(z)$ does not vanish on this line.

Under (a) and (b) the only possible choice of parameter s is $s = 0$, and as pointed out in Remark 3(ii), the form of the corresponding estimator $\hat{f}_{s,h}(x_0)$ coincides with that of the deconvolution estimator in the additive model based on the log-transformed data.

As discussed in Remark 3(iii), the facts (a) and (b) are essential for the proof of the lower bound of Theorem 2, which is achieved on a least favorable two-point testing problem for alternatives $f_X^{(0)}$ and $f_X^{(1)}$ satisfying

$$\int_0^\infty f_X^{(i)}(x)x^{2\alpha}\,\mathrm{d}x = \infty, \quad i = 0, 1, \forall\alpha \ne 0.$$

It turns out, however, that if $\widetilde{f}_X(z)$ is analytic in a strip around $\{z : \mathrm{Re}(z) = 1\}$ then the upper bound of Theorem 1 can be improved in terms of dependence on x_0. As we demonstrate below, this improvement is achieved by the choice of parameter s.

Let $\alpha > 0$, $M > 0$, and consider the functional class

$$\mathscr{F}_{\alpha,M}(A,\beta) := \mathscr{H}_{x_0,r}(A,\beta) \cap \left\{ f : \int_0^\infty x^{2\alpha} f(x)\,\mathrm{d}x \le M \right\}.$$

Note that for $f_X \in \mathscr{F}_{\alpha,M}(A,\beta)$ it holds that

$$\Omega_{f_X} \supset \{z \in \mathbb{C} : 0 \le \mathrm{Re}(z) \le 2\alpha + 1\}.$$

The following statement holds.

Theorem 3. *For arbitrarily small $\epsilon > 0$, let*

$$s_* := \max\left\{ -\alpha, \frac{1}{2}(1-b) + \epsilon \right\}. \tag{3.19}$$

Suppose that Assumption [G1] *holds with $\sigma = 1 - s_*$ and $\gamma > 1$. Let $\hat{f}_{s_*,h_*}(x_0)$ be the estimator associated with kernel K as in Theorem 1 and*

$$s = s_*, \qquad h = h_* := C_5\big[M^{-1}A^2 x_0^{-2s_*+2}\big(x_0^\beta + 1\big)^2 n\big]^{-\frac{1}{2\beta+2\gamma+1}}.$$

If n is large enough so that $h_* < \min\{\ln r, 1\}$, then

$$\mathcal{R}_n\big[\hat{f}_{s_*,h_*}; \mathscr{F}_{\alpha,M}(A,\beta)\big] \le C_6\big[A\big(x_0^\beta + 1\big)\big]^{\frac{2\gamma+1}{2\beta+2\gamma+1}} \big(M x_0^{2s_*-2} n^{-1}\big)^{\frac{\beta}{2\beta+2\gamma+1}}, \tag{3.20}$$

where C_6 depends on β only.

Remark 4. (i) If $\gamma \le 1$ then the result of Theorem 3 holds for a slightly smaller set of functions than $\mathscr{H}_{x_0,r}(A,\beta)$, as discussed in Remark 2(i).

(ii) For x_0 separated away from zero by a constant, the upper bound (3.20) takes the form

$$\mathcal{R}_n\big[\hat{f}_{s_*,h_*}; \mathscr{F}_{\alpha,M}(A,\beta)\big] \le C_8 A^{\frac{2\gamma+1}{2\beta+2\gamma+1}} \big(M x_0^{2\gamma-1+2s_*} n^{-1}\big)^{\frac{\beta}{2\beta+2\gamma+1}}. \tag{3.21}$$

Because $s_* \le 0$, this bound is better than (3.18) in terms of its dependence on x_0, provided $x_0 > 1$. For instance, let η be uniformly distributed random variable on $[0, 1]$; then $\gamma = 1$, $a = 0$ and $b = \infty$. If f_X has bounded second moment, i.e., $f_X \in \mathscr{F}_{1,M}(A,\beta)$, and the condition in (3.17) holds, then in view of (3.19) the best choice of s is $s = s_* = -1$, and the right hand side of (3.21) is proportional to $x_0^{-\beta/(2\beta+3)}$. Thus, the accuracy improves for large x_0. This fact is in contrast to the result of Theorem 1 stated for the functional class $\mathscr{H}_{x_0,r}(A,\beta)$.

3.3.2. *Super-smooth error densities*

Now we turn to the convergence analysis of the risk of $\hat{f}_{s,h}(x_0)$ in the case of super-smooth error densities characterized by the following assumption.

[G2] For some $\sigma \in (a, b)$, there exist constants $c_0 > 0$, $\omega_0 > 0$, $\gamma > 0$, $\nu \in \mathbb{R}$, $B_2 \ge B_1 > 0$ such that

$$\begin{aligned} &\min_{|\omega| \le \omega_0} \big|\widetilde{g}(\sigma + i\omega)\big| \ge c_0 > 0, \\ &B_1|\omega|^\nu e^{-\gamma|\omega|} \le \big|\widetilde{g}(\sigma + i\omega)\big| \le B_2|\omega|^\nu e^{-\gamma|\omega|}, \quad \forall |\omega| \ge \omega_0. \end{aligned} \tag{3.22}$$

The probability densities on $[0,\infty)$ with exponential tails are the prototypes of densities satisfying Assumption [G2].

Example 4 (Gamma distribution). Let $g(x) = \mu^\alpha x^{\alpha-1} e^{-\mu x}/\Gamma(\alpha)$, $\alpha > 0$, $\mu > 0$, $x > 0$; then

$$\widetilde{g}(z) = \mu^{-z+1}\Gamma(z + \alpha - 1)/\Gamma(\alpha), \qquad \mathrm{Re}(z) > -\alpha + 1.$$

As a result $a=-\alpha+1$, $b=\infty$. Furthermore, it is well known [3, Corollary 1.4.4] that for any $\sigma\geq-2$, there exist positive constants C and C' such that uniformly for $|\omega|\geq 2$,

$$C|\omega|^{\sigma-1/2}e^{-|\omega|\pi/2}\leq\left|\Gamma(\sigma+i\omega)\right|\leq C'|\omega|^{\sigma-1/2}e^{-|\omega|\pi/2}. \tag{3.23}$$

Thus, (3.22) is verified for large enough ω_0 with some $c_0=c_0(\omega_0)>0$, $\nu=\sigma+\alpha-3/2$ and $\gamma=\pi/2$.

Example 5 (Half-normal distribution). Let $g(x)=\sqrt{2/\pi}(1/\upsilon)\exp\{-x^2/(2\upsilon^2)\}$ with $\upsilon>0$. As can be easily seen, $g(x)$ is a probability density on $\mathbb{R}_+$ and it holds

$$\widetilde{g}(z)=\pi^{-1/2}(\sqrt{2}\upsilon)^{z-1}\Gamma(z/2).$$

In view of (3.23), Assumption [G2] holds for large enough ω_0 with $\nu=(\sigma-1)/2$ and $\gamma=\pi/4$.

Estimator and bounds on the risk. Now we analyze the accuracy of $\hat{f}_{s,h}(x_0)$ under Assumption [G2]. In this case the kernel K is to be constructed in a different way. Specifically, let $\lambda\geq 2$ be a fixed natural number, and let w be a function defined via its Fourier transform,

$$\widehat{w}(\omega)=\exp\left\{-|\omega|^{2\lambda}/2\lambda\right\}. \tag{3.24}$$

Note that $\int_{-\infty}^{\infty}w(x)\,\mathrm{d}x=1$. For a positive integer number m let

$$K(t)=\sum_{j=1}^{m+1}\binom{m+1}{j}(-1)^{j+1}\frac{1}{j}w\left(\frac{t}{j}\right). \tag{3.25}$$

It is well-known that (3.25) defines kernel K satisfying condition (3.13) (see, e.g., [14]). Although functions w and K depend on the parameter λ, for the sake of brevity we shall not indicate this in our notation. For $h>0$, let $K_h(x,y)$ and $L_{s,h}(x,y)$ be defined by (3.2) and (3.6), respectively. Consider the corresponding estimator

$$\hat{f}_{s,h}(x_0)=\frac{1}{n}\sum_{j=1}^{n}L_{s,h}(x_0,Y_j).$$

Theorem 4. *Suppose that Assumption* [G2] *holds with* $\sigma=1$. *Let* $x_0>0$, *and let* $\hat{f}_{h_*}(x_0)=\hat{f}_{0,h_*}(x_0)$ *be the estimator associated with kernel* K *given in* (3.24) *and* (3.25) *with parameters*

$$m\geq\lfloor\beta\rfloor+1,\qquad h_*=C_1\gamma\left[\ln\left(A^2x_0^{2\beta+2}n\right)\right]^{-1+\frac{1}{2\lambda}}.$$

Then

$$\limsup_{n\to\infty}\left\{\varphi_n^{-1}\mathcal{R}_n\left[\hat{f}_{h_*};\mathscr{H}_{x_0,r}(A,\beta)\right]\right\}\leq C_2, \tag{3.26}$$

where $\varphi_n=A\gamma^{\beta}(\ln n)^{-\beta(1-\frac{1}{2\lambda})}x_0^{\beta}$, *and* $C_2=C_2(\beta,\lambda)$ *depends on* λ *and* β.

Remark 5. Theorem 4 shows that for any fixed $\lambda\geq 2$, the maximal risk of $\hat{f}_{h_*}$ converges to zero at the rate $O((\ln n)^{-\beta(1-(1/2\lambda))})$ as $n\to\infty$. It may seem advantageous to let $\lambda\to\infty$ as $n\to\infty$. However, the constant $C_2(\beta,\lambda)$ on the right hand side of (3.26) explodes as $\lambda\to\infty$.

A simple modification of the proof of Theorem 2 shows that under Assumption [G2] and under suitable condition on the derivative $\widetilde{g}'(1+i\omega)$ (similar to Assumption [G1′]) one has

$$\liminf_{n\to\infty}\left\{\phi_n^{-1}\mathcal{R}_n^*\left[\mathscr{H}_{x_0,r}(A,\beta)\right]\right\}\geq C_3,\quad \phi_n:=A\gamma^{\beta}x_0^{\beta}(\ln n)^{-\beta},$$

where C_3 depends on β only. Thus the estimator $\hat{f}_{h_*}$ can be regarded as nearly rate-optimal. It is worth noting that the result of Theorem 4 remains valid for the class $\mathscr{F}_{\alpha,M}(A,\beta)$, and the choice of the parameter $s\neq 0$ does not lead to improvements in the rate of convergence in terms of its dependence on x_0.

4. Estimation at zero

Now we turn to the problem of estimating $f_X(0)$ in the model (1.1). The following modification of the definition of $\mathscr{H}_{x_0,r}(A,\beta)$ will be considered.

Definition 2. Let $\beta > 0$, $A > 0$ and $r > 0$. We say that $f \in \mathscr{H}_r(A,\beta)$, if f is $\ell = \lfloor \beta \rfloor := \max\{k \in \mathbb{N}_0 : k < \beta\}$ times continuously differentiable on $(0,r]$ and $\max_{k=1,\dots,\ell} |f^{(k)}(x)| \le A$,

$$\left|f^{(\ell)}(x) - f^{(\ell)}(x')\right| \le A\left|x - x'\right|^{\beta-\ell}, \quad \forall x, x' \in (0,r].$$

We define also

$$\bar{\mathscr{H}}_r(A,\beta,M) := \mathscr{H}_r(A,\beta) \cap \left\{f : \sup_{t\in\mathbb{R}_+}\left|f(t)\right| \le M\right\}. \tag{4.1}$$

First we note that if $I_g := \int_0^\infty [g(x)/x]\,\mathrm{d}x < \infty$, i.e., if $\{z : \mathrm{Re}(z) = 0\} \subseteq \Omega_g$, then f_Y is finite at the origin, and in view of (1.2) $f_Y(0) = f_X(0)I_g$. In this case a natural estimator of f_X can be defined as $\hat{f}_X(0) = \hat{f}_Y(0)/I_g$, where $\hat{f}_Y(0)$ is a suitable estimator of $f_Y(0)$, say, a kernel-type estimator with bandwidth h, from direct observations $Y_1,\dots,Y_n$. As a result, under the choice $h \asymp n^{-1/(2\beta+1)}$ (see e.g. Theorem 1.1 in [17]), we get

$$\mathcal{R}_n\left[\hat{f}_h; \mathscr{H}_r(A,\beta)\right] \le O\left(n^{-\beta/(2\beta+1)}\right).$$

It is also clear that this rate is minimax over the class $\mathscr{H}_r(A,\beta)$. Note, however, that the condition $\{z : \mathrm{Re}(z) = 0\} \subseteq \Omega_g$ is too restrictive and does not hold in many situations of interest. For instance, it does not hold for the uniform distribution on $[0,1]$. Thus, in the case when $\{z : \mathrm{Re}(z) = 0\}$ is not a subset of Ω_g, we need to propose an alternative method of estimating $f_X(0)$.

4.1. *Kernel construction and estimator*

In order to construct an estimator of f at zero, we use the following kernel. For a fixed real number $s \ge 0$, consider the function

$$\psi_s(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-s)^2} x^{-s} \exp\left\{-\frac{1}{2}[\ln x]^2\right\}, \quad x \ge 0. \tag{4.2}$$

It is easily checked that $\int_0^\infty \psi_s(x)\,\mathrm{d}x = 1$ and $\widetilde{\psi}_s(s+i\omega) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-s)^2} e^{-\frac{1}{2}|\omega|^2}$. Fix positive integer number m, and define the kernel

$$K_s(x) = \sum_{j=1}^{m+1} \binom{m+1}{j} (-1)^{j+1} \frac{1}{j} \psi_s\left(\frac{x}{j}\right), \quad x \ge 0. \tag{4.3}$$

By construction, K_s satisfies condition (3.13). Another attractive property of the kernel K is that the Mellin transform $\widetilde{K}_s(z)$ decreases at the rate $e^{-\frac{1}{2}|\omega|^2}$ as $|\omega| \to \infty$ along the line $\{z : \mathrm{Re}(z) = s\}$ [see the proof of Theorem 5].

Having defined the function K_s, let us consider its scaled version, $K_{s,h}(x) := (1/h)K_s(x/h)$ for $h > 0$, and note that

$$\widetilde{K}_{s,h}(z) = \int_0^\infty t^{z-1} K_{s,h}(t)\,\mathrm{d}t = h^{z-1}\widetilde{K}_s(z).$$

According to the linear functional strategy, the kernel $L_{s,h}(y)$ corresponding to $K_{s,h}(x)$ is given by

$$\begin{aligned} L_{s,h}(y) &:= \frac{1}{2\pi i}\int_{s-i\infty}^{s+i\infty} \frac{\widetilde{K}_{s,h}(z)}{\widetilde{g}(1-z)} y^{-z}\,\mathrm{d}z \\ &= \frac{1}{2\pi h^{1-s} y^s}\int_{-\infty}^{\infty} \left(\frac{h}{y}\right)^{i\omega} \frac{\widetilde{K}_s(s+i\omega)}{\widetilde{g}(1-s-i\omega)}\,\mathrm{d}\omega, \end{aligned} \tag{4.4}$$

provided that the expression on the right hand side is well defined.

Consider now the following estimator

$$\hat{f}_{s,h}(0)=\frac{1}{n}\sum_{i=1}^{n}L_{s,h}(Y_i). \tag{4.5}$$

The tuning parameters s and h will be specified below in Theorem 5.

4.2. *Bounds on the risk*

First we establish an upper bound on the maximal risk of the estimator $\hat{f}_{s,h}(0)$. It is done under the following assumptions on the error density g.

[G3] For some $p\in[0,1)$, $q\geq 0$ and $\delta\in(0,1)$

$$c_0x^{-p}\big[\ln(1/x)\big]^q\leq g(x)\leq C_0x^{-p}\big[\ln(1/x)\big]^q,\quad x\in(0,\delta). \tag{4.6}$$

Assumption [G3] prescribes behavior of the density g in a vicinity of the origin. If $p<0$ then the integral $\int_0^\infty[g(x)/x]\,\mathrm{d}x$ is finite, and, as discussed above, the problem reduces to the density estimation from direct observations. Moreover, since g is a probability density, it must hold $p<1$. That is why in [G3] we restrict our attention to the case $p\in[0,1)$. Note also that [G3] implies that $\widetilde{g}$ is well defined in the strip $\{z:p<\mathrm{Re}(z)\leq 1\}$, i.e., $\Omega_g\supseteq\{z:p<\mathrm{Re}(z)\leq 1\}$.

In addition to Assumption [G3], we impose some mild conditions on g that guarantee existence of the estimator $\hat{f}_{s,h}(0)$ under the following specific choice of the parameter s,

$$s_*:=\frac{1}{2}(1-p); \tag{4.7}$$

here p is the parameter appearing in Assumption [G3].

[G4] Suppose that $|\widetilde{g}(1-s_*+i\omega)|>0$ for all $\omega\in\mathbb{R}$, and

$$\int_{-\infty}^{\infty}\frac{e^{-\omega^2/2}}{|\widetilde{g}(1-s_*+i\omega)|}\,\mathrm{d}\omega\vee\int_{-\infty}^{\infty}\frac{e^{-\omega^2}}{|\widetilde{g}(1-s_*+i\omega)|^2}\,\mathrm{d}\omega\leq C_1<\infty. \tag{4.8}$$

In addition,

$$\int_{-\infty}^{\infty}\left|\frac{\mathrm{d}^l}{\mathrm{d}\omega^l}\left(\frac{e^{-\omega^2/2}}{\widetilde{g}(1-s_*+i\omega)}\right)\right|^2\mathrm{d}\omega\leq C_2<\infty, \tag{4.9}$$

where $l:=\lceil(q+1)/2\rceil$, and q appears in (4.6).

The conditions of Assumption [G4] are rather mild. First we note that under Assumption [G3] the line $\{z:\mathrm{Re}(z)=1-s_*=\frac{1}{2}(1+p)\}$ belongs to the convergence region of $\widetilde{g}$. The first condition in [G4] bounds from below the rate of decay of $\widetilde{g}$ along this line. It ensures that under the choice $s=s_*$ the integrand in (4.4) is absolutely integrable and square integrable; thus the estimator $\hat{f}_{s_*,h}(0)$ in (4.5) is well defined [see the proof of Theorem 5 for details]. The second condition of [G4] is stated for the derivatives of the integrand in (4.4) and is used to bound the variance of $\hat{f}_{s_*,h}(0)$. Note that (4.8) holds both for the smooth and super-smooth error densities.

We are now in a position to state an upper bound on the risk of the estimator $\hat{f}_{s_*,h}(0)$ under a suitable choice of the bandwidth h.

Theorem 5. *Fix some positive real numbers A, β, M and consider the class of functions $\bar{\mathscr{H}}_r(A,\beta,M)$ defined in (4.1). Let Assumptions [G3] and [G4] hold, and let $\hat{f}_*(0)=\hat{f}_{s_*,h_*}(0)$ denote the estimator (4.5) associated with parameters $m\geq\lfloor\beta\rfloor+1$, $s=s_*$ given by (4.7) and*

$$h=h_*:=\big[MA^{-2}(\ln n)^{q+\varkappa}n^{-1}\big]^{\frac{1}{2\beta+1+p}},\qquad \varkappa:=\begin{cases}0, & p\in(0,1),\\ 1, & p=0.\end{cases} \tag{4.10}$$

Then for n large enough such that $h_*<\min\{r,1\}$ one has

$$\mathcal{R}_n\big[\hat{f}_*;\bar{\mathscr{H}}_r(A,\beta,M)\big]\leq C_3A^{\frac{1+p}{2\beta+1+p}}\big[M(\ln n)^{q+\varkappa}n^{-1}\big]^{\frac{\beta}{2\beta+1+p}},$$

where C_3 may depend on β only.

Remark 6. (i) Note that the upper bound of Theorem 5 holds both for smooth and super-smooth error densities, provided that the mild conditions of Assumption [G4] are fulfilled. This is in contrast to the results on estimating density f_X at a point separated away from zero.

(ii) It is instructive to consider particular cases corresponding to different error densities. For instance, if g is the uniform density on [0, 1], or an exponential density then $p = 0$, $q = 0$ and $\varkappa = 1$. So in these cases the upper bound is of the order $(\ln n/n)^{\beta/(2\beta+1)}$ which is only by a logarithmic factor worse than the standard nonparametric rate.

Our next result is the lower bound on the minimax risk. To that end, we introduce the following condition on g.

[G5] Suppose that $\{z \in \mathbb{C} : 1 \leq \mathrm{Re}(z) \leq 1 + \epsilon\} \subset \Omega_g$ for some $\epsilon > 0$, and

$$\left|\widetilde{g}(1 + \epsilon + i\omega)\right| \leq C_4 < \infty, \quad \forall\omega. \tag{4.11}$$

Assumption [G5] is rather mild; it holds if $\int_0^\infty x^\epsilon g(x)\,\mathrm{d}x \leq C_4$ for some $\epsilon > 0$. Note also that [G5] together with [G3] imply that $\widetilde{g}$ is analytic in the strip $\{z : p < \mathrm{Re}(z) \leq 1 + \epsilon\}$.

Theorem 6. *Let Assumptions* [G3] *and* [G5] *hold, then for the functional class* $\bar{\mathscr{H}}_r(A, \beta, M)$ *with* $M \geq 1$ *one has*

$$\liminf_{n\to\infty}\left\{\phi_n^{-1}\mathcal{R}_n^*\left[\bar{\mathscr{H}}_r(A, \beta, M)\right]\right\} \geq C_5,$$

where

$$\phi_n := A^{\frac{p+1}{2\beta+1+p}}\left[M^{1-p}(\ln n)^{q+\varkappa}n^{-1}\right]^{\frac{\beta}{2\beta+1+p}},$$

and C_5 *depends on* β *only.*

The lower bound on the minimax risk of Theorem 6 matches the bound of Theorem 5 up to a minor discrepancy in terms of dependence on M. Note, however, that in the practically important case of $p = 0$ the bounds coincide. Thus the estimator $\hat{f}_*(0)$ is rate-optimal on the class $\bar{\mathscr{H}}_r(A, \beta, M)$.

5. Numerical experiments

In this section we demonstrate that in many cases of interest the developed estimators are given by analytic formulas and can be easily implemented. We also illustrate numerically theoretical results on performance of the estimators.

5.1. *Estimation outside zero*

First we study numerically the accuracy of the estimator (3.8) for points separated away from zero. Assume that errors (η_i) are beta-distributed with the density

$$g(x) = \nu x^{\nu-1}, \quad 0 \leq x \leq 1, \nu > 0, \tag{5.1}$$

then

$$\widetilde{g}(z) = \nu\int_0^1 x^{\nu-1}x^{z-1}\,\mathrm{d}x = \nu/(\nu + z - 1). \tag{5.2}$$

Furthermore, consider the case of exponentially distributed X, that is, $f_X(x) = e^{-x}$ for $x > 0$. Let $w(x) = e^{-x^2/2}/\sqrt{2\pi}$, and for a fixed natural number m let

$$K(t) = \sum_{j=1}^{m+1}\binom{m+1}{j}(-1)^{j+1}\frac{1}{j}w\left(\frac{t}{j}\right). \tag{5.3}$$

The bilateral Laplace transform of K is defined for any $z \in \mathbb{C}$ and given by

$$\begin{aligned}\check{K}(z) &= \sum_{j=1}^{m+1} \binom{m+1}{j} (-1)^{j+1} \frac{1}{\sqrt{2\pi}\, j} \int_{-\infty}^{\infty} e^{-t^2/(2j^2)-tz}\, \mathrm{d}t \\ &= \sum_{j=1}^{m+1} \binom{m+1}{j} (-1)^{j+1} e^{j^2 z^2/2} = \check{K}(-z).\end{aligned}$$

Let us now compute the kernel $L_{s,h}(x, y)$,

$$L_{s,h}(x,y) := \frac{1}{2\pi i x} \int_{s-i\infty}^{s+i\infty} \left(\frac{x}{y}\right)^z \frac{\check{K}(zh)}{\widetilde{g}(1-z)}\, \mathrm{d}z.$$

Using (5.2), we obtain

$$\begin{aligned}&\frac{1}{2\pi i x} \int_{s-i\infty}^{s+i\infty} \left(\frac{x}{y}\right)^z \frac{e^{j^2h^2z^2/2}}{\widetilde{g}(1-z)}\, \mathrm{d}z \\ &\quad= \frac{1}{2\pi \nu x^{1-s} y^s} \int_{-\infty}^{\infty} e^{iu \ln(x/y)} (\nu - s - iu) e^{j^2h^2(s+iu)^2/2}\, \mathrm{d}u \\ &\quad= \frac{1}{\sqrt{2\pi} x^{1-s} y^s} \exp\left\{\frac{j^2 s^2 h^2}{2} - \frac{1}{2j^2h^2}\left[j^2 s h^2 + \ln(x/y)\right]^2\right\} \\ &\qquad\times \left\{\frac{\nu - s}{(j^2h^2)^{1/2}} + \frac{j^2 s h^2 + \ln(x/y)}{(j^2h^2)^{3/2}}\right\}.\end{aligned}$$

Thus

$$L_{s,h}(x,y) = \frac{1}{\sqrt{2\pi}} \sum_{j=1}^{m+1} \binom{m+1}{j} (-1)^{j+1} \exp\left\{-\frac{\ln^2(x/y)}{2j^2h^2}\right\} \frac{1}{xjh}\left[\nu + \frac{\ln(x/y)}{j^2h^2}\right].$$

Note that the kernel does not depend on s and this corresponds to the fact that the function $\check{K}(zh)/\widetilde{g}(1-z)$ is holomorphic.

In Figure 1 we present box plots of the quantity $|\hat{f}_{h_\star}(x) - f_X(x)|$ for different sample sizes n and different points $x > 0$ over 200 simulation runs, where in each run we construct the estimate $\hat{f}_{h_\star}(x)$ associated with the above kernel $L_{s,h}$ and a precomputed bandwidth $h_\star$. The latter is found by minimizing $\mathbb{E}_N[|\hat{f}_h(x) - f_X(x)|^2]$ over h with the empirical expectation $\mathbb{E}_N$ computed using $N = 300$ independent simulation runs. The left graph in Figure 1 demonstrates convergence of the estimation error for $x_0 = 1$ as the sample sample grows, while the right graph shows dependence of the error for a given sample size $n = 500$ on x_0. As can be seen the error decreases as x_0 grows, which is in accordance with the results of Theorem 3.

5.2. *Estimation at zero*

Now we illustrate behavior of the developed estimator for the case $x_0 = 0$. We consider again beta-distributed errors as in (5.1) and (5.2). Let $w(x) = e^{-x}$, and let K be given by (5.3). Using the fact that $\widetilde{w}(z) = \Gamma(z)$, we have for any $s > \max(0, \nu)$

$$\begin{aligned}\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega y} \frac{\widetilde{w}(s+i\omega)}{\widetilde{g}(1-s-i\omega)}\, \mathrm{d}\omega &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega y} \Gamma(s+i\omega)\left(1 - \frac{s+i\omega}{\nu}\right) \mathrm{d}\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega y} \Gamma(s+i\omega)\, \mathrm{d}\omega - \frac{1}{2\pi\nu} \int_{-\infty}^{\infty} e^{-i\omega y} \Gamma(1+s+i\omega)\, \mathrm{d}\omega.\end{aligned}$$

The well-known identity

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega y} \Gamma(s+i\omega)\, \mathrm{d}\omega = e^{sy} \exp\left(-e^y\right), \qquad y \in \mathbb{R}$$

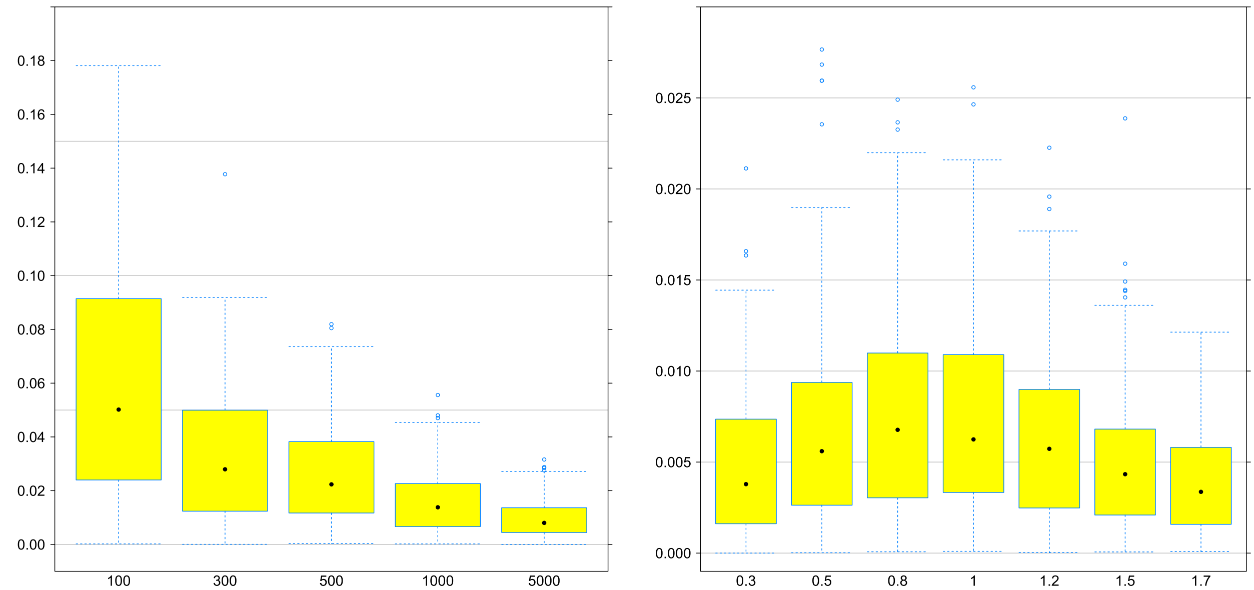

Fig. 1. Left: boxplots of the distance $|\hat{f}_{h_\star}(1) - f_X(1)|$, where the estimate $\hat{f}_{h_\star}(1)$ is based on $n \in \{100, 300, 500, 1000, 5000\}$ observations of the r.v. Y under uniformly distributed errors. Right: boxplots of the distance $|\hat{f}_{h_\star}(x) - f_X(x)|$ for $x \in \{0.3, 0.5, 0.8, 1.0, 1.2, 1.5, 1.7\}$, where the estimate $\hat{f}_{h_\star}(x)$ is based on $n = 500$ observations of the r.v. Y under uniformly distributed errors. The bandwidth $h_\star$ is precomputed using 300 independent runs.

leads to

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega y}\left(1 - \frac{s + i\omega}{\nu}\right)\Gamma(s + i\omega)\,\mathrm{d}\omega = e^{sy}\exp\bigl(-e^{y}\bigr)\left(1 - \frac{e^{y}}{\nu}\right).$$

Then using (4.4) and a straightforward algebra, we obtain

$$L_{s,h}(y) = \sum_{j=1}^{m+1}\binom{m+1}{j}(-1)^{j+1}\frac{1}{jh}\exp\left\{-\frac{y}{jh}\right\}\left(1 - \frac{y}{jh\nu}\right).$$

The corresponding estimator is $\hat{f}_h(0) := \frac{1}{n}\sum_{i=1}^{n} L_{s,h}(Y_i)$.

In our simulation study we take $f_X(x) = 2\exp(-2x)$ so that $f_X(0) = 2$ and the distribution of η as in (5.1) with $\nu \in \{1, \frac{1}{2}\}$. In Figure 2 we present box plots of the quantity $|\hat{f}_h(0) - f_X(0)|$ over 200 simulation runs, where in each run we construct the estimate $\hat{f}_{h_\star}(0)$ using a precomputed bandwidth $h_\star$. The latter is found by minimizing $\mathbb{E}_N[|\hat{f}_{h_\star}(0) - f_X(0)|^2]$ over h with empirical expectation $\mathbb{E}_N$ computed using $N = 300$ independent simulation runs. As expected, in the case $\nu = 1$ the estimator is more accurate than in the case $\nu = 1/2$.

6. Proofs of main results

In the proofs below $c_0, c_1, c_2, \dots$ denote positive constants depending on the parameters appearing in Assumptions [G1]–[G5] and on β only unless specified otherwise.

6.1. *Proof of Theorem 1*

Note that under Assumption [G1] condition (3.14) with $q > \gamma + 1$ guarantees that the estimator $\hat{f}_h(x_0) = \hat{f}_{0,h}(x_0)$ is well-defined. Indeed, under this condition $\widehat{K}(-\cdot h)/\widetilde{g}(1 - i\cdot) \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$.

1^0. The next statement establishes an upper bound on the bias of $\hat{f}_{s,h}(x_0)$.

Lemma 3. *Let $K_h(\cdot,\cdot)$ be given by (3.2), where K satisfies (3.13) with $m \geq \lfloor\beta\rfloor + 1$; then for any $x > 0$ and $h \in (0, \ln r)$*

$$\sup_{f \in \mathscr{H}_{x,r}(A,\beta)}\left|\int_{-\infty}^{\infty} K_h(x,y)f(y)\,\mathrm{d}y - f(x)\right| \leq c_0 A\|K\|_1\left[h^{\beta}|x|^{\beta} + h^{\ell+1}\sum_{k=0}^{\ell}|x|^{k}\right],$$

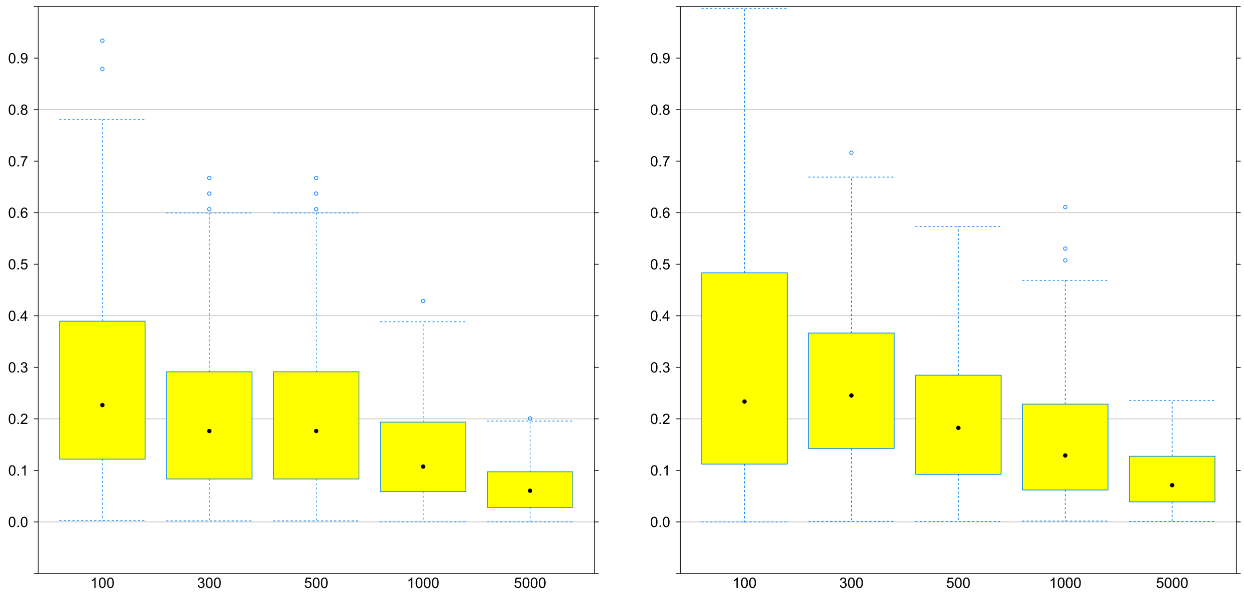

Fig. 2. Boxplots of the distance $|\hat{f}_h(0) - f_X(0)|$, where the estimate $\hat{f}_h(0)$ is based on $n \in \{100, 300, 500, 1000, 5000\}$ observations of the r.v. Y under beta-distributed errors with density (5.1) with parameters $\nu = 1$ (left) and $\nu = 1/2$ (right).

where c_0 depends on β only, and $\|K\|_1 = \int_{-1}^{1} |K(x)|\,\mathrm{d}x$.

The proof of Lemma 3 is given in Section 7.

2^0. Now we derive an upper bound on the variance. Using the Cauchy–Schwarz inequality we obtain

$$\begin{aligned}
&\mathbb{E}_{f_X}\big[L_h^2(x_0, Y_j)\big] \\
&\quad = \frac{1}{4\pi^2 x_0^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x_0|^{i(\omega-\mu)} \widetilde{f}_Y\big(1 - i(\omega-\mu)\big) \frac{\widehat{K}(-\omega h)}{\widetilde{g}(1-i\omega)} \cdot \overline{\frac{\widehat{K}(-\mu h)}{\widetilde{g}(1-i\mu)}}\,\mathrm{d}\omega\,\mathrm{d}\mu \\
&\quad \le \frac{1}{4\pi^2 x_0^2} \int_{-\infty}^{\infty} \big|\widetilde{f}_Y(1-i\mu)\big|\,\mathrm{d}\mu \int_{-\infty}^{\infty} \frac{|\widehat{K}(-\omega h)|^2}{|\widetilde{g}(1-i\omega)|^2}\,\mathrm{d}\omega.
\end{aligned}$$

If $\gamma > 1$ then $\widetilde{f}_Y(1-i\mu)$ is integrable:

$$\begin{aligned}
\int_{-\infty}^{\infty} \big|\widetilde{f}_Y(1-i\mu)\big|\,\mathrm{d}\mu &= \int_{-\infty}^{\infty} \big|\widetilde{f}_X(1-i\mu)\big| \cdot \big|\widetilde{g}(1-i\mu)\big|\,\mathrm{d}\mu \\
&\le \int_{-\infty}^{\infty} \big|\widetilde{g}(1-i\mu)\big|\,\mathrm{d}\mu \le c_1 < \infty,
\end{aligned}$$

where the upper bound in (3.12) has been used. Moreover, in view of (3.14) and the lower bound in (3.12) we have

$$\int_{-\infty}^{\infty} \frac{|\widehat{K}(-\omega h)|^2}{|\widetilde{g}(1-i\omega)|^2}\,\mathrm{d}\omega \le c_2 h^{-2\gamma-1}.$$

Combining these bounds we obtain $\mathrm{var}_{f_X}\{\hat{f}_h(x_0)\} \le c_3 x_0^{-2} h^{-2\gamma-1} n^{-1}$.

On the other hand, Lemma 3 and $h \le 1$ imply that

$$\sup_{f_X \in \mathscr{H}_{x_0,r}(A,\beta)} \big|\mathbb{E}_{f_X}\big[\hat{f}_h(x_0)\big] - f_X(x_0)\big| \le c_3 A\big(x_0^{\beta} + 1\big) h^{\beta}.$$

Then (3.16) follows from substitution of h_* in the bounds for the bias and the variance.

6.2. *Proof of Theorem 2*

The proof is based on the standard technique for proving lower bounds (see [17, Chapter 2]). Recall that for two generic functions u and w on $[0,\infty)$ we write $[w\star u](y):=\int_0^\infty (1/x)w(x)u(y/x)\,\mathrm{d}x$.

0^0. Let $\psi:\mathbb{R}\to\mathbb{R}$ be a function such that its Fourier transform $\widehat{\psi}$ is an infinitely differentiable function satisfying for some $\delta\in(0,\frac14)$

$$\mathcal{F}[\psi;\omega]=\widehat{\psi}(\omega)=\begin{cases}1, & \omega\in[-2+\delta,-1-\delta]\cup[1+\delta,2-\delta],\\ 0, & \omega\in(-\infty,-2]\cup[-1,1]\cup[2,\infty).\end{cases}$$

Let $x_0\ge c_0>0$ for some constant c_0, and define

$$f_X^{(0)}(x):=\frac{1}{\pi x[1+\ln^2(x/x_0)]},\qquad x>0.$$

Define

$$f_X^{(1)}(x)=f_X^{(0)}(x)+\theta\psi_h(x),\qquad \psi_h(x):=\frac{1}{x}\psi\left(\frac{\ln(x/x_0)}{h}\right),$$

where $h\in(0,1)$ and $\theta>0$ are the parameters to be specified.

1^0. First we show that if θ is small enough, $\theta\le\min\{\frac12,c_1h^{-2}\}$ then $f_X^{(1)}$ is a probability density on $[0,\infty)$. Indeed, since $\widehat{\psi}(0)=0$

$$\int_0^\infty\psi_h(x)\,\mathrm{d}x=\int_0^\infty\frac{1}{x}\psi\left(\frac{\ln(x/x_0)}{h}\right)\mathrm{d}x=h\int_{-\infty}^\infty\psi(t)\,\mathrm{d}t=0.$$

Thus, $f_X^{(1)}$ integrates to one. Moreover, by construction ψ is rapidly decreasing as $t\to\infty$; in particular, $|\psi(t)|\le\pi^{-1}\min\{1,c_1t^{-2}\}$, $\forall t\in\mathbb{R}$ with some absolute constant c_1. Therefore, the conditions $\theta\le\frac12$ and $c_1\theta h^2\le\frac12$ imply that $\theta|\psi(t/h)|\le[\pi(1+t^2)]^{-1}$, which, in turn shows that $f_X^{(1)}$ is non-negative. Therefore $f_X^{(1)}$ is the probability density.

2^0. First we note that if $x_0\ge c_0>0$ for some c_0 large enough then $f_X^{(0)}\in\mathscr{H}_{x_0,r}(A/2,\beta)$. Now we show that if $\theta=c_2Ax_0^{\beta+1}h^\beta$ for some constant c_2 then $f_X^{(1)}\in\mathscr{H}_{x_0,r}(A,\beta)$.

For simplicity and without loss of generality assume that β is integer, $\beta\ge1$. Then by the Faá di Bruno formula

$$\begin{aligned}\psi_h^{(\beta)}(x)&=\sum_{j=0}^{\beta}\binom{\beta}{j}\frac{(-1)^jj!}{x^{j+1}}\frac{d^{\beta-j}}{dx^{\beta-j}}\psi\left(\frac{\ln(x/x_0)}{h}\right)\\ &=\sum_{j=0}^{\beta}\binom{\beta}{j}\frac{(-1)^jj!}{x^{j+1}}\sum\frac{(\beta-j)!}{k_1!\cdots k_{\beta-j}!}\psi^{(k)}\left(\frac{\ln(x/x_0)}{h}\right)h^{-k}x^{-(\beta-j)}\prod_{i=1}^{\beta-j}\left[\frac{(-1)^{i+1}}{i!}\right]^{k_i},\end{aligned}$$

where the second summation is over all partitions of $\beta-j$, and $k:=k_1+\cdots+k_{\beta-j}$, $k_1+2k_2+\cdots+(\beta-j)k_{\beta-j}=\beta-j$. It follows from this expression and the fact that $h<1$ that

$$\left|\psi_h^{(\beta)}(x)\right|\le c_3x^{-\beta-1}h^{-\beta}\max_{k=1,\dots,\beta}\left|\psi^{(k)}\left(\frac{\ln(x/x_0)}{h}\right)\right|,\qquad\forall x>0,$$

where c_3 depends on β only. Since ψ is an infinite differentiable rapidly decreasing function, we obtain

$$\left|\psi_h^{(\beta)}(x)\right|\le c_4x_0^{-\beta-1}h^{-\beta},\qquad r^{-1}x_0\le x\le rx_0,$$

where c_4 depends on β. Then setting $\theta=c_2Ax_0^{\beta+1}h^\beta$, by choice of c_2 we obtain $f_X^{(1)}\in\mathscr{H}_{x_0,r}(A,\beta)$.

3^0. Next we bound the χ^2-divergence between $f_Y^{(1)}$ and $f_Y^{(0)}$. We have

$$\begin{aligned}f_Y^{(0)}(y)=\left[f_X^{(0)}\star g\right](y)&=\frac{1}{\pi y}\int_0^\infty\frac{g(x)}{1+[\ln(y/x_0)-\ln(x)]^2}\,\mathrm{d}x\\ &\ge\frac{1}{\pi y[1+2\ln^2(y/x_0)]}\int_0^\infty\frac{g(x)}{1+2\ln^2(x)}\,\mathrm{d}x\ge\frac{c_5}{y[1+2\ln^2(y/x_0)]}.\end{aligned}$$

Furthermore,

$$f_Y^{(1)}(y) - f_Y^{(0)}(y) = \theta[g \star \psi_h](y) = \theta \int_0^\infty \frac{1}{x} g(x)\psi_h(y/x)\,\mathrm{d}x$$
$$= \frac{\theta}{2\pi y} \int_{-\infty}^{\infty} \widetilde{g}(1+i\omega)\widetilde{\psi}_h(1+i\omega) y^{-i\omega}\,\mathrm{d}\omega, \tag{6.1}$$

where in the second line we have applied the inverse Mellin transform formula. By definition of ψ_h,

$$\widetilde{\psi}_h(1+i\omega) = \int_0^\infty x^{i\omega}\psi_h(x)\,\mathrm{d}x = \int_0^\infty x^{i\omega-1}\psi\left(\frac{\ln(x/x_0)}{h}\right)\mathrm{d}x$$
$$= hx_0^{i\omega}\int_{-\infty}^{\infty} e^{ith\omega}\psi(t)\,\mathrm{d}t = hx_0^{i\omega}\widehat{\psi}(-\omega h).$$

Substituting this expression in (6.1) we obtain

$$f_Y^{(1)}(y) - f_Y^{(0)}(y) = \frac{\theta h}{2\pi y}\int_{-\infty}^{\infty} \widetilde{g}(1+i\omega)\widehat{\psi}(-\omega h) e^{-i\omega \ln(y/x_0)}\,\mathrm{d}\omega =: \frac{\theta h}{2\pi y}\rho\big(\ln(y/x_0)\big).$$

The χ^2-divergence between $f_Y^{(1)}$ and $f_Y^{(0)}$ is bounded as follows

$$\chi^2\big(f_Y^{(1)}, f_Y^{(0)}\big) = \int_0^\infty \frac{(f_Y^{(1)}(y) - f_Y^{(0)}(y))^2}{f_Y^{(0)}(y)}\,\mathrm{d}y$$
$$\le c_6\theta^2h^2\int_0^\infty \big[1 + 2\ln^2(y/x_0)\big]\frac{1}{y}\rho^2\big(\ln(y/x_0)\big)\,\mathrm{d}y = c_6\theta^2h^2\int_{-\infty}^{\infty}\big(1+2t^2\big)\rho^2(t)\,\mathrm{d}t.$$

By Parseval's identity, definition of ψ and Assumption [G1]

$$\int_{-\infty}^{\infty}\rho^2(t)\,\mathrm{d}t = \int_{-\infty}^{\infty}\big|\widetilde{g}(1+i\omega)\big|^2\big|\widehat{\psi}(-\omega h)\big|^2\,\mathrm{d}\omega \le 2\int_{1/h}^{2/h}\big|\widetilde{g}(1+i\omega)\big|^2\,\mathrm{d}\omega \le c_7 h^{2\gamma-1}. \tag{6.2}$$

Moreover, using Assumptions [G1] and [G1′]

$$\int_{-\infty}^{\infty} t^2\rho^2(t)\,\mathrm{d}t = \int_{-\infty}^{\infty}\left|\frac{\mathrm{d}}{\mathrm{d}\omega}\widetilde{g}(1+i\omega)\widehat{\psi}(-\omega h)\right|^2\mathrm{d}\omega$$
$$\le 2\int_{-\infty}^{\infty}\big|\widetilde{g}'(1+i\omega)\big|^2\big|\widehat{\psi}(-\omega h)\big|^2\,\mathrm{d}\omega + 2\int_{-\infty}^{\infty}\big|\widetilde{g}(1+i\omega)\big|^2\big|\widehat{\psi}'(-\omega h)\big|^2 h^2\,\mathrm{d}\omega$$
$$\le c_8h^{2\gamma-1} + c_9h^{2\gamma+1}.$$

Combining these bounds with (6.2) for h small enough we obtain

$$\chi^2\big(P^{(1)}, P^{(0)}\big) \le c_9\theta^2h^{2\gamma+1} = c_{10}A^2x_0^{2\beta+2}h^{2\beta+2\gamma+1}.$$

4^0. Now we complete the proof. Let

$$h = h_* := c_{11}x_0^{-\frac{2\beta+2}{2\beta+2\gamma+1}}\big(A^2n\big)^{-\frac{1}{2\beta+2\gamma+1}}.$$

With this choice $\theta = c_2Ax_0^{\beta+1}h_*^{\beta} \le \frac{1}{2}$ for n large enough so $f_X^{(1)} \in \mathscr{H}_{x_0,r}(A,\beta)$. We obtain $\chi^2(f_Y^{(1)}, f_Y^{(0)}) \le 1/n$ so that the hypotheses $f_X = f_X^{(0)}$ and $f_X = f_X^{(1)}$ are indistinguishable from the observations $Y_1,\ldots,Y_n$. Moreover, with this choice of the parameter h

$$\big|f_X^{(1)}(x_0) - f_X^{(0)}(x_0)\big| = \theta\big|\psi_{h_*}(x_0)\big| = c_{12}Ax_0^{\beta+1}h_*^{\beta}x_0^{-1}\big|\psi(0)\big|$$
$$= c_{13}A^{\frac{2\gamma+1}{2\beta+2\gamma+1}}x_0^{\frac{\beta(2\gamma-1)}{2\beta+2\gamma+1}}n^{-\frac{\beta}{2\beta+2\gamma+1}}.$$

This completes the proof of the theorem.

6.3. *Proof of Theorem 3*

The bound on bias of $\hat{f}_{s,h}(x_0)$ given in Lemma 3 remains intact. We consider only the variance term. For

$$L_{s,h}(x,y) = \frac{1}{2\pi x}\int_{-\infty}^{\infty}\left(\frac{x}{y}\right)^{s+i\omega}\frac{\check{K}(-(s+i\omega)h)}{\widetilde{g}(1-s-i\omega)}\,\mathrm{d}\omega$$

we have

$$\begin{aligned}
&\mathbb{E}_{f_X}\big[L_{s,h}^2(x_0,Y_j)\big]\\
&\quad= \frac{1}{4\pi^2 x_0^{2-2s}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_0^{i(\omega-\mu)}\widetilde{f}_Y\big(1-2s-i(\omega-\mu)\big)\frac{\check{K}(-(s+i\omega)h)}{\widetilde{g}(1-s-i\omega)}\cdot\overline{\frac{\check{K}(-(s+i\mu)h)}{\widetilde{g}(1-s-i\mu)}}\,\mathrm{d}\omega\,\mathrm{d}\mu\\
&\quad\le \frac{1}{4\pi^2 x_0^{2-2s}}\int_{-\infty}^{\infty}\big|\widetilde{f}_Y(1-2s-i\mu)\big|\,\mathrm{d}\mu\int_{-\infty}^{\infty}\left|\frac{\check{K}(-(s+i\omega)h)}{\widetilde{g}(1-s-i\omega)}\right|^2\mathrm{d}\omega.
\end{aligned}$$

Since $f_X\in\mathscr{F}_{\alpha,M}(A,\beta)$, $|\widetilde{f}_X(1-2s-i\mu)|\le 1+M<\infty$ for all $\mu\in\mathbb{R}$ and $-\alpha\le s\le 0$. For such s

$$\int_{-\infty}^{\infty}\big|\widetilde{f}_Y(1-2s-i\mu)\big|\,\mathrm{d}\mu \le (1+M)\int_{-\infty}^{\infty}\big|\widetilde{g}(1-2s-i\mu)\big|\,\mathrm{d}\mu\le c_1(1+M),$$

provided that $a<1-2s<b$. Setting $s=s_*=\max\{-\alpha,\frac{1}{2}(1-b)+\epsilon\}$ for any $\epsilon>0$ we obtain

$$\mathbb{E}_{f_X}\big[L_{s_*,h}^2(x_0,Y_j)\big]\le\frac{c_2(1+M)}{4\pi^2 x_0^{2-2s_*}}\int_{-\infty}^{\infty}\left|\frac{\check{K}(-(s_*+i\omega)h)}{\widetilde{g}(1-s_*-i\omega)}\right|^2\mathrm{d}\omega. \tag{6.3}$$

Furthermore,

$$\check{K}\big(-(s_*+i\omega)h\big)=\int_{-1}^{1}K(x)e^{s_*hx}e^{i\omega hx}\,\mathrm{d}x=\mathcal{F}[v_{s_*,h};-\omega h]=\widehat{v}_{s_*,h}(-\omega h),$$

where $v_{s,h}(x):=K(x)e^{-shx}\mathbf{1}_{[-1,1]}(x)$. Therefore

$$\begin{aligned}
\int_{-\infty}^{\infty}\left|\frac{\check{K}(-(s_*+i\omega)h)}{\widetilde{g}(1-s_*-i\omega)}\right|^2\mathrm{d}\omega &= \int_{-\infty}^{\infty}\left|\frac{\widehat{v}_{s_*,h}(-\omega h)}{\widetilde{g}(1-s_*-i\omega)}\right|^2\mathrm{d}\omega\\
&\le\frac{c_3}{h^{2\gamma+1}}\int_{-\infty}^{\infty}\big|\widehat{v}_{s_*,h}(\omega)\big|^2\big(1+|\omega|^{2\gamma}\big)\,\mathrm{d}\omega,
\end{aligned}\tag{6.4}$$

where c_3 may depend on s_*. In view of (3.14), $v_{s,h}$ is q times continuously differentiable on its support, and $v_{s,h}^{(q)}(x)=\sum_{j=0}^{q}\binom{q}{j}K^{(j)}(x)(-sh)^{q-j}e^{-shx}$. Therefore

$$\begin{aligned}
\big\|v_{s_*,h}^{(q)}\big\|_2 &\le\sum_{j=0}^{q}\binom{q}{j}e^{2|s_*|h}|s_*h|^{q-j}\left[\int_{-1}^{1}\big|K^{(j)}(x)\big|^2\,\mathrm{d}x\right]^{1/2}\\
&\le c_4\max_{j=0,\dots,q}\big\|K^{(j)}\big\|_2\le c_5C_K.
\end{aligned}$$

Taking into account that $q>\gamma+1$ and combining this inequality with (6.4) and (6.3) we obtain

$$\operatorname{var}_{f_X}\big\{\hat{f}_{s_*,h}(x_0)\big\}\le c_6(1+M)x_0^{-2+2s_*}h^{-2\gamma-1}n^{-1}.$$

This bound together with the bound on the bias leads to the announced result.

6.4. *Proof of Theorem 4*

The proof goes along the same lines as the proof of Theorem 1. In the proof below $c_1, c_2, \dots$ stand for positive constants depending on β and λ only.

It is immediate to verify that

$$\big|\widehat{K}(\omega)\big| \le c_1 \exp\{-\omega^{2\lambda}/2\lambda\}, \quad \forall \omega \in \mathbb{R}. \tag{6.5}$$

This fact together with Assumption [G2] guarantees that the estimator $\hat{f}_{h_*}(x_0)$ is well-defined. In addition, by [11, Chapter IV, §7] as $|t| \to \infty$

$$\begin{aligned} w(t) &= 2\sqrt{\frac{1}{2\lambda-1}}|t|^{-(\lambda-1)/(2\lambda-1)} \exp\left\{-\left(\frac{2\lambda-1}{2\lambda}\right)\sin\left(\frac{\pi}{2(2\lambda-1)}\right)|t|^{2\lambda/(2\lambda-1)}\right\} \\ &\quad \times \left[\cos\left(\frac{2\lambda-1}{2\lambda}|t|^{2\lambda/(2\lambda-1)}\cos\left(\frac{\pi}{2(2\lambda-1)}\right)\right) + O\big(|t|^{-2\lambda/(2\lambda-1)}\big)\right]. \end{aligned}$$

Therefore, it follows from (3.25) that for large $|t|$ one has

$$\big|K(t)\big| \le c_2|t|^{-(\lambda-1)/(2\lambda-1)} \exp\{-c_3|t|^{2\lambda/(2\lambda-1)}\}. \tag{6.6}$$

First we bound the bias of the estimator $\hat{f}_{h_*}(x_0)$. To that end we note that the proof of Lemma 3 applies verbatim; the only difference is that now the integration in (7.10) is over the whole real line because K is not compactly supported. However, since K is a bounded function and in view of (6.6) we have

$$\int_{-\infty}^{\infty} |t|^{\ell+1} e^{\ell|th|}\big|K(t)\big|\,\mathrm{d}t \le \int_{-\infty}^{\infty} |t|^{\ell+\frac{\lambda}{2\lambda-1}} \exp\{\ell|t| - c_3|t|^{-\frac{2\lambda}{2\lambda-1}}\}\,\mathrm{d}t \le c_4.$$

This inequality and reasoning of the proof of Lemma 3 yield

$$\sup_{f\in\mathscr{H}_{x,r}(A,\beta)}\left|\int_{-\infty}^{\infty} K_h(x,y)f(y)\,\mathrm{d}y - f(x)\right| \le c_5 A\left[h^\beta|x|^\beta + h^{\ell+1}\sum_{k=0}^{\ell}|x|^k\right]. \tag{6.7}$$

To bound the variance we follow the lines of the proof of Theorem 1. In particular, in view of (6.5) and Assumption [G2] we have for small enough h

$$\mathbb{E}_{f_X}\big[L_h^2(x_0, Y_j)\big] \le \frac{c_6}{x_0^2}\int_{-\infty}^{\infty} \frac{|\widehat{K}(-\omega h)|^2}{|\widetilde{g}(1-i\omega)|^2}\,\mathrm{d}\omega \le \frac{c_6}{x_0^2}\exp\{c_7\big(\gamma h^{-1}\big)^{2\lambda/(2\lambda-1)}\}. \tag{6.8}$$

Indeed, in view of (6.5) and [G2] we have

$$\int_{|\omega|\le\omega_0} \frac{|\widehat{K}(-\omega h)|^2}{|\widetilde{g}(1-i\omega)|^2}\,\mathrm{d}\omega \le c_1^2 c_0^{-2},$$

where constants ω_0 and c_0 appear in [G2]. Let $\omega_1 := c_8(\lambda\gamma)^{1/(2\lambda-1)}h^{-2\lambda/(2\lambda-1)}$ for sufficiently large constant c_8 that can depend on ν. Then we have

$$\begin{aligned} \int_{|\omega|>\omega_0} \frac{|\widehat{K}(-\omega h)|^2}{|\widetilde{g}(1-i\omega)|^2}\,\mathrm{d}\omega &\le c_9\int_{|\omega|>\omega_0} |\omega|^{-2\nu}\exp\{2|\omega|\gamma - |\omega h|^{2\lambda}/\lambda\}\,\mathrm{d}\omega \\ &\le c_{10}\left\{\int_{\omega_0\le|\omega|\le\omega_1} \exp\{c_{11}|\omega|\gamma\}\,\mathrm{d}\omega + \int_{|\omega|\ge\omega_1} \exp\{-c_{12}|\omega h|^{2\lambda}\}\,\mathrm{d}\omega\right\} \\ &\le c_{13}\exp\{c_{12}\big(\gamma h^{-1}\big)^{2\lambda/(2\lambda-1)})\}, \end{aligned}$$

which leads to (6.8). Then the result of the theorem follows from balancing the bounds in (6.7) and (6.8).

6.5. *Proof of Theorem 5*

In the proof below $c_1, c_2, \dots$ stand for positive constants; they can depend on parameters appearing in assumptions [G3] and [G4] and on parameter β only. The proof proceeds in steps.

1^0. First we show that under the premise of the theorem the estimator $\hat{f}_{s_*,h}(0)$ is well defined. It follows from the definition of function $\psi_s(x)$ that

$$\widetilde{\psi}_s(s+i\omega) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-s)^2} \int_0^\infty x^{s+i\omega-1} x^{-s} \exp\left\{-\frac{1}{2}[\ln x]^2\right\} \mathrm{d}x$$
$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-s)^2} e^{-\frac{1}{2}\omega^2};$$

therefore

$$\int_0^\infty \psi_s\left(\frac{x}{j}\right) x^{s+i\omega-1}\,\mathrm{d}x = j^{s+i\omega}\widetilde{\psi}_s(s+i\omega) = \frac{j^{s+i\omega}}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-s)^2} e^{-\frac{1}{2}\omega^2},$$

and

$$\widetilde{K}_s(s+i\omega) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-s)^2} e^{-\omega^2/2} \sum_{j=1}^{m+1} \binom{m+1}{j} (-1)^{j+1} j^{s-1+i\omega}.$$

The last expression implies that

$$\left|\widetilde{K}_s(s+i\omega)\right| \le c_1 m^s e^{-\frac{1}{2}(1-s)^2} e^{-\omega^2/2}, \tag{6.9}$$

where c_1 depends on m only. Next we observe that Assumption [G3] implies $1-s_* = 1-\frac{1}{2}(1-p) \in \Omega_g$, so that $\widetilde{g}(1-s_*+i\omega)$ is well defined. Then in view of (6.9) and condition (4.8) of Assumption [G4], $\widetilde{K}_{s_*}(s_*+i\cdot)/\widetilde{g}(1-s_*-i\cdot) \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$ so that $\hat{f}_{s_*,h}(0)$ is well defined.

2^0. Our next step is to prove the following statement about local behavior of the density f_Y near the origin. This result is instrumental in establishing an upper bound on the variance term.

Lemma 4. *Let Assumption* [G3] *hold, and assume that* $f_X(t) \le M, \forall t$.

(i) *If* $p=0$ *then for all* $y \le \delta$

$$f_Y(y) \le C_1(1+M)|\ln y|^{q+1} + M\delta^{-1},$$

where C_1 *depends on* q *only.*

(ii) *If* $p \in (0,1)$ *then for all* $y \le \delta$

$$f_Y(y) \le C_2\big(1+Mp^{-1}\big) y^{-p} |\ln y|^q + M\delta^{-1},$$

where C_2 *depends on* q *only.*

The proof of the lemma is given in Section 7.

3^0. Now we are ready to establish an upper bound on the variance term. Define

$$\rho_s(x) := \frac{1}{2\pi} \int_{-\infty}^\infty e^{-i\omega x} \frac{\widetilde{K}_s(s+i\omega)}{\widetilde{g}(1-s-i\omega)}\,\mathrm{d}\omega.$$

With this notation $L_{s,h}(y) = h^{s-1} y^{-s} \rho_s(\ln(y/h))$ [cf. (4.4)], and therefore

$$\mathbb{E}_{f_X}\big[L_{s,h}^2(Y)\big] = \frac{1}{h^{2-2s}} \int_0^\infty \frac{|\rho_s(\ln(y/h))|^2}{y^{2s}} f_Y(y)\,\mathrm{d}y.$$

Now we bound the last integral which can be written as a sum J_1+J_2, where

$$J_1 := \frac{1}{h^{2-2s}} \int_0^\delta \frac{|\rho_s(\ln(y/h))|^2}{y^{2s}} f_Y(y)\,\mathrm{d}y, \qquad J_2 := \frac{1}{h^{2-2s}} \int_\delta^\infty \frac{|\rho_s(\ln(y/h))|^2}{y^{2s}} f_Y(y)\,\mathrm{d}y.$$

Using Lemma 4 for $p = 0$ and $s = s_* = \frac{1}{2}$ by straightforward algebra we obtain

$$J_1 \le \frac{c_1 M}{h} \int_0^{\delta} y^{-1} \big|\rho_{s_*}\big(\ln(y/h)\big)\big|^2 \big[|\ln y|^{q+1} + \delta^{-1}\big]\,dy$$
$$\le \frac{c_2 M}{h}\left(|\ln h|^{q+1} \int_{-\infty}^{\infty} \big|\rho_{s_*}(t)\big|^2\,dt + \int_{-\infty}^{\infty} \big|\rho_{s_*}(t)\big|^2 |t|^{q+1}\,dt\right) \le c_3 M h^{-1} |\ln h|^{q+1},$$

where the last inequality follows from Parseval's identity, properties of the kernel K_s and condition [G4] [cf. (4.8) and (4.9)]. If $p \in (0,1)$ then using Lemma 4 for $s = s_* = \frac{1}{2}(1-p)$ we have similarly

$$J_1 \le \frac{c_4 M}{h^{2-2s_*}} \int_0^{\delta_0} \frac{1}{y} \big|\rho_{s_*}\big(\ln(y/h)\big)\big|^2 \frac{y^{-p}|\ln y|^{q+\varkappa}}{y^{2s_*-1}}\,dy$$
$$\le \frac{c_5 M}{h^{1+p}}\left(|\ln h|^{q} \int_{-\infty}^{\infty} \big|\rho_{s_*}(t)\big|^2\,dt + \int_{-\infty}^{\infty} \big|\rho_{s_*}(t)\big|^2 |t|^{q}\,dt\right) \le c_6 M h^{-1-p}\big[\ln(1/h)\big]^{q}.$$

Combining the last two upper bounds on J_1 in cases $p = 0$ and $p \in (0,1)$ we can write

$$J_1 \le c_7 M h^{-1-p}\big[\ln(1/h)\big]^{q+\varkappa},$$

where $\varkappa$ is defined in (4.10).

In order to bound J_2 we note that (4.8) implies $|\rho_{s_*}(x)| \le c_8 < \infty$, $\forall x$; therefore

$$J_2 \le c_8^2 h^{-1-p} \int_{\delta_0}^{\infty} y^{p-1} f_Y(y)\,dy \le c_9 h^{-1-p}.$$

Combining the bounds on J_1 and J_2 we obtain

$$\mathbb{E}_{f_X}\big[L^2_{s_*,h}(Y)\big] \le c_{10} M h^{-1-p}\big[\ln(1/h)\big]^{q+\varkappa}.$$

4^0. We proceed with bounding the bias of $\hat{f}_{s_*,h}(0)$. By construction of $K_{s_*,h}(x)$ we have

$$\int_0^{\infty} (1/h) K_{s_*}(x/h)\big[f_X(x) - f_X(0)\big]\,dx$$
$$= \int_0^{\infty} K_{s_*}(u)\big[f_X(uh) - f_X(0)\big]\,dx$$
$$= \int_0^{r/h} K_{s_*}(u)\left[\sum_{j=1}^{\ell-1} \frac{1}{j!} f_X^{(j)}(0)(uh)^j + \frac{1}{\ell!} f_X^{(\ell)}(\xi u h)(uh)^{\ell}\right] du$$
$$+ \int_{r/h}^{\infty} K_{s_*}(u)\big[f_X(uh) - f_X(0)\big]\,du.$$

Since $f_X(x) \le M$, $\forall x$, by (4.2) and (4.3)

$$\left|\int_{r/h}^{\infty} K_{s_*}\big[f_X(uh) - f_X(0)\big]\,du\right| \le 2M \sum_{j=1}^{m+1} \binom{m+1}{j} \int_{r/h}^{\infty} \psi_{s_*}(x)\,dx$$
$$\le c_1 M \exp\big\{-c_2\big[\ln(r/h)\big]^2\big\}.$$

Furthermore, it is readily verified that for small enough h

$$\left|\int_0^{r/h} K_{s_*}(u) u^j\,du\right| = \left|\int_{r/h}^{\infty} K_{s_*}(u) u^j\,du\right| \le c_3 \exp\left\{-c_4\left[\ln\left(\frac{r}{mh}\right)\right]^2\right\}.$$

Using these facts we finally obtain that

$$\left|\int_0^\infty (1/h)K_{s_*}(x/h)\big[f_X(x)-f_X(0)\big]\,\mathrm{d}x\right|$$
$$\le c_5Ah^\beta + c_6(1+M)\exp\left\{-c_7\left[\ln\left(\frac{r}{mh}\right)\right]^2\right\}.$$

We complete the proof by noting that the choice $h=h_*$ indicated in the statement of the theorem provides a balance for the bounds on the bias and on the variance.

6.6. *Proof of Theorem 6*

The proof is based on the standard technique for proving lower bounds (see [17, Chapter 2]). Throughout the proof constants $c_0, c_1, \ldots$ may depend only on β and parameters appearing in Assumptions [G3] and [G5].

1^0. Let $M_0=\pi M/4$, and without loss of generality assume that $M_0\ge 1$. Let

$$f_X^{(0)}(x) := \frac{2M_0}{\pi(1+M_0x)(1+\ln^2(1+M_0x))}, \quad x\ge 0.$$

It is evident that $f_X^{(0)}(x)\le M/2$, $\forall x$ and $f_X^{(0)}\in\bar{\mathscr{H}}_r(A,\beta,M)$ provided that A is large enough.

For $h>0$ define

$$f_X^{(1)}(x)=f_X^{(0)}(x)+c_0Ah^\beta\varphi(x/h), \qquad \varphi(x):=(1-x)e^{-x}, \quad x\ge 0.$$

In what follows parameter h will be chosen going to zero as $n\to\infty$; in the subsequent proof we use this fact. It is evident that function $f_X^{(1)}$ is a probability density, and under appropriate choice of constant c_0 and for h small enough it belongs to $\bar{\mathscr{H}}_r(A,\beta,M)$. We note also that

$$\widetilde{\varphi}(z)=\int_0^\infty x^{z-1}\varphi(x)\,\mathrm{d}x=\Gamma(z)-\Gamma(z+1), \quad z\in\Omega_\varphi=\big\{z:\mathrm{Re}(z)>0\big\}. \tag{6.10}$$

Our current goal is to bound the χ^2-divergence between the corresponding densities of observations $f_Y^{(0)}$ and $f_Y^{(1)}$. For any s such that $\{z:\mathrm{Re}(z)=s\}\subseteq\Omega_g\cap\Omega_\varphi$ we have

$$\begin{aligned}
f_Y^{(1)}(y)-f_Y^{(0)}(y) &= c_0Ah^\beta\int_0^\infty \frac{1}{x}\varphi\left(\frac{y}{hx}\right)g(x)\,\mathrm{d}x\\
&= \frac{c_0Ah^\beta}{2\pi i}\int_{s-i\infty}^{s+i\infty}\left(\frac{h}{y}\right)^z\widetilde{\varphi}(z)\widetilde{g}(z)\,\mathrm{d}z\\
&= \frac{c_0Ah^{\beta+s}}{2\pi y^s}\int_{-\infty}^\infty\left(\frac{h}{y}\right)^{i\omega}\widetilde{\varphi}(s+i\omega)\widetilde{g}(s+i\omega)\,\mathrm{d}\omega = c_0Ah^{\beta+s}y^{-s}\rho_s\big(\ln(y/h)\big),
\end{aligned}$$

where we have used the Mellin transform inversion formula, and we have denoted

$$\rho_s(t):=\frac{1}{2\pi}\int_{-\infty}^\infty e^{-i\omega t}\widetilde{\varphi}(s+i\omega)\widetilde{g}(s+i\omega)\,\mathrm{d}\omega=\frac{e^{st}}{2\pi}\int_{s-i\infty}^{s+i\infty}e^{-zt}\widetilde{\varphi}(z)\widetilde{g}(z)\,\mathrm{d}z. \tag{6.11}$$

Thus

$$\chi^2\big(f_Y^{(0)},f_Y^{(1)}\big)=c_0^2A^2h^{2(\beta+s)}\int_0^\infty\frac{y^{-1}\rho_s^2(\ln(y/h))}{y^{2s-1}f_Y^{(0)}(y)}\,\mathrm{d}y, \tag{6.12}$$

and now we will bound the integral on the right hand side under a particular choice of parameter s.

2^0. Let $s=s_*:=\frac{1}{2}(p+1)$. Note that by the upper bound in (4.6) and by definition of s_*

$$\big|\widetilde{g}(s_*+i\omega)\big|\le\int_0^\infty x^{s_*-1}g(x)\,\mathrm{d}x=\int_0^\infty x^{(p-1)/2}g(x)\,\mathrm{d}x\le c_1;$$

thus $\{z : \mathrm{Re}(z) = s_*\} \subseteq \Omega_g$. Let $\nu := \frac{1}{2}(1-p)+\epsilon$, where ϵ is given in Assumption [G5]. Then $s_* + \nu = 1+\epsilon$, and according to Assumptions [G3] and [G5], function $e^{-zt}\widetilde{g}(z)\widetilde{\varphi}(z)$ is analytic in $\{z : s_* \le \mathrm{Re}(z) \le s_*+\nu\}$. Therefore the line of integration in the last integral on the right hand side of (6.11) can be replaced by $\{z : \mathrm{Re}(z) = s_*+\nu\}$. This yields

$$\begin{aligned}\rho_{s_*}(t) &= \frac{e^{s_*t}}{2\pi}\int_{s_*+\nu-i\infty}^{s_*+\nu+i\infty} e^{-zt}\widetilde{g}(z)\widetilde{\varphi}(z)\,\mathrm{d}z \\ &= \frac{e^{-\nu t}}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\widetilde{g}(1+\epsilon+i\omega)\widetilde{\varphi}(1+\epsilon+i\omega)\,\mathrm{d}\omega.\end{aligned}$$

Then it follows from Assumption [G5] that

$$\left|\rho_{s_*}(t)\right| \le c_2 e^{-\nu t}\int_{-\infty}^{\infty}\left|\widetilde{\varphi}(1+\epsilon+i\omega)\right|\mathrm{d}\omega \le c_3 e^{-\nu t}, \tag{6.13}$$

where the last inequality follows from (6.10) and bounds on the Gamma function as presented in (3.23) in Example 4.

3^0. Now we derive lower bounds on $f_Y^{(0)}(y)$. Note that $f_X^{(0)}(x) = M_0\bar{f}_X^{(0)}(M_0x)$ where

$$\bar{f}_X^{(0)}(x) := \frac{2}{\pi(1+x)(1+\ln^2(1+x))}, \quad x \ge 0.$$

Therefore $f_Y^{(0)}(y) = M_0\bar{f}_Y^{(0)}(M_0y)$, $\bar{f}_Y^{(0)}(y) := [\bar{f}_X^{(0)} \star g](y)$ and the lower bounds on $f_Y^{(0)}(y)$ can be obtained in an evident way from the corresponding bounds on $\bar{f}_Y^{(0)}(y)$.

First we note that the lower bound in (4.6) and the arguments as in the proof of (7.14) in Lemma 4, yield for all $y < \delta/2$

$$\int_y^\delta \left[g(t)/t\right]\mathrm{d}t \ge c_4 y^{-p}|\ln y|^{q+\varkappa}, \tag{6.14}$$

where $\varkappa$ is defined in (4.10). In view of (6.14) for $y < \delta/2$

$$\begin{aligned}\bar{f}_Y^{(0)}(y) &\ge \int_y^\delta \frac{2g(x)}{\pi x(1+y/x)(1+\ln^2(1+y/x))}\,\mathrm{d}x \\ &\ge \frac{1}{\pi(1+\ln^2(2))}\int_y^\delta \frac{g(x)}{x}\,\mathrm{d}x \ge c_5 y^{-p}|\ln y|^{q+\varkappa}.\end{aligned}$$

Thus,

$$f_Y^{(0)}(y) \ge c_5 M_0^{1-p} y^{-p}\left|\ln(M_0 y)\right|^{q+\varkappa}, \quad \forall y < \delta/(2M_0). \tag{6.15}$$

On the other hand, for any y we have

$$\begin{aligned}\bar{f}_Y^{(0)}(y) &= \int_0^\infty \frac{2g(x)}{\pi x(1+y/x)(1+\ln^2(1+y/x))}\,\mathrm{d}x \\ &\ge \int_0^1 \frac{2g(x)}{\pi(x+y)(1+2\ln^2(x+y)+2\ln^2(x))}\,\mathrm{d}x \\ &\ge \frac{2}{\pi(1+y)(1+2\ln^2(1+y))}\int_0^1 \frac{g(x)}{1+2\ln^2(x)}\,\mathrm{d}x \\ &\ge \frac{c_6}{(1+y)(1+2\ln^2(1+y))},\end{aligned}$$

so that

$$f_Y^{(0)}(y) \ge \frac{c_6 M_0}{(1+M_0y)(1+2\ln^2(1+M_0y))}, \quad \forall y. \tag{6.16}$$

4^0. Now we bound from above the integral on the right hand side of (6.12).

Let $\xi \in (h, \delta/(2M_0))$ be a parameter that will be specified later; then we can write the integral on the right hand side of (6.12) in the following form

$$\int_0^\infty \frac{y^{-1}\rho_s^2(\ln(y/h))}{y^{2s-1}f_Y^{(0)}(y)}\,\mathrm{d}y = \int_0^\xi \frac{y^{-1}\rho_s^2(\ln(y/h))}{y^{2s-1}f_Y^{(0)}(y)}\,\mathrm{d}y + \int_\xi^\infty \frac{y^{-1}\rho_s^2(\ln(y/h))}{y^{2s-1}f_Y^{(0)}(y)}\,\mathrm{d}y =: I_1 + I_2. \tag{6.17}$$

Our current goal is to bound I_1 and I_2 when $s = s_* = \frac{1}{2}(p+1)$.

Using (6.15) we obtain

$$I_1 \le \frac{c_6}{M_0^{1-p}} \int_0^\xi \frac{y^{-1}\rho_{s_*}^2(\ln(y/h))}{|\ln(M_0 y)|^{q+\varkappa}}\,\mathrm{d}y \le \frac{c_6|\ln(M_0\xi)|^{-q-\varkappa}}{M_0^{1-p}} \int_{-\infty}^{\ln(\xi/h)} \rho_{s_*}^2(t)\,\mathrm{d}t. \tag{6.18}$$

It follows from (6.11) that

$$\int_{-\infty}^\infty \rho_{s_*}^2(t)\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^\infty \big|\widetilde{\varphi}(s_*+i\omega)\big|^2\big|\widetilde{g}(s_*+i\omega)\big|^2\,\mathrm{d}\omega \le c_7\int_{-\infty}^\infty \big|\widetilde{\varphi}(s_*+i\omega)\big|^2\,\mathrm{d}\omega = c_7\int_0^\infty x^{2s_*-1}\varphi^2(x)\,\mathrm{d}x \le c_8,$$

where the equality in the last line follows from the Parseval identity (2.2), and the last inequality is by definition of φ. This inequality together with (6.18) leads to

$$I_1 \le c_9 M_0^{-1+p}\big|\ln(M_0\xi)\big|^{-q-\varkappa}. \tag{6.19}$$

Now consider the integral I_2 on the right hand side of (6.17). Using (6.16) we write (remind that $2s_*-1=p$)

$$\begin{aligned} I_2 &\le \frac{c_{10}}{M_0}\int_\xi^\infty y^{-p-1}\rho_{s_*}^2\big(\ln(y/h)\big)(1+M_0y)\big[1+\ln^2(1+M_0y)\big]\,\mathrm{d}y \\ &= \frac{c_{10}}{M_0}\bigg\{\int_\xi^\infty y^{-p-1}\big[1+\ln^2(1+M_0y)\big]\rho_{s_*}^2\big(\ln(y/h)\big)\,\mathrm{d}y \\ &\quad + M_0\int_\xi^\infty y^{-p}\big[1+\ln^2(1+M_0y)\big]\rho_{s_*}^2\big(\ln(y/h)\big)\,\mathrm{d}y\bigg\} =: \frac{c_{10}}{M_0}\big\{I_2^{(1)}+I_2^{(2)}\big\}. \end{aligned} \tag{6.20}$$

Applying (6.13), using a simple inequality $\ln(1+x) \le \ln 2 + |\ln x|$, $x \ge 0$, and assuming that h is small so that $M_0h \le 1$ we derive

$$\begin{aligned} I_2^{(1)} &= h^{-p}\int_{\ln(\xi/h)}^\infty e^{-pt}\rho_{s_*}^2(t)\big[1+\ln^2\big(1+M_0he^t\big)\big]\,\mathrm{d}t \\ &\le c_{11}h^{-p}\int_{\ln(\xi/h)}^\infty e^{-(p+\nu)t}e^{-\nu t}\big(1+t^2\big)\,\mathrm{d}t \\ &\le c_{12}h^\nu\xi^{-p-\nu}\int_0^\infty e^{-\frac{1}{2}(1-p)t}\big(1+t^2\big)\,\mathrm{d}t \le c_{13}h^\nu\xi^{-p-\nu}, \end{aligned} \tag{6.21}$$

and similarly

$$I_2^{(2)} \le c_{14}M_0h^{-p+1}\int_{\ln(\xi/h)}^\infty e^{t(1-p)}e^{-2\nu t}\big[1+t^2\big]\,\mathrm{d}t \le c_{15}M_0h^{-p+1}, \tag{6.22}$$

where we have used that $\nu = \frac{1}{2}(1-p)+\epsilon$. Combining inequalities (6.22), (6.21), (6.20) and (6.19) we conclude that for small enough h and for $\xi \in (h, \delta/(2M_0))$ one has

$$\int_0^\infty \frac{y^{-1}\rho_{s_*}^2(\ln(y/h))}{y^{2s-1}f_Y^{(0)}(y)}\,\mathrm{d}y \le \frac{c_{16}}{M_0}\big\{M_0^p\big[\ln(M_0/\xi)\big]^{-q-\varkappa} + h^\nu\xi^{-p-\nu} + M_0h^{-p+1}\big\}.$$

Let $\nu_0 \in (0, \nu)$; then we set $\xi = h^{(\nu-\nu_0)/(p+\nu)}$. First, we note that with this choice $\xi \ge h$ as required. Second, it is immediately verified that the second term in the figure brackets on the right hand side of the previous display formula is bounded above by h^{ν_0}, and the first term is dominant as $h \to 0$. Combining this result with (6.12) we conclude that for h small enough

$$\chi^2\big(f_Y^{(0)}, f_Y^{(1)}\big) = c_{16}A^2M^{-1+p}h^{2(\beta+s_*)}\big[\ln(1/h)\big]^{-q-\varkappa},$$

where we took into account that $M_0 = \pi M/4$.

5^0. Now we complete the proof of the theorem. Let

$$h = h_* = \big[c_{17}A^{-2}M^{1-p}(\ln n)^{q+\varkappa}n^{-1}\big]^{1/(2\beta+1+p)}.$$

With this choice and appropriately small constant c_{16} the χ^2-divergence $\chi^2(f_Y^{(0)}, f_Y^{(1)})$ is less than $1/n$, and the hypotheses $f_X = f_X^{(0)}$ and $f_X = f_X^{(1)}$ cannot be distinguished from the observations. Under these circumstances

$$\big|f_X^{(0)}(0) - f_X^{(1)}(0)\big| = c_1Ah_*^{\beta} = c_{18}A^{\frac{p+1}{2\beta+1+p}}\big[M^{1-p}(\ln n)^{q+\varkappa}n^{-1}\big]^{\frac{\beta}{2\beta+1+p}}.$$

This completes the proof.

7. Proofs of auxiliary results

7.1. *Proof of Lemma* 1

Considering the integral (1.2) for $y \ge 0$ and $y < 0$ and using notation (2.3) we obtain

$$f_Y^+(y) = \int_0^\infty \frac{1}{x} f_X^+(y/x)g^+(x)\,\mathrm{d}x - \int_0^\infty \frac{1}{x} f_X^-(y/x)g^-(x)\,\mathrm{d}x \tag{7.1}$$

$$f_Y^-(y) = -\int_0^\infty \frac{1}{x} f_X^+(y/x)g^-(x)\,\mathrm{d}x + \int_0^\infty \frac{1}{x} f_X^-(y/x)g^+(x)\,\mathrm{d}x. \tag{7.2}$$

Applying the Mellin transform to the both sides of (7.1)–(7.2), we have

$$\begin{aligned}\widetilde{f}_Y^+(z) &= \widetilde{f}_X^+(z)\widetilde{g}^+(z) - \widetilde{f}_X^-(z)\widetilde{g}^-(z),\\ \widetilde{f}_Y^-(z) &= -\widetilde{f}_X^+(z)\widetilde{g}^-(z) + \widetilde{f}_X^-(z)\widetilde{g}^+(z).\end{aligned} \tag{7.3}$$

Note that the line $\{z : \mathrm{Re}(z) = 1\}$ is in the strip of analyticity of $\widetilde{f}_X^\pm$ and $\widetilde{g}^\pm$ because f_X and g are probability densities. Thus the Mellin transforms in (7.3) are well-defined in an infinite strip containing the line $\{z : \mathrm{Re}(z) = 1\}$.

The system of equations (7.3) has a unique solution $(\widetilde{f}_X^+(z), \widetilde{f}_X^-(z))$ if and only if

$$\left|\det\begin{bmatrix}\widetilde{g}^+(z) & -\widetilde{g}^-(z)\\ -\widetilde{g}^-(z) & \widetilde{g}^+(z)\end{bmatrix}\right| = \big|\big[\widetilde{g}^+(z)\big]^2 - \big[\widetilde{g}^-(z)\big]^2\big| \ne 0.$$

Under this condition, with $\widetilde{f}_X^+(z)$ and $\widetilde{f}_X^-(z)$ satisfying (7.3) in the common region of analyticity containing the line $\{z : \mathrm{Re}(z) = 1\}$, functions f_X^+ and f_X^- are uniquely determined by the inversion formula

$$f_X^\pm(x) = \frac{1}{2\pi}\int_{-\infty}^\infty x^{-(1+iv)}\widetilde{f}_X^\pm(1+iv)\,\mathrm{d}v.$$

Therefore the necessary and sufficient conditions for identifiability are

$$\widetilde{g}^+(z) - \widetilde{g}^-(z) = \int_0^\infty x^{z-1}\big[g(x) - g(-x)\big]\,\mathrm{d}x \ne 0, \quad \widetilde{g}^+(z) + \widetilde{g}^-(z) \ne 0 \tag{7.4}$$

for almost all z in the common strip of analyticity of $\widetilde{g}^+$ and $\widetilde{g}^-$. Note that $\widetilde{g}^+(z) + \widetilde{g}^-(z)$ is an analytic function; therefore the second condition in (7.4) holds for any density g. Then the statement of the lemma follows from the uniqueness property of the Mellin transform.

7.2. *Proof of Lemma* 2

By (1.2) we have

$$\int_{-\infty}^{\infty} L_{s,h}(x,y) f_Y(y)\,\mathrm{d}y = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} L_{s,h}(x,ty)g(t)\,\mathrm{d}t\right] f_X(y)\,\mathrm{d}y;$$

therefore, in order to prove (3.5) it suffices to show that $L_{s,h}(\cdot,\cdot)$ solves the equation

$$\int_{-\infty}^{\infty} L_{s,h}(x,ty)g(t)\,\mathrm{d}t = K_h(x,y). \tag{7.5}$$

To this end, we will show that for any fixed x the one-sided Mellin transforms of expressions on the both sides of (7.5) coincide in a common vertical strip of the complex plane. This will imply the lemma statement.

It follows from (3.2) that for $x>0$

$$\int_0^{\infty} y^{z-1} K_h(x,y)\,\mathrm{d}y = x^{z-1}\int_{-\infty}^{\infty} K(t)e^{thz}\,\mathrm{d}t = x^{z-1}\check{K}(-zh), \tag{7.6}$$

and for $x<0$

$$\begin{aligned}\int_{-\infty}^{0} (-y)^{z-1} K_h(x,y)\,\mathrm{d}y &= -(-x)^{z-1}\int_{-\infty}^{\infty} K(t)e^{thz}\,\mathrm{d}t\\ &= -(-x)^{z-1}\check{K}(-zh).\end{aligned} \tag{7.7}$$

Let

$$L^{+}_{s,h}(\cdot,y) := \begin{cases} L_{s,h}(\cdot,y), & y>0\\ 0, & y<0,\end{cases} \qquad L^{-}_{s,h}(\cdot,y) := \begin{cases} L_{s,h}(\cdot,-y), & y>0\\ 0, & y<0.\end{cases}$$

Remind that with this notation, $L_{s,h}(\cdot,y) = L^{+}_{s,h}(\cdot,y)$ for $y\ge 0$ and $L_{s,h}(\cdot,y) = L^{-}_{s,h}(\cdot,-y)$ for $y<0$. Integrating the left hand side of (7.5) we obtain

$$\begin{aligned}&\int_0^{\infty} y^{z-1}\int_{-\infty}^{\infty} L_{s,h}(x,ty)g(t)\,\mathrm{d}t\,\mathrm{d}y\\
&\quad= \int_0^{\infty} y^{z-1}\int_{-\infty}^{0} L_{s,h}(x,ty)g(t)\,\mathrm{d}t\,\mathrm{d}y + \int_0^{\infty} y^{z-1}\int_0^{\infty} L_{s,h}(x,ty)g(t)\,\mathrm{d}t\,\mathrm{d}y\\
&\quad= \int_0^{\infty} y^{z-1}\int_0^{\infty} L_{s,h}(x,-ty)g(-t)\,\mathrm{d}t\,\mathrm{d}y + \int_0^{\infty} y^{z-1}\int_0^{\infty} L_{s,h}(x,ty)g(t)\,\mathrm{d}t\,\mathrm{d}y\\
&\quad= \widetilde{L}^{-}_{s,h}(x,z)\widetilde{g}^{-}(1-z) + \widetilde{L}^{+}_{s,h}(x,z)\widetilde{g}^{+}(1-z),\end{aligned}$$

where we denoted $\widetilde{L}^{+}_{s,h}(x,z) = \mathcal{M}[L^{+}_{s,h}(x,\cdot);z]$ and $\widetilde{L}^{-}_{s,h}(x,z) = \mathcal{M}[L^{-}_{s,h}(x,\cdot);z]$. Similarly,

$$\begin{aligned}&\int_{-\infty}^{0} (-y)^{z-1}\int_{-\infty}^{\infty} L_{s,h}(x,ty)g(t)\,\mathrm{d}t\,\mathrm{d}y\\
&\quad= \int_0^{\infty} y^{z-1}\int_{-\infty}^{0} L_{s,h}(x,-ty)g(t)\,\mathrm{d}t\,\mathrm{d}y + \int_0^{\infty} y^{z-1}\int_0^{\infty} L_{s,h}(x,-ty)g(t)\,\mathrm{d}t\,\mathrm{d}y\\
&\quad= \int_0^{\infty} y^{z-1}\int_0^{\infty} L_{s,h}(x,ty)g(-t)\,\mathrm{d}t\,\mathrm{d}y + \int_0^{\infty} y^{z-1}\int_0^{\infty} L_{s,h}(x,-ty)g(t)\,\mathrm{d}t\,\mathrm{d}y\\
&\quad= \widetilde{L}^{+}_{s,h}(x,z)\widetilde{g}^{-}(1-z) + \widetilde{L}^{-}_{s,h}(x,z)\widetilde{g}^{+}(1-z).\end{aligned}$$

Comparing these expressions with (7.6) and (7.7), we set

$$\widetilde{L}^{-}_{s,h}(x,z)\widetilde{g}^{-}(1-z) + \widetilde{L}^{+}_{s,h}(x,z)\widetilde{g}^{+}(1-z) = \begin{cases} x^{z-1}\check{K}(-zh), & x>0,\\ 0, & x<0,\end{cases} \tag{7.8}$$

and

$$\widetilde{L}^{+}_{s,h}(x,z)\widetilde{g}^{-}(1-z)+\widetilde{L}^{-}_{s,h}(x,z)\widetilde{g}^{+}(1-z)=\begin{cases}0, & x>0\\ -(-x)^{z-1}\check{K}(-zh), & x<0.\end{cases} \tag{7.9}$$

It is immediate to verify that solution to equations (7.8)–(7.9) is given by

$$\widetilde{L}^{+}_{s,h}(x,z)=\frac{\check{K}(-zh)}{[\widetilde{g}^{+}(1-z)]^2-[\widetilde{g}^{-}(1-z)]^2}\times\begin{cases}x^{z-1}\widetilde{g}^{+}(1-z), & x>0\\ (-x)^{z-1}\widetilde{g}^{-}(1-z), & x<0,\end{cases}$$

$$\widetilde{L}^{-}_{s,h}(x,z)=\frac{\check{K}(-zh)}{[\widetilde{g}^{+}(1-z)]^2-[\widetilde{g}^{-}(1-z)]^2}\times\begin{cases}x^{z-1}\widetilde{g}^{-}(1-z), & x>0\\ -(-x)^{z-1}\widetilde{g}^{+}(1-z), & x<0.\end{cases}$$

Applying the inverse Mellin transform we obtain

$$L^{+}_{s,h}(x,y)=\frac{1}{2\pi i x}\int_{s-i\infty}^{s+i\infty}\left(\frac{x}{y}\right)^{z}\frac{\check{K}(-zh)\widetilde{g}^{+}(1-z)}{[\widetilde{g}^{+}(1-z)]^2-[\widetilde{g}^{-}(1-z)]^2}\,\mathrm{d}z,\quad x>0,y>0,$$

$$L^{+}_{s,h}(x,y)=-\frac{1}{2\pi i x}\int_{s-i\infty}^{s+i\infty}\left(\frac{-x}{y}\right)^{z}\frac{\check{K}(-zh)\widetilde{g}^{-}(1-z)}{[\widetilde{g}^{+}(1-z)]^2-[\widetilde{g}^{-}(1-z)]^2}\,\mathrm{d}z,\quad x<0,y>0,$$

and

$$L^{-}_{s,h}(x,y)=-\frac{1}{2\pi i x}\int_{s-i\infty}^{s+i\infty}\left(\frac{x}{y}\right)^{z}\frac{\check{K}(-zh)\widetilde{g}^{-}(1-z)}{[\widetilde{g}^{+}(1-z)]^2-[\widetilde{g}^{-}(1-z)]^2}\,\mathrm{d}z,\quad x>0,y>0,$$

$$L^{-}_{s,h}(x,y)=\frac{1}{2\pi i x}\int_{s-i\infty}^{s+i\infty}\left(\frac{-x}{y}\right)^{z}\frac{\check{K}(-zh)\widetilde{g}^{+}(1-z)}{[\widetilde{g}^{+}(1-z)]^2-[\widetilde{g}^{+}(1-z)]^2}\,\mathrm{d}z,\quad x<0,y>0.$$

Comparing these with (3.4) and taking into account that $L_{s,h}(x,y)=L^{+}_{s,h}(x,y)$ when $y\ge 0$ and $L_{s,h}(x,y)=L^{-}_{s,h}(x,-y)$ when $y<0$ for fixed x, we complete the proof.

7.3. *Proof of Lemma* 3

Below $c_1,c_2,\dots$ stand for positive constants depending on ℓ only. By the change of variables, $t=\frac{1}{h}\ln(y/x)$, we have

$$\int\frac{1}{xh}K\left(\frac{\ln(y/x)}{h}\right)f(y)\,\mathrm{d}y-f(x)=\int_{-1}^{1}K(t)\big[w_x(th)-w_x(0)\big]\,\mathrm{d}t,$$

where we have denoted $w_x(t):=e^t f(xe^t)$. Since $f\in\mathscr{H}_{x,r}(A,\beta)$, the function $w_x(\cdot)$ is ℓ times continuously differentiable on $[-\ln r,\ln r]$. Expanding $w_x(\cdot)$ in Taylor's series around zero we have for any $t\in[-\ln r,\ln r]$

$$w_x(t)=w_x(0)+\sum_{k=1}^{\ell-1}\frac{1}{k!}w_x^{(k)}(0)t^k+\frac{1}{\ell!}w_x^{(\ell)}(\xi t)t^{\ell},\quad \xi=\xi(t)\in[0,1].$$

Therefore if $h<\ln r$ then

$$\left|\int_{-1}^{1}K(t)\big[w_x(th)-w_x(0)\big]\,\mathrm{d}t\right|\le\frac{h^{\ell}}{\ell!}\int_{-1}^{1}|t|^{\ell}\big|K(t)\big|\big|w_x^{(\ell)}(\xi th)-w_x^{(\ell)}(0)\big|\,\mathrm{d}t. \tag{7.10}$$

It follows from the Faá di Bruno formula that for $s>0$

$$\begin{aligned}w_x^{(\ell)}(s)&=\sum_{\pi\in\Pi}\big[\big(uf(xu)\big)^{(|\pi|)}\big]_{u=e^s}e^{|\pi|s}\\&=\sum_{\pi\in\Pi}\big[e^s x^{|\pi|}f^{(|\pi|)}\big(xe^s\big)+|\pi|x^{|\pi|-1}f^{(|\pi|-1)}\big(xe^s\big)\big]e^{|\pi|s},\end{aligned}$$

where the summation runs over the set Π of all partitions of the set $\{1, \dots, \ell\}$, and $|\pi|$ is the number of subsets in partition π. Thus

$$\begin{aligned} w_x^{(\ell)}(\xi th) - w_x^{(\ell)}(0) &= \sum_{\pi\in\Pi} e^{(|\pi|+1)\xi th} x^{|\pi|}\big[f^{(|\pi|)}\big(xe^{\xi th}\big) - f^{(|\pi|)}(x)\big] \\ &\quad + \sum_{\pi\in\Pi} x^{|\pi|} f^{(|\pi|)}(x)\big[e^{(|\pi|+1)\xi th} - 1\big] \\ &\quad + \sum_{\pi\in\Pi} e^{|\pi|\xi th}|\pi|x^{|\pi|-1}\big[f^{(|\pi|-1)}\big(xe^{\xi th}\big) - f^{(|\pi|-1)}(x)\big] \\ &\quad + \sum_{\pi\in\Pi}\big(e^{|\pi|\xi th} - 1\big)|\pi|x^{|\pi|-1} f^{(|\pi|-1)}(x). \end{aligned}$$

In view of $f \in \mathscr{H}_{x,r}(A, \beta)$ and by elementary inequality $|e^x - 1| \le |x|e^{|x|}$,

$$\bigg|\sum_{\pi\in\Pi} e^{(|\pi|+1)\xi th} x^{|\pi|}\big[f^{(|\pi|)}\big(xe^{\xi th}\big) - f^{(|\pi|)}(x)\big]\bigg| \le c_1 A|x|^\beta |th|^{\beta-\ell} e^{(\beta+1)|th|},$$

$$\bigg|\sum_{\pi\in\Pi} x^{|\pi|} f^{(|\pi|)}(x)\big[e^{(|\pi|+1)\xi th} - 1\big]\bigg| \le c_2 A|th|e^{(\ell+1)|th|}\sum_{k=1}^{\ell}|x|^k,$$

$$\bigg|\sum_{\pi\in\Pi} e^{|\pi|\xi th}|\pi|x^{|\pi|-1}\big[f^{|\pi|-1}\big(xe^{\xi th}\big) - f^{(|\pi|-1)}(x)\big]\bigg| \le c_3 A|th|e^{\ell|th|}\sum_{k=1}^{\ell}|x|^k,$$

$$\bigg|\sum_{\pi\in\Pi}\big(e^{|\pi|\xi th} - 1\big)|\pi|x^{|\pi|-1} f^{(|\pi|-1)}(x)\bigg| \le c_4 A|th|e^{\ell|th|}\sum_{k=0}^{\ell-1}|x|^k.$$

Combining these inequalities and substituting them in (7.10) completes the proof.

7.4. *Proof of Lemma* 4

We have

$$\begin{aligned} f_Y(y) &= \int_0^\infty \frac{1}{x} f_X(y/x)g(x)\,\mathrm{d}x \\ &= \int_0^y \frac{1}{x} f_X(y/x)g(x)\,\mathrm{d}x + \int_y^\infty \frac{1}{x} f_X(y/x)g(x)\,\mathrm{d}x =: I_1 + I_2. \end{aligned}$$

Using [G3] for any $y \le \delta$ and $p \in [0, 1)$ we obtain

$$\begin{aligned} I_1 &\le C_0 \int_0^y \frac{1}{x} f_X(y/x)x^{-p}|\ln x|^q\,\mathrm{d}x = C_0 y^{-p}\int_1^\infty t^{p-1} f_X(t)\big|\ln(y/t)\big|^q\,\mathrm{d}t \\ &\le 2^{(q-1)_+} C_0 y^{-p}|\ln y|^q\bigg[\int_1^\infty t^{p-1} f_X(t)\,\mathrm{d}t + \int_1^\infty t^{p-1} f_X(t)|\ln t|^q\,\mathrm{d}t\bigg] \\ &\le c_1 y^{-p}|\ln y|^q. \end{aligned} \tag{7.11}$$

Since $f_X(t) \le M$, $\forall t \ge 0$,

$$\begin{aligned} I_2 &\le M\int_y^\infty \frac{g(x)}{x}\,\mathrm{d}x \le M\delta^{-1} + M\int_y^\delta \frac{g(x)}{x}\,\mathrm{d}x \\ &\le M\delta^{-1} + C_0 M\int_y^\delta x^{-p-1}|\ln x|^q\,\mathrm{d}x. \end{aligned} \tag{7.12}$$

If $p = 0$ then the last integral on the right hand side is bounded from above by $|\ln y|^{q+1}$, and

$$I_2 \le M\delta^{-1} + MC_0|\ln y|^{q+1}.$$

This inequality together with (7.11) completes the proof of statement (i).

Now we bound the expression on the right hand side of (7.12) in the case $p \in (0, 1)$. Using the following formula (see, e.g., [9, 616.2])

$$\int x^{p-1}(\ln x)^q \,\mathrm{d}x = \frac{1}{p}x^p(\ln x)^q - \frac{q}{p}\int x^{p-1}(\ln x)^{q-1}\,\mathrm{d}x, \quad \forall p \ne 0, q \ne -1,$$

we obtain

$$\int_y^\delta x^{-p-1}|\ln x|^q \,\mathrm{d}x = \int_{1/\delta}^{1/y} t^{p-1}(\ln t)^q \,\mathrm{d}t = \left[\frac{t^p(\ln t)^q}{p}\right]_{1/\delta}^{1/y} - \frac{q}{p}\int_{1/\delta}^{1/y} t^{p-1}(\ln t)^{q-1}\,\mathrm{d}t. \tag{7.13}$$

Hence it follows from (7.13) that

$$\int_y^\delta x^{p-1}|\ln x|^q \,\mathrm{d}x \le \left[\frac{t^p(\ln t)^q}{p}\right]_{1/\delta}^{1/y} \le \frac{y^{-p}}{p}\big[\ln(1/y)\big]^q. \tag{7.14}$$

Thus, we obtain

$$I_2 \le M\delta^{-1} + C_0Mp^{-1}y^{-p}|\ln y|^q.$$

Combining this inequality with (7.11) we complete the proof of (ii).

Acknowledgements

This article was prepared within the framework of the HSE University Basic Research Program and funded by the Russian Academic Excellence Project 5-100. Results of Sections 4 and 5 have been obtained under support of the RSF grant No. 18-11-00132. The first author gratefully acknowledges the financial support by the German Research Foundation (DFG) through the Collaborative Research Center 832. The second author gratefully acknowledges the financial support by the Israel Science Foundation (ISF) research grant 361/15. The authors are grateful to an anonymous referee for careful reading and insightful comments that lead to substantial improvements in the paper.

References

[1] K. A. Andersen and M. B. Hansen. Multiplicative censoring: Density estimation by a series expansion approach. *J. Statist. Plann. Inference* **98** (2001) 137–155. MR1860231 https://doi.org/10.1016/S0378-3758(00)00237-8
[2] R. S. Anderssen. On the use of linear functionals for Abel–type integral equations. In *The Application and Numerical Solution of Integral Equations* 195–221. R. Anderssen, F. De Hoog and M. Lucas (Eds). Sijthoff and Noordhof International Publishers, 1980. MR0582991
[3] G. E. Andrews, R. Askey and R. Roy. *Special Functions*, **71**. Cambridge University Press, Cambridge, 1999. MR1688958 https://doi.org/10.1017/CBO9781107325937
[4] M. Asgharian, M. Carone and V. Fakoor. Large–sample study of the kernel density estimators under multiplicative censoring. *Ann. Statist.* **40** (2012) 159–187. MR3013183 https://doi.org/10.1214/11-AOS954
[5] D. Belomestny, F. Comte and V. Genon-Catalot. Nonparametric Laguerre estimation in the multiplicative censoring model. *Electron. J. Stat.* **10** (2) (2016) 3114–3152. MR3571964 https://doi.org/10.1214/16-EJS1203
[6] D. Belomestny and J. Schoenmakers. Statistical Skorohod embedding problem: Optimality and asymptotic normality. *Statist. Probab. Lett.* **104** (2015) 169–180. MR3360720 https://doi.org/10.1016/j.spl.2015.05.015
[7] E. Brunel, F. Comte and V. Genon-Catalot. Nonparametric density and suivival function estimation in the multiplicative censoring model. *TEST* **25** (2016) 570–590. MR3531843 https://doi.org/10.1007/s11749-016-0479-1
[8] F. Comte and C. Dion. Nonparametric estimation in a multiplicative censoring model with symmetric noise. *J. Nonparametr. Stat.* **28** (2016) 768–801. MR3555457 https://doi.org/10.1080/10485252.2016.1225737
[9] H. B. Dwight. *Tables of Integrals and Other Mathematical Data*, 4th edition. The Macmillan Company, 1961. MR0129577
[10] J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** (1991) 1257–1272. MR1126324 https://doi.org/10.1214/aos/1176348248
[11] M. V. Fedoryuk. *Asymptotics. Integrals and Series*. Nauka, Moscow, 1987. (in Russian). MR0950167
[12] M. Goldberg. A method of adjoints for solving some ill-posed equations of the first kind. *Appl. Math. Comput.* **5** (1979) 123–130. MR0522852 https://doi.org/10.1016/0096-3003(79)90002-X

[13] S. J. Iturria, R. J. Carrol and D. Firth. Polynomial regression and estimating functions in the presence of multiplicative measurement error. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** (3) (1999) 547–561. MR1707860 https://doi.org/10.1111/1467-9868.00192

[14] G. Kerkyacharian, O. Lepski and D. Picard. Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121** (2) (2001) 137–170. MR1863916 https://doi.org/10.1007/PL00008800

[15] W.-L. Loh and C.-H. Zhang. Estimating mixing densities in exponential family models for discrete variables. *Scand. J. Stat.* **24** (1997) 15–32. MR1436620 https://doi.org/10.1111/1467-9469.t01-1-00046

[16] R. B. Paris and D. Kaminski. *Asymptotics and Mellin–Barnes Integrals*, **85**. Cambridge University Press, Cambridge, 2001. MR1854469 https://doi.org/10.1017/CBO9780511546662

[17] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York, 2009. MR2724359 https://doi.org/10.1007/b13794

[18] B. Van Es and P. Spreij. Estimation of a multivariate stochastic volatility density by kernel deconvolution. *J. Multivariate Anal.* **102** (3) (2011) 683–697. MR2755024 https://doi.org/10.1016/j.jmva.2010.12.003

[19] B. Van Es, P. Spreij and H. Van Zanten. Nonparametric volatility density estimation. *Bernoulli* **9** (3) (2003) 451–465. MR1997492 https://doi.org/10.3150/bj/1065444813

[20] Y. Vardi. Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika* **76** (4) (1989) 751–761. MR1041420 https://doi.org/10.1093/biomet/76.4.751

[21] Y. Vardi and C.-H. Zhang. Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.* **20** (2) (1992) 1022–1039. MR1165604 https://doi.org/10.1214/aos/1176348668

[22] R. Wong. *Asymptotic Approximations of Integrals*. Society for Industrial and Applied Mathematics, Philadelphia, 2001. MR1851050 https://doi.org/10.1137/1.9780898719260

[23] C.-H. Zhang. Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** (1990) 806–831. MR1056338 https://doi.org/10.1214/aos/1176347627

[24] C.-H. Zhang. On estimating mixing densities in discrete exponential family models. *Ann. Statist.* **23** (1995) 929–945. MR1345207 https://doi.org/10.1214/aos/1176324629