Ann Inst Stat Math (2016) 68:25–75
DOI 10.1007/s10463-014-0487-4

Minimax theory of nonparametric hazard rate estimation: efficiency and adaptation

Sam Efromovich

Received: 21 January 2014 / Revised: 4 June 2014 / Published online: 30 September 2014

Abstract The problem of nonparametric estimation of the hazard rate function is considered and the theory of sharp minimax estimation for two global and two local Sobolev classes is developed. Several interesting outcomes are as follows: (i) Classical global and local function classes imply different sharp constants of the MISE convergence. This is in contrary to the density estimation where sharp constants are the same. (ii) Two global classes imply different sharp constants and correspondingly require using different linear estimates. (iii) Two local classes imply the same sharp constant, and nonetheless require different linear estimates to attain this constant. (iv) A sharp-minimax data-driven estimator is proposed that adapts to the smoothness of the hazard rate and to an unknown underlying function class, and it is tested for small samples via a numerical study.

Keywords Asymptotic · Coefficient of difficulty · Global and local minimax · MISE · Small sample

1 Introduction

Let X be a nonnegative continuous random variable. It can be a lifetime, or the time to an event of interest (which can be the time of failure of a device, or the time of an illness relapse, or the time of repairing a strain break in DNK), or an insurance loss, or a commodity price. In all these cases, it is of interest to assess the risk associated with X via the so-called *hazard rate* function

S. Efromovich (✉)
Department of Mathematical Sciences, The University of Texas at Dallas, Richardson, TX 75080, USA
e-mail: efrom@utdallas.edu

$$h(x) := \lim_{v \to 0} \frac{\Pr(x \leq X \leq x + v | X \geq x)}{v} = \frac{p(x)}{G(x)}, \quad G(x) > 0, \quad x \geq 0, \tag{1}$$

where $p(x)$ is the probability density of X and $G(x) := \int_x^\infty p(u)\mathrm{d}u$ is the survivor function. If one thinks about X as a time to an event-of-interest, then $h(x)\mathrm{d}x$ represents the instantaneous likelihood that the event occurs within the interval $(x, x + \mathrm{d}x)$ given that the event has not occurred at time x. The hazard rate quantifies the trajectory of imminent risk, and it may be referred to by other names in different sciences, for instance as the failure rate in reliability theory and the force of mortality in sociology.

Let us remember some classical properties and examples of the hazard rate that will be used in the paper. The hazard rate, similarly to the probability density or the survivor function, characterizes the random variable X. Namely, if the hazard rate is known, then the corresponding probability density is

$$p(x) = h(x)\mathrm{e}^{-\int_0^x h(v)\mathrm{d}v}, \tag{2}$$

and the survivor function is

$$G(x) = \mathrm{e}^{-\int_0^x h(v)\mathrm{d}v}. \tag{3}$$

The preceding identity follows from integrating both sides of $h(x) = -[\mathrm{d}G(x)/\mathrm{d}x]/G(x)$ and then using $G(0) = 1$, and (2) follows from (1) and the verified (3). A familiar example is the constant hazard rate of an exponential random variable (the rate is equal to the reciprocal of the mean), and inverse is also valid—a constant hazard rate implies exponential distribution. A constant hazard rate has coined the name *memoryless* for exponential distribution. Another interesting example is Weibull distribution where, depending on the shape parameter, the hazard rate can decrease or increase. One of the important properties of the hazard rate is that hazard rate of the minimum of two independent random variables is equal to the sum of their hazard rates. As a result, we can conclude that there is no bona fide restriction on the shape of a hazard rate function. On the other hand, similarly to the probability density, the hazard rate is nonnegative and has the same smoothness as the corresponding density. The last but not the least remark is that all these examples and properties can be rephrased and correspondingly referred to when one is interested in the study of a hazard rate over some interval, say $[a, a + 1]$ with $a \geq 0$ and $G(a) > 0$. Indeed, for $x \in [a, a + 1]$ identity (3) yields $G(x) = G(a)\mathrm{e}^{-\int_a^x h(v)\mathrm{d}v}$, and this explains the remark made. For instance, if $h(x) = C$ is constant on $[a, a+1]$ then $p(a+z) = CG(a)\mathrm{e}^{-Cz}$, $z \in [0, 1]$ and we may refer to the distribution (random variable) as memoryless on the interval. Because the hazard rate is not integrable on its support, that is a hazard rate must satisfy $\int_0^\infty h(x)\mathrm{d}x = \infty$, it is natural to study hazard rate estimation over a finite interval, which without loss of generality can be $[a, a + 1]$, $a \geq 0$.

Now we are in a position to formulate the aim of this paper. Based on a sample $X_1, X_2, \ldots, X_n$ of size n from the random variable of interest X, we would like to: (i) Estimate hazard rate $h(x)$ of X over an interval $[a, a + 1]$, $a \geq 0$ under a minimax mean integrated squared error (MISE) criterion; (ii) Find sharp lower bounds for a minimax MISE; (iii) Propose a data-driven estimate that attains lower bounds—a so-called adaptive and efficient estimate.

Nonparametric estimation of the hazard rate is a familiar topic in the literature. Different types of estimators, including kernel, classical orthogonal series and modern wavelet methods, have been proposed. A number of adaptive, to smoothness of an underlying hazard rate function, procedures motivated by known ones for the probability density have been developed. The interested reader can find a relevant discussion and thorough reviews in a number of classical and more recent publications including Watson and Leadbetter (1964), Rice and Rosenblatt (1976), Prakasa Rao (1983), Cox and Oakes (1984), Silverman (1986), Patil (1993, 1997), Gonzalez-Manteiga et al. (1996), Dögler and Rüschendorf (2002), Wu and Wells (2003), Huber and MacGibbon (2004), Wang (2005), Müller and Wang (2007). Interesting results, including both estimation and testing, have been obtained for the case of known restrictions on the shape of hazard rate, see a discussion in Jankowski and Wellner (2009).

The literature review shows that, despite the interest in developing the theory of nonparametric hazard rate estimation, so far all publications have been devoted to either consistent or rate-minimax estimation of the hazard rate with no results about sharp minimax estimation when the best rate and constant are studied simultaneously. Why? This is a good question because theory of a closely related problem of sharp minimax estimation of the probability density has been known for almost 30 years. Namely, a sharp minimax lower bound for estimation of differentiable densities under the MISE criterion was obtained in Efromovich and Pinsker (1982) for a global Sobolev function class. Furthermore, Pinsker (1980)'s estimator, proposed for the filtering model and based on a known underlying Sobolev class, has happened to be also sharp minimax for the density. A bit later Efromovich (1985) suggested a data-driven estimator that attained the sharp constant. Because the estimator used data to adjust its performance to unknown smoothness of an underlying density, it was called adaptive. Golubev (1991) made an important addition to those results by showing that even if a global Sobolev class is replaced by a local one, where all considered densities are near (in L_∞-norm) a given pivotal density, the sharp minimax lower bound remains the same. A practical significance of his result is that the pivot does not affect the sharp constant and known sharp estimators are still optimal under the local minimax.

Knowing these theoretical results for the density, together with the fact that the hazard rate is the density divided by the survivor function (which is smoother than the density and can be estimated with the parametric rate), it is reasonable to hope that creating a sharp minimax theory for the hazard rate is a straightforward undertaking. The paper shows that this is not the case. The main issue arises when one wants to follow the approach of Efromovich and Pinsker (1982) for obtaining a lower bound for the minimax MISE. The approach yields a bound which is too low for all known estimators, and then either the lower bound should be increased or a new estimator is proposed. The paper shows that the former is the way to solve the issue. Remember that a minimax lower bound is obtained via replacing a minimax risk by an appropriate Bayes one, and the prior distribution used in the density case has happened to be too simplistic for the hazard rate due to the presence of the "nuisance" survivor function. Furthermore, that relatively smooth and related to the estimand nuisance function (the survival function) creates several interesting phenomena unknown in the density estimation literature. To warm up the reader, let us present some. (i) Sharp constants are different for the Pinsker's global and Golubev's local function classes while they are the same for the probability density (as well as regression and spectral density) setting. In other words, the pivot affects the sharp constant and then requires using special shrinkage coefficients in the Pinsker's estimator. Let us stress that no longer the same Pinsker's estimator is sharp minimax for these two classes. (ii) Different sharp constants are obtained for two global Sobolev classes that are often used interchangeably when the minimax rate is of interest. This shows that sharp-minimax estimation of the hazard rate is sensitive to minor changes in an underlying functional class. (iii) Two local Sobolev classes imply the same sharp minimax constant but require using different estimates to attain the constant. Furthermore, for one local class, the Pinsker estimator is not even rate minimax. This is the first known example when this estimator fails to deliver a minimax convergence. Other interesting findings will be presented shortly.

As it is clear from the above-mentioned results, to create a feasible minimax theory for the hazard rate, it is prudent to consider an array of functional classes with the same smoothness, here the same number $\alpha \in \{1, 2, \dots\}$ of derivatives. Specifically, two global and two local function classes, defined in Sect. 2, will be considered. Furthermore, let us explain how a game theory can help us to understand a minimax approach applied to an array of functional classes. Rules of the corresponding minimax game are as follows. There are three players: the dealer, nature and the statistician. The game is defined by: (a) An array $\{\mathcal{S}_i(P),\ P \in \mathcal{P}_i, i = 1, 2, \dots\}$ of function classes $\mathcal{S}_i$; (b) The set of assumptions (rules) $\mathcal{P}_i$ about possible parameters P for each class; (c) The risk used to measure results of the game. This information is known to all players. The game begins with the dealer dealing a particular function class and its parameters, satisfying rules of the game, to nature. Then for each n nature chooses a hazard rate h from the dealt class and generates a sample of size n from this hazard rate. The dealer and the statistician, using the sample, estimate h. The dealer knows everything apart of an underlying hazard rate h chosen by nature, the statistician knows the sample and all rules of the game. Nature tries to select the most difficult hazard rate h for estimation, and the dealer and the statistician try to estimate it with the smallest risk.

Now let us complement this definition of the minimax game by several comments.

Remark 1 The dealer has an advantage of knowing the dealt class and its parameters and, therefore, the dealer's risk may serve as a lower bound (benchmark) for the statistician. All known in the literature minimax nonparametric lower bounds are bounds for dealer-estimators, that is, estimators proposed by the dealer. A lower bound is called minimax (more precisely asymptotically rate- or sharp-minimax depending on either only rate or both rate and constant are studied) if it is attained, as the sample size n increases, by a dealer-estimate. A familiar example is the rate-minimax bound $n^{-2\alpha/(2\alpha+1)}$ for the MISE convergence of α-fold differentiable functions, with the most traditional assumption being the twice-differentiability ($\alpha = 2$ and the rate is $n^{-4/5}$); see Efromovich (1999) and, for the hazard rate case, a review in Wang (2005). The literature may refer to those procedures as estimates, but in our terminology they should be referred to as dealer-estimates. If the statistician can suggest its own estimate (that is a data-driven estimate) which attains the dealer's lower bound, then the estimate is called rate- or sharp-minimax adaptive, and in the case of sharp-minimaxity the estimate can be also referred to as efficient. Of course, the art of obtaining a meaningful lower bound is to propose a dealer whose performance can be matched by the statistician. We shall continue this discussion shortly in Remark 2.

Remark 2 Let us make two more comments. Under a minimax approach, a feasible choice of an array of functional classes is important. One of the familiar criticisms of the minimax methodology is that it cares only about the best strategy of nature (here the worst hazard rate for estimation); see Lehmann (1983). To overcome this drawback and to make a minimax methodology more practical, in parametric inference it has been proposed to consider both a global minimax over a large set of parameters and a local minimax over a sequence of classes converging (in some norm) to a pivotal parameter; see a discussion in Ibragimov and Khasminskii (1981). In what follows we will use this approach and expand it upon the nonparametric inference. The second comment is about the notion of adaptivity. As it has been explained in Remark 1, this notion is traditionally used for an estimator that adapts its performace toward unknown parameters of an underlying function class (like the smoothness of an estimated curve). So far there has been no need to consider a broader notion of adaptation because for density and regression settings, dealer's lower bounds for global and local minimax approaches have coincided. As we shall see shortly in Sect. 2, this is not the case for hazard rate estimation. As a result, a data-driven hazard rate estimator should adapt to an unknown underlying function class (the new task) and unknown parameters of that class (the traditional task). One of the ways to solve the problem is via aggregation of minimax procedures for all considered classes, but the paper proposes a much simpler approach.

Remark 3 If a spectator is allowed to observe a minimax game and to know both the dealt function class and the hazard rate chosen by nature, then the spectator has an advantage over the dealer. In the nonparametric literature, thanks to David Donoho and Ian Johnstone, a spectator is called an oracle, and oracle's estimates may inspire construction of a good data-driven estimator. As we shall see shortly, the oracle approach becomes handy when a data-driven estimator should adapt to an array of function classes.

The paper is organized as follows. Section 2 is devoted to lower bounds for four function classes motivated by the same Sobolev smoothness. Section 3 is devoted to upper bounds. Section 4 is devoted to a numerical study. Section 5 presents discussion of obtained results. Proofs are deferred to Sect. 6.

2 Lower minimax bounds

Lower minimax bounds are calculated for two global and two local Sobolev classes. All these classes have the same Sobolev's order α of smoothness (here and in what follows α is a positive integer and the considered functions are α-fold differentiable), and this will allow us to analyze the effect of a functional class on sharp-minimax estimation. The studied risk is the MISE over the unit interval $[a, a+1]$, $a \geq 0$ (remember our discussion in the Introduction). Remember that for α-fold differentiable functions the familiar minimax rate $n^{-2\alpha/(2\alpha+1)}$ of the MISE convergence is well known for different statistical models including hazard rate, density and regression; see Ibragimov and Khasminskii (1981), Silverman (1986), Efromovich (1999), Zhao (2000), Wang (2005), and Johnstone (2011).

In what follows $\varphi_0(x) = 1$ and $\varphi_j(x) = 2^{1/2}\cos(\pi j(x-a))$, $j > 0$ are elements of the classical cosine basis on the interval ($[a, a+1]$), α is a positive integer number, $\theta_j = \int_a^{a+1} \varphi_j(x)h(x)\mathrm{d}x$ are Fourier coefficients of $h(x)$ on $[a, a+1]$. $E_h\{\cdot\}$ denotes the expectation given a hazard rate function h (remember that the hazard rate characterizes a random variable), and $o_s(1)$ are generic sequences that are vanishing as $s \to \infty$.

The first function class is a classical global Sobolev class of order α,

$$\begin{aligned}
&\mathcal{S}_1(\alpha, Q_0, Q, G(a)) \\
&\quad = \Bigg\{ h:\ \ h(x) = \sum_{j=0}^{\infty} \theta_j \varphi_j(x),\ x \in [a, a+1];\ h(x) \geq 0,\ x \geq 0; \\
&\qquad \mathrm{e}^{-\int_0^a h(x)\mathrm{d}x} = G(a);\quad \sum_{j=1}^{\infty} (\pi j)^{2\alpha} \theta_j^2 \leq Q < \infty,\ 0 < \theta_0 \leq Q_0 < \infty \Bigg\}. \qquad (4)
\end{aligned}$$

Note that $G(a)$ is the value of the survivor function at the left end of the studied unit interval, and if $a = 0$ then $G(0) = 1$. In what follows $G^{-k}(x) := [G(x)]^{-k}$.

Theorem 1 (Global Minimax I) *The following lower bound for the minimax MISE of dealer-estimators is valid*:

$$\begin{aligned}
&\inf_{\check{h}^*} \sup_{h \in \mathcal{S}_1(\alpha, Q_0, Q, G(a))} E_h \left\{ \int_a^{a+1} (\check{h}^*(x) - h(x))^2 \mathrm{d}x \right\} \\
&\quad \geq P(\alpha, Q)([G^{-1}(a)(\mathrm{e}^{Q_0} - 1)]n^{-1})^{2\alpha/(2\alpha+1)}(1 + o_n(1)), \qquad (5)
\end{aligned}$$

where the infimum is taken over all possible dealer-estimators $\check{h}^*$*, based on a sample* $X_1, \ldots, X_n$ *from a distribution with the hazard rate* h *and parameters* $(\alpha, Q_0, Q, G(a))$*, and*

$$P(u, v) := v^{1/(2u+1)}(2u + 1)^{1/(2u+1)} \left[\frac{u}{\pi(u + 1)} \right]^{2u/(2u+1)}. \qquad (6)$$

Following the terminology of Efromovich (1999), $P(u, v)$ is referred to as the Pinsker function, and factor $G^{-1}(a)(\mathrm{e}^{Q_0} - 1)$ is called the coefficient of difficulty of estimation of the hazard rate over the interval $[a, a + 1]$ for the global Sobolev class.

Now let us introduce a second global functional class which is also well known in the literature:

$$\mathcal{S}_2(\alpha, c_1, c_2, Q, G(a))$$
$$= \left\{ h : \ h(x) = \sum_{j=0}^{\infty} \theta_j \varphi_j(x), \ x \in [a, a+1]; \ \ \mathrm{e}^{-\int_0^a h(x)\mathrm{d}x} = G(a), \ 0 < G_0(a) \right.$$
$$\left. \leq 1, h(x) \geq 0, \ x \geq 0; \ \sum_{j=0}^{\infty} [c_1 + c_2(\pi j)^{2\alpha}]\theta_j^2 \leq Q < \infty, \ c_1 > 0, \ c_2 > 0 \right\}. \quad (7)$$

The difference between the two global Sobolev classes is that the latter treats θ_0 (the average value of the hazard rate over $[a, a+1]$) similarly to other Fourier coefficients.

Set Q_0^* to be a positive root of the equation $g(z) = Q$ where $g(z) := c_1 z(z - \alpha^{-1} + \alpha^{-1}\mathrm{e}^z)$. Function $g(z)$ is strictly increasing in $z > 0$ and, therefore, Q_0^* is unique and well defined.

Theorem 2 (Global Minimax II) *The following lower bound for the minimax MISE of dealer-estimators is valid:*

$$\inf_{\check{h}^*} \sup_{h \in \mathcal{S}_2(\alpha, c_1, c_2, Q, G(a))} E_h \left\{ \int_a^{a+1} (\check{h}^*(x) - h(x))^2 \mathrm{d}x \right\}$$
$$\geq P(\alpha, [Q - c_1(Q_0^*)^2]/c_2)([G^{-1}(a)(\mathrm{e}^{Q_0^*} - 1)]n^{-1})^{2\alpha/(2\alpha+1)}(1 + o_n(1)), \quad (8)$$

where the infimum is taken over all possible dealer-estimators $\check{h}^*$ based on a sample $X_1, \ldots, X_n$ from a distribution with the hazard rate $h(x)$ and parameters $(\alpha, c_1, c_2, Q, G(a))$, and function $P(u, v)$ is defined in (2.3).

Note the interesting structure of the sharp constant in (8). To the best of the author's knowledge, this is the first known sharp result for the global Sobolev class (7).

Now let us consider two local classes where estimated functions are not far from a pivotal one. We begin with a local Sobolev class whose motivation goes back to Golubev (1991) where the density estimation has been considered. For our setting let $h_0(x)$ denote the pivot chosen by the dealer, and define a Golubev's local Sobolev class as

$$\mathcal{S}_3(\alpha, Q, h_0, \rho, \beta)$$
$$:= \left\{ h : \quad h(x) = h_0(x) + \sum_{j=1}^{\infty} \theta_j \varphi_j(x) I(x \in [a, a+1]) \ ; \right.$$
$$\sum_{j=1}^{\infty} (\pi j)^{2\alpha} \theta_j^2 \leq Q < \infty, \quad \sup_{x \in [a, a+1]} \left| \sum_{j=1}^{\infty} \theta_j \varphi_j(x) \right| < \rho; \ \ \mathrm{e}^{-\int_0^a h_0(v)\mathrm{d}v}$$
$$=: G_0(a), \ \ 0 < G(a) \leq 1 \ ;$$
$$\left. \inf_{x \in [a, a+1]} h_0(x) \geq \rho > 0; \ \sum_{j=1}^{\infty} j^{2\alpha+\beta} \left[\int_a^{a+1} h_0(x)\varphi_j(x)\mathrm{d}x \right]^2 < \infty, \beta > 0 \right\}. \quad (9)$$

The underlying idea of this class is that all considered functions are not farther than ρ in $L_\infty([a, a+1])$-norm from the pivot. This is why the class is called local. The last line in (9) indicates that the pivot should be smoother than a "regular" function from the class. Golubev (1991) shows that the pivot does not affect the sharp constant in the probability density, regression and spectral density estimation problems. Next theorem shows that the pivot does affect the constant in the hazard rate case.

Theorem 3 (Golubev's Local Minimax) *The following lower bound for the Golubev local minimax MISE holds:*

$$\inf_{\check{h}^*} \sup_{h \in \mathcal{S}_3(\alpha, Q, h_0, \rho, \beta)} E_h \left\{ \int_a^{a+1} (\check{h}^*(x) - h(x))^2 \mathrm{d}x \right\}$$
$$\geq P(\alpha, Q)([G_0^{-1}(a)(\mathrm{e}^{\int_a^{a+1} h_0(u)\mathrm{d}u} - 1)]n^{-1})^{2\alpha/(2\alpha+1)}(1 + o_n(1)), \quad (10)$$

where the infimum is taken over all possible dealer-estimators $\check{h}^*$ based on a sample $X_1, \ldots, X_n$ from a distribution with the hazard rate $h(x)$ and parameters $(\alpha, Q, h_0, \rho, \beta)$, and $P(\alpha, Q)$ is defined in (6).

There are several issues, related to the classical Golubev's local minimax, that we would like to explore further. First, we are still far from the ideal (remember Remark 2) case of a single underlying hazard rate function. What will be if a low-frequency part of the pivot is also the low-frequency part of an underlying hazard rate? Also, in (9) the parameter ρ, controlling how close a hazard rate must be to the pivot, may be as small as desired, but can it vanish as sample size increases? Also, can the pivot be less smooth than a regular function from a local class? A weakly restricted local Sobolev class, defined below, takes into account these issues. Set

$$\begin{aligned}
&\mathcal{S}_4(\alpha, Q, h_0, \rho_n, M_n) \\
&:= \Bigg\{ h : \ h(x) = h_0(x) I(x \notin [a, a+1]) + \Bigg[\sum_{j=0}^{M_n - 1} \int_a^{a+1} h_0(u)\varphi_j(u)\mathrm{d}u \varphi_j(x) \\
&\quad + \sum_{j \geq M_n} \theta_j \varphi_j(x) \Bigg] I(x \in [a, a+1]), h(x) \geq 0 \, ; \\
&\quad \times \sum_{j \geq M_n} (\pi j)^{2\alpha} \theta_j^2 \leq Q < \infty, \quad \sup_{x \in [a, a+1]} \Bigg| \sum_{j \geq M_n} \theta_j \varphi_j(x) \Bigg| < \rho_n \, ; \\
&\quad \times \mathrm{e}^{-\int_0^a h_0(v)\mathrm{d}v} =: G_0(a) > 0, \ \sum_{j=0}^{\infty} \Bigg| \int_a^{a+1} \varphi_j(x) h_0(x) \mathrm{d}x \Bigg| < \infty, \quad \min_{x \in [a, a+1]} h_0(x) > 0 \, ; \\
&\quad \times \rho_n = o_n(1), \ \rho_n \geq \ln(n) n^{-\alpha/(2\alpha+1)}; \quad M_n^{-1} = o_n(1), \ 1 \leq M_n \leq n^{1/(2\alpha+1)} / \ln^2(n) \Bigg\}.
\end{aligned} \quad (11)$$

Let us comment on the weakly restricted local class. Low-frequency (in Fourier frequency domain) parts of all functions from this class are equal to the low-frequency part of the pivot and, therefore, known to the dealer. The cardinality of known low-frequency Fourier coefficients is M_n and it increases to infinity as $n \to \infty$. Note that M_n may increase as fast as $n^{1/(2\alpha+1)}/\ln^2(n)$. Further, all functions from the class are not farther than $\rho_n' := \rho_n + 2^{1/2}\sum_{j=M_n}^{\infty} |\int_a^{a+1} \varphi_j(v)h_0(v)\mathrm{d}v|$, in $L_\infty([a, a+1])$-norm, from the pivot. According to the third and fourth lines in (11), $\rho_n' = o_n(1)$, and thus the class shrinks toward the pivot. Finally, the third line in (11) imposes a rather mild restriction on smoothness of the pivot, in particular any Lipschitz function of order $1/2 + \rho$, $\rho > 0$ satisfies the restriction.

Theorem 4 (Weakly Restricted Local Minimax) *The following lower bound for the weakly restricted local minimax MISE holds*:

$$\inf_{\check{h}^*} \sup_{h \in \mathcal{S}_4(\alpha, Q, h_0, \rho_n, M_n)} E_h\left\{\int_a^{a+1} (\check{h}^*(x) - h(x))^2 \mathrm{d}x\right\}$$
$$\geq P(\alpha, Q)([G_0^{-1}(a)(\mathrm{e}^{\int_a^{a+1} h_0(u)\mathrm{d}u} - 1)]n^{-1})^{2\alpha/(2\alpha+1)}(1 + o_n(1)), \quad (12)$$

where the infimum is taken over all possible dealer-estimators $\check{h}^*$ based on a sample $X_1, \ldots, X_n$ from a distribution with the hazard rate h and parameters $(\alpha, Q, h_0, \rho_n, M_n)$, and $P(\alpha, Q)$ is defined in (6).

The important outcome of this theorem is that despite the fact that the weakly restricted local Sobolev class shrinks toward the pivot, all its functions have the same low-frequency component known to the dealer, and the pivot may be rougher than a typical function from the class, the sharp minimax constant is the same as for the Golubev's local class. This answers the above-raised questions about the Golubev's class, but then is this function class of any interest on its own? As we shall see shortly, the answer is "yes" because for the weakly restricted class the dealer no longer can use a Pinsker's estimate to attain the sharp constant.

3 Efficient and adaptive estimation

The aim of this section is twofold. First, we are verifying that lower bounds for the MISE of dealer-estimators, presented in Sect. 2, are sharp, that is, they are attainable by dealer-estimators. Second, a data-driven estimator is proposed which adapts to both the smoothness of an underlying hazard rate and an underlying functional class. That estimator will be constructed using an oracle approach and the blockwise-shrinkage methodology.

We begin with verification that all four lower bounds of Sect. 2 are sharp and attainable by dealer-estimators. Following the methodology of Efromovich and Pinsker (1982), introduce a family of Pinsker's dealer-estimators of an underlying hazard rate $h(x)$ for $x \in [a, a+1]$,

$$\check{h}(x, J(n), \alpha, \{\hat{\theta}_j\}) := \sum_{j=0}^{J(n)} [1 - (j/J(n))^\alpha]\hat{\theta}_j \varphi_j(x), \quad J(n) \in \{0, 1, \ldots, n\}. \quad (13)$$

Here $\varphi_0(x) = 1, \varphi_j(x) = \sqrt{2}\cos(\pi j(x-a)), j = 1, 2, \ldots$ are elements of the cosine basis on $[a, a+1]$,

$$\hat{\theta}_j := \sum_{l=1}^{n} \varphi_j(X_l)\eta_l^{-1} I(X_l \in [a, a+1]) \tag{14}$$

are estimates of Fourier coefficients $\theta_j := \int_a^{a+1} h(x)\varphi_j(x)\mathrm{d}x = E_h\{\varphi_j(X)G^{-1}(X)I(X \in [a, a+1])\}$ of an underlying hazard rate motivated by the plug-in sample mean estimator,

$$\eta_l := \sum_{s=1}^{n} I(X_s \geq X_l) \tag{15}$$

is the antirank of X_l (note that it is at least 1 and, therefore, its reciprocal always exists, and η_l/n is a feasible estimate of $G(X_l)$), and $I(\cdot)$ is the indicator. To get a particular Pinsker's estimator, the dealer should specify $J(n)$. Let us stress that so far for all settings, involving the probability density, regression and spectral density, a Pinsker's dealer-estimator (13) has implied sharp-minimax estimation whenever $\hat{\theta}_j$ was chosen correctly. Set

$$b_n := [n(2\alpha+1)(\alpha+1)/(\alpha\pi^{2\alpha})]^{1/(2\alpha+1)}, \tag{16}$$

remember that $\lfloor z \rfloor$ denotes the largest integer smaller than z, Q_0^* is defined in the paragraph above Theorem 2, and define specific dealer-estimators for four functional classes $\mathcal{S}_i, i = 1, \ldots, 4$ considered in Sect. 2,

$$\check{h}_1(x) := \check{h}(x, J_1, \alpha, \{\hat{\theta}_j\}) \quad \text{where} \quad J_1 := \lfloor [QG(a))/(\mathrm{e}^{Q_0} - 1)]^{1/(2\alpha+1)} b_n \rfloor, \tag{17}$$

$$\check{h}_2(x) := \check{h}(x, J_2, \alpha, \{\hat{\theta}_j\}) \quad \text{where} \quad J_2 := \lfloor [(Q - c_1(Q_0^*)^2)c_2^{-1}G(a)/(\mathrm{e}^{Q_0^*} - 1)]^{1/(2\alpha+1)} b_n \rfloor, \tag{18}$$

$$\check{h}_3(x) := \check{h}(x, J_3, \alpha, \{\hat{\theta}_j\}) \quad \text{where} \quad J_3 := \lfloor [QG_0(a)/(\mathrm{e}^{\int_a^{a+1} h_0(v)\mathrm{d}v} - 1)]^{1/(2\alpha+1)} b_n \rfloor, \tag{19}$$

$$\check{h}_4(x) := \sum_{j=0}^{M_n - 1} \hat{\theta}_j \varphi_j(x) + \sum_{j=M_n}^{J_3} (1 - (j/J_3)^{\alpha})\hat{\theta}_j \varphi_j(x). \tag{20}$$

Note that, despite the fact that sharp constants are the same for the two local function classes $\mathcal{S}_3$ and $\mathcal{S}_4$, the dealer-estimator (20), proposed for the weakly restricted local class $\mathcal{S}_4$, is different from the Pinsker's estimator (19) proposed for the Golubev's local class, and furthermore it is not even a Pinsker's estimator. On the other hand, it has several features of the dealer-estimator $\check{h}_3$, namely the same cutoff J_3 and the same shrinkage coefficients on high frequencies. As we shall see below, a modification of Pinsker's estimator for class $\mathcal{S}_4$ is necessary.

Theorem 5 (Sharpness of Lower Minimax Bounds of Sect. 2) *Consider settings of Theorems i, $i = 1, 2, 3, 4$ and denote considered in these theorems function classes and presented lower bounds for minimax MISEs as $\mathcal{S}_i$ and R_i, respectively. Then,* *the dealer-estimators $\check{h}_i$, defined in* (17)–(20)*, are sharp-minimax for corresponding function classes S_i, that is,*

$$\sup_{h \in \mathcal{S}_i} E_h \left\{ \int_a^{a+1} (\check{h}_i(x) - h(x))^2 \mathrm{d}x \right\} = R_i(1 + o_n(1)), \quad i \in \{1, 2, 3, 4\}. \tag{21}$$

Furthermore, a Pinsker's dealer-estimator (13) *is not rate-minimax (its MISE converges slower than $n^{-2\alpha/(2\alpha+1)}$) for a weakly restricted local Sobolev class $\mathcal{S}_4 = \mathcal{S}_4(\alpha, Q, h_0, \rho_n, M_n)$ whenever the pivot $h_0(x)$ is not Sobolev of order α on the interval $[a, a+1]$. That is, if*

$$\sum_{j=1}^{\infty} j^{2\alpha} \left[\int_a^{a+1} h_0(x) \varphi_j(x) \mathrm{d}x \right]^2 = \infty, \tag{22}$$

then

$$n^{2\alpha/(2\alpha+1)} \min \sup_{h \in \mathcal{S}_4(\alpha, Q, h_0, \rho_n, M_n)} E_h \left\{ \int_a^{a+1} (\check{h}(x, J(n), \alpha, \{\hat{\theta}_j\}) - h(x))^2 \mathrm{d}x \right\}$$
$$\to \infty \quad \text{as } n \to \infty, \tag{23}$$

where the minimum is taken over $J(n) \in \{0, 1, \ldots, n\}$.

We conclude that the four lower bounds of Sect. 2 are sharp and attainable by dealer-estimates, and the weakly restricted local Sobolev class requires a dealer-estimator which is different from a classical Pinsker's one.

Remark 4 A series estimate may take on negative values. Then, an L_2-projection on a class of nonnegative functions makes the estimate bona fide and reduces the MISE. The interested reader can find the projection algorithm in Efromovich (1999, p. 63).

Now let us present an oracle-estimator motivated by the blockwise-shrinkage methodology developed in Efromovich (1985) for the case of probability density estimation. Let $\{B_k, k = 1, 2, \ldots\}$ be a partition of nonnegative integers (frequencies of the cosine basis $\{\varphi_j(x), j = 0, 1, \ldots\}$ into non-overlapping blocks of cardinality (length) L_k such that $\max(j : j \in B_k) < \min(j : j \in B_{k+1})$. Only to be specific, set $L_k = 1$ for $k = 1, 2, \ldots, \lfloor \ln(n) \rfloor$ and $L_k = \lfloor (1 + \ln^{-1}(n)/\ln(\ln(n)))^k \rfloor + 1$ for $k > \ln(n)$, and also introduce a sequence of integers K_n such that K_n is the smallest positive integer satisfying $\sum_{k=1}^{K_n} L_k > n^{1/3} \ln(\ln(n))$. Note that the largest length L_{K_n} is of order $n^{1/3}/\ln(n)$, K_n is of order $\ln^2(n)\ln(\ln(n))$, and the total number of estimated Fourier coefficients is of order $n^{1/3}\ln(\ln(n))$. The latter is due to the assumed restriction $\alpha \geq 1$ which implies that the effect of not estimated Fourier coefficients on the MISE is of order $[n^{1/3}\ln(\ln(n))]^{-2\alpha} = o_n(1)n^{-2\alpha/(2\alpha+1)}$.

Now we can define the oracle-estimator [remember that the statistic $\hat{\theta}_j$ is defined in (14)]

$$\tilde{h}^*(x, h) := \sum_{k=1}^{K_n} \frac{\Theta_k}{\Theta_k + d^* n^{-1}} \sum_{j \in B_k} \hat{\theta}_j \varphi_j(x), \tag{24}$$

where

$$\Theta_k := \Theta_k(h) := L_k^{-1} \sum_{j \in B_k} \left[\int_a^{a+1} h(v)\varphi_j(v)\mathrm{d}v \right]^2 \tag{25}$$

and

$$d^* := d^*(h) := \int_a^{a+1} h(v)G^{-1}(v)\mathrm{d}v. \tag{26}$$

Note that Θ_k, $k = 1, 2, \ldots$ and d^* may be referred to as Sobolev's functionals and a coefficient of difficulty.

Theorem 6 *The oracle-estimator $\tilde{h}^*(x, h)$, defined in (24), is sharp-minimax for the four functional classes S_i considered in Sect. 2, that is, using notation of Theorem 5,*

$$\sup_{h \in \mathcal{S}_i} E_h \left\{ \int_a^{a+1} (\tilde{h}^*(x, h) - h(x))^2 \mathrm{d}x \right\} \leq R_i(1 + o_n(1)), \quad i \in \{1, 2, 3, 4\}. \tag{27}$$

It is an attractive idea to mimic the oracle-estimator by a data-driven estimator which uses estimates of unknown functionals Θ_k and d^*. Set

$$\hat{h}(x) := \sum_{k=1}^{K_n} \left[1 - \frac{\hat{d}n^{-1}}{L_k^{-1} \sum_{j \in B_k} \hat{\theta}_j^2} \right] I \left(L_k^{-1} \sum_{j \in B_k} \hat{\theta}_j^2 \right.$$
$$\left. > \ (\hat{d} + 1/\ln(n))n^{-1} \right) \sum_{j \in B_k} \hat{\theta}_j \varphi_j(x), \tag{28}$$

where $\hat{\theta}_j$ is defined in (14) and (remember that the antirank η_l is defined in (15) and $\eta_l \geq 1$)

$$\hat{d} := n \sum_{l=1}^{n} \eta_l^{-2} I(X_l \in [a, a+1]). \tag{29}$$

Theorem 7 *Data-driven estimator $\hat{h}(x)$, defined in (28), is sharp-minimax for the four function classes considered in Sect. 2, that is, using notation of Theorem 5,*

$$\sup_{h \in \mathcal{S}_i} E_h \left\{ \int_a^{a+1} (\hat{h}(x) - h(x))^2 \mathrm{d}x \right\} = R_i(1 + o_n(1)), \quad i \in \{1, 2, 3, 4\}. \tag{30}$$

In Sect. 6, the interested reader can find an oracle inequality (150) for the proposed estimator.

4 Numerical study

Plug-in estimation is the classical method used in nonparametric statistics for estimation of the hazard rate. It is based on definition (1) and the fact that estimations of the probability density and survivor function are two classical and oldest problems in nonparametric statistics. The plug-in estimation procedure goes back to Watson and Leadbetter (1964) when appropriate estimates $\tilde{p}$ for the density and $\tilde{G}$ for the survivor function are plugged in (1) to get the estimate

$$\tilde{h}(x) = \frac{\tilde{p}(x)}{\tilde{G}(x)}. \tag{31}$$

Let us note that some authors refer to this estimate as external, see Nielsen and Linton (1995). Because the survivor function is estimated with the parametric accuracy, the asymptotic MISE of plug-in estimator (31) is defined by the term $E\{\int_a^{a+1}(\tilde{p}(x) - p(x))^2 G^{-2}(x)\mathrm{d}x\}$. As a result, given $G(a+1)$ is separated from zero, the plug-in (external) estimator inherits minimax rates of the density estimator.

The aim of the numerical study is to compare a good plug-in estimator with the proposed one. Following Silverman (1986, s.6.5.1), we are using $\tilde{p}$ which is a kernel estimator with Gaussian kernel $K(x)$. The optimal variable bandwidth $w(x)$ for the density kernel estimator is chosen according to the golden rule

$$w(x) = w(x, p, K) = \frac{[p(x)\int_{-\infty}^{\infty} K^2(t)dt]^{1/5}}{[p^{(2)}(x)\int_{-\infty}^{\infty} t^2 K(t)dt]^{2/5}} n^{-1/5}. \tag{32}$$

In our numerical study, an underlying density is known, and it will be used by the kernel estimator to calculate (32). As a result, we can refer to the plug-in estimator as an oracle-estimator. Furthermore, to avoid boundary issues that are critical for a kernel estimator, we will study estimation over an interval $[a, b]$ with $a > 0$. Finally, the empirical survivor function is used in the denominator of (31).

The used underlying hazard rate functions are from Weibull distribution,

$$h(x) = \alpha\lambda^{-1}x^{\alpha-1}I(x > 0). \tag{33}$$

Here α is the shape and λ is the scale of the Weibull distribution. Note that if $\alpha = 1$ then this distribution becomes exponential (memoryless).

We are conducting 48 experiments with different shapes and scales of Weibull distribution, different intervals $[a, b]$ on which the MISE is evaluated, and different sample sizes. For each experiment, 500 simulations are generated, and for each sample the empirical integrated squared errors of the plug-in oracle (ISEO) and the proposed estimate (ISEE) are calculated. Then, the median ratio (over 500 simulations) of ISEO/ISEE is shown in Table 1. Also, for each experiment, the table contains the expected number of observations fallen within the studied interval $[a, b]$.

As we see, the proposed estimator performs well with respect to the oracle-estimator. Furthermore, please look at the average number of observations fallen within a studied interval $[a, b]$. These numbers are relatively small with respect to n, and they shed light on complexity of the studied problem.

Table 1 Results of Monte Carlo simulations for Weibull hazard rate with parameters α (shape) and λ (scale)

α	λ	a	b	n					
				100	200	300	400	500	1000
1	2	1	3	1.18/38	1.11/77	0.99/115	1.07/153	1.12/191	1.35/383
1	2	1	4	1.30/47	1.46/94	1.49/141	1.51/188	1.81/236	1.84/471
1	4	1	4	0.98/41	1.12/82	1.13/123	1.21/164	1.22/205	1.24/411
1	4	2	7	1.23/43	1.32/87	1.40/130	1.42/173	1.49/216	1.50/433
0.7	2	0.8	6	1.16/48	1.08/95	1.01/142	1.07/190	1.15/237	1.18/475
0.7	4	0.8	6	0.91/46	0.92/92	0.97/137	0.99/183	1.08/229	1.23/458
1.2	2	1	4	1.24/54	1.13/109	1.07/164	1.08/218	1.07/273	1.05/546
1.5	0.5	4	3	1.12/82	1.14/165	1.18/247	1.17/329	1.18/412	1.15/824

For each experiment, which is defined by the two parameters of the distribution, sample size n, and the interval of estimation $[a, b]$, 500 samples are generated and then for each sample the plug-in-oracle estimate and the proposed estimate are calculated and then the corresponding integrated squared errors over $[a, b]$ are calculated and denoted as ISEO and ISEE, respectively. Each entry in the table is written as A/B where A is the median of 500 ratios ISEO/ISEE and B is the average number of observations fallen within the considered interval $[a, b]$

5 Discussion

Hazard rate estimation The problem of nonparametric estimation of the hazard rate function, which by definition is the probability density divided by the survivor function, has been always in the shadow of nonparametric density estimation because the survivor function is estimated with the parametric rate. As a result, a good density estimate, divided by the empirical survivor function, should produce a correspondingly good hazard rate estimate. This natural approach works perfectly well if the rate of a risk convergence is of interest. Does it work out if the sharp constant is of interest? The paper shows that in this case a special estimator is needed to attain the sharp constant. The proposed data-driven estimator is simultaneously sharp minimax for a number of underlying functional classes and it adapts to unknown smoothness of an underlying hazard rate. Furthermore, the study of local functional classes has allowed us to understand how an underlying hazard rate affects the accuracy of estimation. The numerical study confirms theoretical conclusions and the proposed estimator can be recommended for small sample sizes.

Minimax theory A nonparametric estimation problem is traditionally studied via considering a single function class. In the case of a sharp minimax estimation under the MISE criteria, that class is typically either the Pinsker (1980)'s global Sobolev class [here it is the class (4)] or the Golubev (1991)'s local Sobolev class [here it is the class (9)]. The difference between the classes is that the local one studies functions that are close in L_∞-norm to a pivot. Golubev (1991) proved that for the probability density, regression and spectral density models the local and global approaches imply the same sharp constant and then the same Pinsker's estimate is sharp minimax for global and local function classes. The outcome changes for the hazard rate problem where global and local approaches yield different sharp constants. This outcome can be explained by the fact that the hazard rate is the ratio of the density and the survivor function, and in that ratio both the numerator and the denominator depend on the pivot. To shed a new light on this result, two additional Sobolev classes (one global and one local) with the same smoothness are considered. The two global classes imply different sharp constants, and then different shrinkage coefficients should be used by corresponding Pinsker's estimators. On the contrary, two local classes imply the same sharp constant and then different dealer-estimators are required to attain the constant. Furthermore, for one of the local functional class the Pinsker's estimator is no longer even rate minimax. This is the first known case in the minimax theory when the Pinsker's estimator is not minimax. The practical conclusion of the theory is that a data-driven hazard rate estimator should adapt to smoothness of the hazard rate and an underlying functional class.

Minimax game To reflect upon and take into account the fact that an array of functional classes is considered, it is useful to modify traditional explanation of the minimax approach as a game between nature and the statistician where nature chooses a least favorable function, generates a corresponding sample and then the statistician uses the sample to estimate that function. If several function classes are considered, then someone else should choose an underlying class and deal it to nature. Remember that in some card games the dealer deals cards and, therefore, it is natural, in addition to nature and the statistician, to introduce the dealer who deals a functional class to nature. There is an extra benefit from considering the trio of players. Namely, when we say that a minimax lower bound indicates a specific rate, then whose rate is it? It is definitely not the rate for nature who knows the estimand, and it is also not the rate for the statistician because any lower bound is deduced for a pseudo-estimator that knows parameters of an underlying class. The correct answer is that all known lower bounds are lower bounds for dealer-estimators.

Adaptive data-driven estimation Typically adaptive estimation is considered as a separate problem from establishing sharp-minimax lower bounds. Here, due to the new requirement of adaptation to an underlying functional class, it becomes a part of the minimax theory. It is shown that a single data-driven estimator, based on the blockwise-shrinkage methodology of Efromovich (1985), is sharp minimax for the four functional classes and it can be used in place of specific dealer-estimators for each function class.

Least favorable hazard rate It is shown in Sect. 6 that in the minimax game nature always chooses a hazard rate corresponding to a random variable X which is the minimum of two independent random variables: one is memoryless and another has a vanishing hazard rate. Distributions of the two random variables are defined by an underlying functional class in the minimax game. This outcome allows us to conclude that, even for the case of a global functional class, nature uses a local approach. Proofs of lower bounds indicate the following difference between the least favorable distributions for the density and hazard rate estimation. For the density setting, nature randomly chooses a density by assigning a Gaussian distribution to Fourier coefficients with respect to a basis on the interval of interest $[a, a + 1]$. For the hazard rate, nature performs a more sophisticated play. Nature divides $[a, a + 1]$ into a number of subintervals (this number increases as the sample size increases), then nature spreads the Sobolev's power among those subintervals according to a special algorithm, and only then for each subinterval nature uses an approach similar to the density case. This remark sheds light on the dramatic difference between nature's plays for the hazard rate and density settings.

Nuisance function One of the interesting topics in the nonparametric statistics is the effect of nuisance functions on estimation. Traditional nuisance functions, like the scale function in nonparametric regression, do not depend on the estimand and the main issue is the possibility to estimate a nuisance function with sufficient accuracy. In the dual settings of density and hazard rate estimation, the survivor function may be looked at as the nuisance function. There is clearly no issue with smoothness of the survivor function, but the paper shows that it makes a significant difference in the theory of sharp minimax estimation due to its affect on sharp constants and sharp dealer-estimators for different functional classes. The observed phenomenon can be explained by the fact that each of the three functions—probability density, hazard rate and survivor function—uniquely defines an underlying distribution of the data.

Practical implications There is an interesting relationship between the asymptotic theory and its practical feasibility for small samples. The presented numerical study, as well as known results of Efromovich (1999) for other models, indicates that for sample sizes up to several hundreds only a low-frequency part of the underlying hazard rate should be estimated. As a result, approaches of the global minimax and the local minimax are appropriate for these samples. For moderate and large sample sizes, a weakly restricted local minimax is more appropriate because these sample sizes allow us to look at higher frequencies. Furthermore, for large sample sizes knowing or not knowing a low-frequency part of the hazard rate has no significant effect on the estimation. We may conclude that each minimax setting has its own applied merits. Another important remark is that, as the coefficient of difficulty indicates, the hazard rate estimation may be dramatically more complicated problem than the density estimation. A practitioner should always pay attention to this fact.

Future topics Let us mention five related topics to be explored in the future. The first one is hazard rate estimation under a shape restriction. There is a vast literature devoted to this topic (see a review in Jankowski and Wellner 2009) where either consistency or rates of a risk convergence are studied. Here local minimax approaches look reasonable and will allow us to establish sharp minimax lower bounds for bona fide estimators. The second topic is a classical one and it is hazard rate estimation with censored data, see a discussion in Antoniadis et al. (1999) and Brunel and Comte (2005). In the first paper, wavelet methodology is explored and in the second a finite sample adaptation of estimators without a priori assumption on the regularity of an underlying hazard rate is considered. This is the setting where a sharp minimax approach should shine because the theory will allow us to understand how an underlying hazard rate, together with censoring mechanism, affects the sharp constant and coefficient of difficulty. The third topic, which is a new one, is hazard rate estimation with missing data. The fourth topic is estimation of the conditional hazard rate, and an interesting discussion and examples can be found in Spierdijk (2008), where a kernel estimator with plug-in bandwidth based on a reference distribution is proposed, and a data-drive adaptive procedure is given in Comte et al. (2011). Let us stress that so far no result about efficient and adaptive estimation for these four topics is known. Finally, in some applications the cumulative hazard $\Lambda(x) := \int_0^x h(v)\mathrm{d}v$ is of interest, see a review in Spierdijk

(2008). The author conjectures that the proposed minimax methodology may be used to develop the theory of second-order efficient estimation of the cumulative hazard.

6 Proofs

Remember that $\varphi_0(x) = 1$, $\varphi_j(x) = \sqrt{2}\cos(\pi j(x-a))$, $j \geq 1$ is the classical cosine basis on the unit interval $[a, a+1]$, $o_n(1)$ and Cs denote generic vanishing sequences and positive constants. We also use notation $o_n^*(1)$ and C^* to stress the fact that these generic vanishing sequences and positive constants do not depend on all other parameters considered in a proof. $\lfloor x \rfloor$ denotes the smallest positive integer larger than x. In what follows it is assumed that $n \geq 30$ so all sequences in n are well defined.

We begin with the proof of Theorem 1, which is also used to verify other lower bounds. This explains its rather general approach.

Proof of Theorem 1 Remember that we are establishing a lower bound for a dealer-estimate that knows all parameters of an underlying function class. In particular, the dealer knows a pivotal hazard rate. For now, let $h_0(x)$ be a known pivotal hazard rate function such that $\inf_{x\in[a,a+1]} h_0(x) > 0$, $\int_a^{a+1} h_0^2(x)\mathrm{d}x < \infty$ and $\mathrm{e}^{-\int_0^a h_0(u)\mathrm{d}u} = G(a)$. This function will be specified later for each particular underlying function class.

On first glance, it looks reasonable to verify the lower bound for the hazard rate following the known proof for the density in Efromovich and Pinsker (1982). To follow that proof, we should verify that in a minimax game nature chooses a least favorable hazard rate among functions $h(x) = h_0(x) + [\sum_{j=1}^{J_n} \theta_j \varphi_j(x)] I(x \in [a, a+1])$ where $\{\theta_j, j = 1, 2, \ldots\}$ satisfy some specific restrictions. A direct calculation shows that this approach produces a lower bound which is too small. In other words, presence of the "nuisance" survival function $G(x) = G(x, h)$ allows nature to choose a smarter way to choose a least favorable hazard rate. Namely, as we shall see shortly, nature divides $[a, a+1]$ into $s = s_n \to \infty$, $n \to \infty$ subintervals and then defines its own Sobolev's class on each subinterval. The main challenge of the proof is to understand how nature spreads the power of an underlying Sobolev class over the subintervals, and this nature's strategy makes the proof more complicated and challenging than in the density case considered in Efromovich and Pinsker (1982).

The proof requires intensive notation presented and commented on in the next paragraph.

Notation Set $s := s_n := 1 + \lfloor \ln(\ln(n)) \rfloor$, and this will be the total number of subintervals used by nature. For $k = 0, 1, \ldots, s$ set $\mathcal{I}_{sk} := [h_0(a + k/s)\mathrm{e}^{\int_0^{a+k/s} h_0(v)\mathrm{d}v}]^{-1}$, $\mathcal{I}_s := [\sum_{k=0}^{s-1} \mathcal{I}_{sk}^{-1}]^{-1}$, $Q_{sk} := (1 - 1/s)\mathcal{I}_s \mathcal{I}_{sk}^{-1} Q$, $J := \lfloor [n(2\alpha+1)(\alpha+1)s^{-2\alpha}(\alpha\pi^{2\alpha})^{-1}(1-s^{-1})Q\mathcal{I}_s]^{1/(2\alpha+1)} \rfloor$, $J_* = \lfloor J/\ln(n) \rfloor$, $\varphi_{skj}(x) = (2s)^{1/2}\cos(\pi j[s(x-a)-k])$. Let $\phi(n, v)$, $v \in (-\infty, \infty)$ be a sequence of flattop non-negative kernels defined on a real line such that for a given n the kernel $\phi(n, v)$ is zero beyond $(0, 1)$, it is α-fold continuously differentiable on $(-\infty, \infty)$, $0 \leq \phi(n, v) \leq 1$, $\phi(n, v) = 1$ for $2(\ln(n))^{-2} \leq v \leq 1 - 2(\ln(n))^{-2}$, and $|\phi^{(\alpha)}(n, v)| \leq C(\ln(n))^{2\alpha}$. For instance, such a kernel may be constructed using the so-called mollifiers discussed in Efromovich (1999, ch.7). Then set $\phi_{sk}(x) := \phi(n, s(x-a)-k)$, and note this kernel is zero beyond $(a+k/s, a+(k+1)/s)$ and it is equal to 1 on $[a+k/s+2[s\ln^2(n)]^{-1}, a+(k+1)/s-2[s\ln^2(n)]^{-1}]$ except of two boundary intervals of the total length $4[s\ln^2(n)]^{-1}$. This special kernel is used to "sew" functions defined on adjoint subintervals $[a+k/s, a+(k+1)/s]$, so over interval $[a, a+1]$ considered functions are sufficiently smooth. Now let us introduce several parametric function classes. For a $k \in \{0, 1, \dots, s\}$ set $\vec{\nu}_{sk} := \{\nu_{skJ_*}, \dots, \nu_{skJ}\}$ and define a parametric function class

$$\mathcal{H}_{sk} := \Bigg\{ f_{sk} : \ f_{sk}(x|\vec{\nu}_{sk}) := \sum_{j=J_*}^{J} \nu_{skj}\varphi_{skj}(x) \ , \\ \times \sum_{j=J_*}^{J} (\pi j)^{2\alpha}\nu_{skj}^2 \le s^{-2\alpha}Q_{sk}, \ |f_{sk}(x|\vec{\nu}_{sk})| \\ \le [s^4\ln(n)]^{1/2}n^{-\alpha/(2\alpha+1)}, \ x \ge 0 \Bigg\}. \tag{34}$$

"Sewing" these classes together with the help of flattop kernels $\phi_{sk}(x)$, we may define a parametric class

$$\mathcal{H}_s := \{h : \ h(x|\vec{\nu}_s) = h_0(x) + \sum_{k=0}^{s-1} f_{sk}(x|\vec{\nu}_{sk})\phi_{sk}(x), \ f_{sk} \in \mathcal{H}_{sk}, \ h(x) \ge 0, x \ge 0\}, \tag{35}$$

where $\vec{\nu}_s = (\vec{\nu}_{s0}, \dots, \vec{\nu}_{s(s-1)})$. Note that a function from this class is defined by a vector-parameter $\vec{\nu}_s$, and it is convenient to introduce sets of parameters corresponding to the above-defined function classes. For each functional class $\mathcal{H}_{sk}$, the corresponding set of vector-parameters $\vec{\nu}_{sk}$ is $V_{sk} := \dot{V}_{sk} \cap \ddot{V}_{sk}$ where $\dot{V}_{sk} := \{\vec{\nu}_{sk} : \sum_{j=J_*}^{J}(\pi j)^{2\alpha}\nu_{skj}^2 \le s^{-2\alpha}Q_{sk}\}$ and $\ddot{V}_{sk} := \{\vec{\nu}_{sk} : \max_{x\in[a,a+1]}|\sum_{j=J_*}^{J}\nu_{skj}\varphi_{skj}(x)| \le [s^4\ln(n)]^{1/2}n^{-\alpha/(2\alpha+1)}\}$. For $\mathcal{H}_s$, the corresponding set of parameters is $V_s := \prod_{k=0}^{s-1} V_{sk} := V_{s0} \times \dots \times V_{s(s-1)}$. In what follows we may also use negative subscripts to indicate that a specific part of a vector or a set is skipped, like in $V_{-sr} := \prod_{k\in\{0,\dots,s-1\}\setminus r} V_{sk}$, or $\vec{\nu}_{-s1} := (\vec{\nu}_{s0}, \vec{\nu}_{s2}, \dots, \vec{\nu}_{s(s-1)})$. With this understanding of negative subscripts in mind, set $f_{-skj}(x|\vec{\nu}_{-skj}) := \sum_{r\in\{J_*,\dots,J\}\setminus\{j\}} \nu_{skj}\varphi_{skj}(x)$ where $\vec{\nu}_{-skj}$ denotes vector $\vec{\nu}_{sk}$ with its jth component being skipped. Further, set

$$V_{skj} := \left\{ \vec{\nu}_{-skj} : \ \max_{x\in[a,a+1]} |f_{-skj}(x|\vec{\nu}_{-skj})| \le (1/2)[s^4\ln(n)]^{1/2}n^{-\alpha/(2\alpha+1)} \right\} \\ \times \{\nu_{skj} : |\nu_{skj}| < s^2 n^{-1/2}\} \\ =: V_{-skj} \times \{\nu_{skj} : |\nu_{skj}| < s^2 n^{-1/2}\}.$$

Denote by $\vec{v}_{-skj,0}$ a vector $\vec{v}_s$ with its element v_{skj} being replaced by zero. A dealer-estimate is denoted as $\hat{h}$, and note that it may depend on the pivot h_0 which is known to the dealer. As a result, the dealer also knows function $\hat{f}(x) := \hat{h}(x) - h_0(x)$. Furthermore, in a similar way we can write for $h \in \mathcal{H}_s$,

$$\begin{aligned} h(x) &=: h(x|\vec{v}_s) = h_0(x) + \sum_{k=0}^{s-1} \sum_{j=J_*}^{J} v_{skj}\varphi_{skj}(x)\phi_{sk}(x) = h_0(x) \\ &\quad + \Bigg[\Bigg[\sum_{k\in\{0,\ldots,s-1\}\setminus r} \sum_{j=J_*}^{J} v_{skj}\varphi_{skj}(x)\phi_{sk}(x)\Bigg] \\ &\quad + \Bigg[\sum_{j\in\{J_*(k),\ldots,J(k)\}\setminus i} v_{srj}\varphi_{srj}(x)\phi_{sr}(x)\Bigg] + v_{sri}\varphi_{sri}(x)\phi_{sr}(x)\Bigg] \\ &=: h_0(x) + \dot{f}_s(x|\vec{v}_s) =: h_0(x) + [\dot{f}_{-sr}(x|\vec{v}_{-sr}) + \dot{f}_{-sri}(x|\vec{v}_{-sri})\phi_{sr}(x) \\ &\quad + v_{sri}\varphi_{sri}(x)\phi_{sr}(x)], \quad \vec{v}_s \in V_s. \end{aligned} \tag{36}$$

Note that in $\dot{f}$ the dot above the function indicates that the function is "sewed" (smoothed) by the flattop kernel. Finally, $q > 3$ and $b > 3$ are constants, $o_t^*(1)$ denote generic sequences in t which vanish as $t \to \infty$ uniformly over all other parameters considered in the proof, $\dot{f}^{(l)}(x|\vec{v}) := \partial^\alpha \dot{f}(x|\vec{v})/\partial x^\alpha$, $g(v|\sigma) := (2\pi\sigma^2)^{-1/2}e^{-v^2/(2\sigma^2)}$ denotes the normal density with zero mean and standard deviation σ, and $g(\vec{v}_{sk}|\vec{\tau}_{sk}) := \prod_{j=J_*}^{J} g(v_{skj}|\tau_{skj})$, where $\tau_{skj} := [n^{-1}(1-3q^{-1})\mathcal{I}_{sk}^{-1}\max(q^{-1}, \min(q, (J/j)^\alpha - 1))]^{1/2}$.

The proof of the lower bound consists of several steps. Step 1 is to check that for all sufficiently large n (note that all assertions are verified for sufficiently large n) $\mathcal{H}_s$ is a subset of an underlying Sobolev class. This is the place where we are considering a specific Sobolev space and define a corresponding pivot h_0 on the interval of interest $[a, a+1]$, and beyond this interval $h_0(x)$ is any hazard rate satisfying restrictions of an underlying function class. In Theorem 1, we are using $h_0(x) := Q_0(1 - 1/s^2)$ for $x \in [a, a+1]$, and the only other restriction on the pivot is that $e^{-\int_0^a h_0(v)dv} = G(a)$. For other Sobolev classes, the pivot will be defined in their own proofs. We need to check that $\mathcal{H}_s \subset \mathcal{S}_1(\alpha, Q_0, Q, G(a))$ which will follow from (remember notation $\dot{f}_s(x|\vec{v}_s) := h(x|\vec{v}_s) - h_0(x)$ introduced in (36))

$$\begin{aligned} &\sup_{\vec{v}_s\in V_s} \sum_{j=1}^{\infty} (\pi j)^{2\alpha} \left[\int_a^{a+1} (h(x) - h_0(x))\varphi_j(x)dx\right]^2 \\ &= \sup_{\vec{v}_s\in V_s} \sum_{j=1}^{\infty} (\pi j)^{2\alpha} \left[\int_a^{a+1} \dot{f}_s(x|\vec{v}_s)\varphi_j(x)dx\right]^2 \le Q. \end{aligned} \tag{37}$$

To verify (37), we note that $\dot{f}$ and its α derivatives are periodic on $[a, a+1]$ due to the used flattop kernels. Then, it is verified via integration by parts (or see Efromovich

1999) that the following Parseval identity for $\dot{f}_s^{(\alpha)}$ holds:

$$\sum_{j=1}^{\infty}(\pi j)^{2\alpha}\left[\int_a^{a+1}\dot{f}_s(x|\vec{v}_s)\varphi_j(x)\mathrm{d}x\right]^2=\int_a^{a+1}\left[\dot{f}_s^{(\alpha)}(x|\vec{v}_s)\right]^2\mathrm{d}x. \tag{38}$$

As a result, to check (37) it suffices to evaluate the integral in the right side of (38). Write

$$\int_a^{a+1}[\dot{f}_s^{(\alpha)}(x|\vec{v}_s)]^2\mathrm{d}x=\sum_{k=0}^{s-1}\int_{a+k/s}^{a+(k+1)/s}[(f_{sk}(x|\vec{v}_{sk})\phi_{sk}(x))^{(\alpha)}]^2\mathrm{d}x,\quad \vec{v}_s\in V_s. \tag{39}$$

Using Leibnitz's rule, together with the Cauchy inequality $[a_1+a_2]^2\le a_1^2(1+\gamma)+a_2^2(1+\gamma^{-1})$, $\gamma>0$ and notation $\mathbf{C}_l^{\alpha}:=\alpha!/((\alpha-l)!l!)$, we can write for any positive γ,

$$\begin{aligned}
&\int_{a+k/s}^{a+(k+1)/s}[(f_{sk}(x|\vec{v}_{sk})\phi_{sk}(x))^{(\alpha)}]^2\mathrm{d}x\\
&=\int_{a+k/s}^{a+(k+1)/s}\left[\sum_{l=0}^{\alpha}\mathbf{C}_l^{\alpha}f_{sk}^{(\alpha-l)}(x|\vec{v}_{sk})\phi_{sk}^{(l)}(x)\right]^2\mathrm{d}x\\
&=\int_{a+k/s}^{a+(k+1)/s}[f_{sk}^{(\alpha)}(x|\vec{v}_{sk})\phi_{sk}(x)]^2\mathrm{d}x(1+\gamma)\\
&\quad+\int_{a+k/s}^{a+(k+1)/s}\left[\sum_{l=1}^{\alpha}\mathbf{C}_l^{\alpha}f_{sk}^{(\alpha-l)}(x|\vec{v}_{sk})\phi_{sk}^{(l)}(x)\right]^2\mathrm{d}x(1+\gamma^{-1})\\
&=:A_{sk}+B_{sk}.
\end{aligned} \tag{40}$$

Using definition of kernels $\phi_{sk}(x)$ we get

$$\max_{0\le l\le\alpha}\int_{a+k/s}^{a+(k+1)/s}(\phi_{sk}^{(l)}(x))^2\mathrm{d}x<C^*(s(\ln(n))^2)^{2\alpha}, \tag{41}$$

and for $0<l\le\alpha$, $\vec{v}_{sk}\in V_{sk}$ and $x\in[a+k/s,a+(k+1)/s]$ we can write using Cauchy–Schwarz inequality,

$$\begin{aligned}
|f_{sk}^{(\alpha-l)}(x|\vec{v}_{sk})|^2&=\left|\sum_{j=J_*}^{J}v_{skj}\varphi_{skj}^{(\alpha-l)}(x)\right|^2\\
&\le Cs^{2(\alpha-l)+1}\left(\sum_{j=J_*}^{J}j^{2\alpha}v_{skj}^2\right)\left(\sum_{j=J_*(k)}^{J(k)}j^{-2l}\right)\\
&\le C^*J^{-2l+1}[\ln(n)]^{2l}.
\end{aligned}$$

The obtained inequality, together with definition of J, implies that the term B_{sk}, defined in (40), can be evaluated as

$$B_{sk} = (1+\gamma^{-1})o_n^*(1)n^{-1/[2(2\alpha+1)]}. \tag{42}$$

Further, using $|\phi_{sk}(x)| \le 1$ we get

$$\int_{a+k/s}^{a+(k+1)/s} [f_{sk}^{(\alpha)}(x|\vec{v}_{sk})\phi_{sk}(x)]^2 \mathrm{d}x \le \int_{a+k/s}^{a+(k+1)/s} (f_{sk}^{(\alpha)}(x|\vec{v}_{sk}))^2 \mathrm{d}x \le Q_{sk},\ \vec{v}_{sk} \in V_{sk}. \tag{43}$$

Now set $\gamma = 1/\ln(n)$, and using (40), (42), (43) and $\sum_{k=0}^{s-1} Q_{sk} = Q(1-s^{-1})$ we get

$$\sup_{\vec{v}_s \in V_s} \int_a^{a+1} [\dot{f}_s^{(\alpha)}(x|\vec{v}_s)]^2 \mathrm{d}x \le \sum_{k=0}^{s-1} Q_{sk} + o_n^*(1)s^{-1} = Q - s^{-1}[Q + o_n^*(1)].$$

This result, together with (38) and (39), proves (37) for all sufficiently large n.

Second step is to convert the studied MISE of dealer-estimate $\hat{f}(x) = \hat{h}(x) - h_0(x)$ into the sum of MSEs of corresponding Fourier coefficients $\hat{v}_{skj} = \int_{a+k/s}^{a+(k+1)/s} \hat{f}(x)\varphi_{skj}(x)\mathrm{d}x$. Using a Cauchy-type inequality $(c+d)^2 \ge (1-s^{-1})c^2 - sd^2$, as well as definition of the kernel $\phi_{sk}(x)$, definition of the set $\ddot{V}_s \subset V_{sk}$, and the Bessel inequality, we can write for $\vec{v}_s \in V_s$,

$$\begin{aligned}
&\int_{a+k/s}^{a+(k+1)/s} (\hat{f}(x|\vec{v}_{sk}) - f_{sk}(x|\vec{v}_{sk})\phi_{sk}(x))^2 \mathrm{d}x \\
&\ge \int_{a+k/s}^{a+(k+1)/s} \Big[(1-s^{-1})(\hat{f}(x) - f_{sk}(x|\vec{v}_{sk}))^2 \\
&\quad - s f_s^2(x|\vec{v}_{sk})(1-\phi_{sk}(x))^2\Big] \mathrm{d}x \ge (1-s^{-1}) \sum_{j=J_*}^{J} (\hat{v}_{skj} - v_{skj})^2 \\
&\quad - s^4 \ln(n) n^{-2\alpha/(2\alpha+1)} [4(\ln^2(n)s)^{-1}].
\end{aligned} \tag{44}$$

In the last inequality, we used the definition of $\ddot{V}_{sk}$ and that on $[a+k/s, a+(k+1)/s]$ function $1-\phi_{sk}(x)$ is zero apart of two subintervals of the total length $4(\ln^2(n)s)^{-1}$ where it is nonnegative and less than 1.

In step 3, we combine results of the first two steps and get

$$\begin{aligned}
&\inf_{\hat{h}} \sup_{h \in S_1(\alpha, Q_0, Q, G(a))} E_h \left\{ \int_a^{a+1} (\hat{h}(x) - h(x|\vec{v}_s))^2 \mathrm{d}x \right\} \\
&\ge (1-s^{-1}) \inf_{\hat{\vec{v}}_s} \sup_{\vec{v}_s \in V_s} \sum_{k=0}^{s-1} \sum_{j=J_*}^{J} E_{\vec{v}_s} \left\{ (\hat{v}_{skj} - v_{skj})^2 \right\} \\
&\quad + o_n(1) n^{-2\alpha/(2\alpha+1)} =: (1-s^{-1}) R_s + o_n(1) n^{-2\alpha/(2\alpha+1)}.
\end{aligned} \tag{45}$$

Note that in the second expectation, we used subscript $\vec{v}_s$ in place of h because for the dealer (remember that we are establishing a lower bound for the dealer's MISE) knowing $\vec{v}_s$ is equivalent to knowing $h(x|\vec{v}_s)$.

Step 4 is to bound from below the supremum over hazard rates by an integral with a Gaussian probability measure for the parameters defining considered hazard rates. Write

$$\begin{aligned}
R_s &\geq \inf_{\hat{\vec{v}}_s} \int_{V_s} E_{\vec{v}_s} \left\{ \sum_{k=0}^{s-1} \sum_{j=J_*}^{J} (\hat{v}_{skj} - v_{skj})^2 \right\} \left[\prod_{r=0}^{s-1} g(\vec{v}_{sr}|\vec{\tau}_{sr}) \right] \mathrm{d}\vec{v}_s \\
&= \inf_{\hat{\vec{v}}_s} \sum_{k=0}^{s-1} \int_{V_{-sk}} \left[\int_{V_{sk}} \left[E_{\vec{v}_s} \left\{ \sum_{j=J_*}^{J} (\hat{v}_{skj} \right.\right.\right. \\
&\quad \left.\left. - v_{skj})^2 \right\} g(\vec{v}_{sk}|\vec{\tau}_{sk}) \mathrm{d}\vec{v}_{sk} \right] \left[\prod_{r\in\{0,\ldots,s-1\}\backslash k} g(\vec{v}_{sr}|\vec{\tau}_{sr}) \right] \Bigg] \mathrm{d}\vec{v}_{-sk} \\
&\geq \sum_{k=0}^{s-1} \int_{V_{-sk}} \left[\inf_{\hat{\vec{v}}_{sk}} \int_{V_{sk}} \left[E_{\vec{v}_s} \left\{ \sum_{j=J_*}^{J} (\hat{v}_{skj} \right.\right.\right. \\
&\quad \left.\left. - v_{skj})^2 \right\} g(\vec{v}_{sk}|\vec{\tau}_{sk}) \mathrm{d}\vec{v}_{sk} \right] \prod_{r\in\{0,\ldots,s-1\}\backslash k} g(\vec{v}_{sr}) \Bigg] \mathrm{d}\vec{v}_{-sk} \\
&\geq \sum_{k=0}^{s-1} \int_{V_{-sk}} \left[\inf_{\hat{\vec{v}}_{sk}\in V_{sk}} \int_{V_{sk}} \left[E_{\vec{v}_s} \left\{ \sum_{j=J_*}^{J} (\hat{v}_{skj} \right.\right.\right. \\
&\quad \left.\left. - v_{skj})^2 \right\} g(\vec{v}_{sk}|\vec{\tau}_{sk}) \mathrm{d}\vec{v}_{sk} \right] \prod_{r\in\{0,\ldots,s-1\}\backslash k} g(\vec{v}_{sr}) \Bigg] \mathrm{d}\vec{v}_{-sk} \\
&=: \sum_{k=0}^{s-1} \int_{V_{-sk}} R_{sk}(\vec{v}_{-sk}) \left[\prod_{r\in\{0,\ldots,s-1\}\backslash k} g(\vec{v}_{sr}|\vec{\tau}_{sr}) \right] \mathrm{d}\vec{v}_{-sk}.
\end{aligned} \tag{46}$$

Note that in the fourth line of (46), the infimum over all possible estimates $\hat{\vec{v}}_{sk}$ can be replaced by the infimum over $\hat{\vec{v}}_{sk} \in V_{sk}$ because the estimand $\vec{v}_{sk}$ belongs to the set V_{sk}.

Step 5 is to replace in R_{sk} the integration over V_{sk} by integration over a larger set $\ddot{V}_{sk}$ because this simplifies analysis of the term. Using $V_{sk} = \ddot{V}_{sk} \backslash (\ddot{V}_{sk} \cap \dot{V}_{sk}^c)$ (here $\dot{V}_{sk}^c$ denotes a set complementary to $\dot{V}_{sk}$), we write

$$R_{sk}(\vec{v}_{-sk}) \geq \inf_{\hat{\vec{v}}_{sk}\in V_{sk}} \int_{\ddot{V}_{sk}} E_{\vec{v}_s} \left\{ \sum_{j=J_*}^{J} (\hat{v}_{skj} - v_{skj})^2 \right\} g(\vec{v}_{sk}|\vec{\tau}_{sk}) \mathrm{d}\vec{v}_{sk}$$

$$- \sup_{\hat{\vec{v}}_{sk} \in V_{sk}} \int_{\dot{V}_{sk}^c} E_{\vec{v}_s} \left\{ \sum_{j=J_*}^{J} (\hat{v}_{skj} - v_{skj})^2 \right\} g(\vec{v}_{sk}|\vec{\tau}_{sk}) \mathrm{d}\vec{v}_{sk}. \quad (47)$$

Let us show that the second term in (47) is of order $o_n^*(1)n^{-2\alpha/(2\alpha+1)}$. By the Cauchy inequality $\sum_{j=J_*}^{J}(\hat{v}_{skj} - v_{skj})^2 \leq 2\sum_{j=J_*}^{J} \hat{v}_{skj}^2 + 2\sum_{j=J_*}^{J} v_{skj}^2$, and for $\hat{\vec{v}}_{sk} \in V_{sk}$ we have $\sum_{j=J_*}^{J} \hat{v}_{skj}^2 \leq s^{-2\alpha}[\ln(n)]^{2\alpha} J^{-2\alpha} Q_{sk}$. Using these relations and Cauchy-Schwarz inequality, we can write

$$\begin{aligned}
&\sup_{\hat{\vec{v}}_{sk} \in V_{sk}} \int_{\dot{V}_{sk}^c} E_{\vec{v}_s} \left\{ \sum_{j=J_*}^{J} (\hat{v}_{skj} - v_{skj})^2 \right\} g(\vec{v}_{sk}|\vec{\tau}_{sk}) \mathrm{d}\vec{v}_{sk} \\
&\quad \leq 2s^{-2\alpha}[\ln(n)]^{2\alpha} J^{-2\alpha} Q_{sk} \int_{\dot{V}_{sk}^c} g(\vec{v}_{sk}|\vec{\tau}_{sk}) \mathrm{d}\vec{v}_{sk} \\
&\quad\quad +2\left[\int_{\dot{V}_{sk}^c} \left[\sum_{j=J_*}^{J} v_{skj}^2 \right]^2 g(\vec{v}_{sk}|\vec{\tau}_{sk}) d\vec{v}_{sk} \right]^{1/2} \left[\int_{\dot{V}_{sk}^c} g(\vec{v}_{sk}|\vec{\tau}_{sk}) d\vec{v}_{sk} \right]^{1/2}. \quad (48)
\end{aligned}$$

Remember that $g(\vec{v}_{sk}|\vec{\tau}_{sk}) = \prod_{j=J_*}^{J} g(v_{skj}|\tau_{skj})$. Then using $[\sum_{j=J_*}^{J} v_{skj}^2]^2 \leq (J - J_* + 1)\sum_{j=J_*}^{J} v_{skj}^4$, we get

$$\begin{aligned}
&\int_{\dot{V}_{sk}^c} \left[\sum_{j=J_*}^{J} v_{skj}^2 \right]^2 g(\vec{v}_{sk}|\vec{\tau}_{sk}) \mathrm{d}\vec{v}_{sk} \leq (J - J_* + 1) \sum_{j=J_*}^{J} \int_{-\infty}^{\infty} v_{skj}^4 g(v_{skj}|\tau_{skj}) dv_{skj} \\
&\quad = 3(J - J_* + 1) \sum_{j=J_*}^{J} \tau_{skj}^4 \leq 3q^2 \mathcal{I}_{sk}^{-2} J^2 n^{-2}. \quad (49)
\end{aligned}$$

Now we are evaluating integral $\int_{\dot{V}_{sk}^c} g(\vec{v}_{sk}|\vec{\tau}_{sk}) d\vec{v}_{sk}$. To make some elementary calculations, we will use the following relation:

$$\sum_{r=k}^{K-1} \psi(r) \leq \int_k^K \psi(x)\mathrm{d}x \leq \sum_{r=k+1}^{K} \psi(r), \quad (50)$$

which holds for any nonnegative and not decreasing function $\psi(x)$ and any pair of integers k and K such that $k < K - 1$. Using

$$\sum_{j=J_*}^{J} (\pi j)^{2\alpha} \tau_{skj}^2 = \pi^{2\alpha} \mathcal{I}_{sk}^{-1} n^{-1} (1 - 3q^{-1}) \sum_{j=J_*}^{J} j^{2\alpha} \max(q^{-1}, \min(q, (J/j)^\alpha - 1)), \quad (51)$$

and (50) we can write for all sufficiently large n (large uniformly over k),

$$\sum_{j=J_*}^{J} j^{2\alpha}\max(q^{-1},\min(q,(J/j)^{\alpha}-1)) \le q^{-1}\sum_{J(1+q^{-1})^{-1/\alpha}<j\le J} j^{2\alpha}$$
$$+\sum_{j=J_*}^{J}[J^{\alpha}j^{\alpha}-j^{2\alpha}]$$
$$\le q^{-1}(2\alpha+1)^{-1}J^{2\alpha+1}[1-(1+q^{-1})^{-(2\alpha+1)/\alpha}$$
$$+(1+\alpha)^{-1}J^{\alpha}[J^{\alpha+1}-(J_*-1)^{\alpha+1}]-(2\alpha+1)^{-1}[(J-1)^{2\alpha+1}-J_*^{2\alpha+1}]$$
$$\le \frac{J^{2\alpha+1}\alpha}{(\alpha+1)(2\alpha+1)}+q^{-1}(2\alpha+1)^{-1}J^{2\alpha+1}\le ns^{-2\alpha}\pi^{-2\alpha}\mathcal{I}_sQ[1+3q^{-1}]. \tag{52}$$

Using (52) in (51), we conclude that for all sufficiently large n

$$\sum_{j=J_*}^{J}(\pi j)^{2\alpha}\tau_{skj}^2 \le (1-s^{-1})(1-3q^{-1})(1+3q^{-1})s^{-2\alpha}I_{sk}^{-1}I_sQ \le (1-9q^{-2})s^{-2\alpha}Q_{sk}. \tag{53}$$

Let Z_{skj} be independent normal random variables with zero mean and variance τ_{skj}^2, that is, $g(z|\tau_{skj}^2)$ is the probability density of Z_{skj}. Then, the integral of interest can be written as a corresponding probability, and then (53) and the Chebyshev inequality can be used to evaluate the probability. Following these steps, we write

$$\int_{\dot{V}_{sk}^c} g(\vec{v}_{sk}|\vec{\tau}_{sk})d\vec{v}_{sk} = \Pr\left(\sum_{j=J_*}^{J}(\pi j)^{2\alpha}Z_{skj}^2 \ge s^{-2\alpha}Q_{sk}\right)$$
$$\le \Pr\left(\sum_{j=J_*}^{J}(\pi j)^{2\alpha}(Z_{skj}^2-\tau_{skj}^2)\ge 9q^{-2}s^{-2\alpha}Q_{sk}\right)$$
$$< \frac{\sum_{j=J_*}^{J}(\pi j)^{4\alpha}\tau_{skj}^4}{[q^{-2}s^{-2\alpha}Q_{sk}]^2} < \frac{\pi^{4\alpha}(4\alpha+1)^{-1}J^{4\alpha+1}q^2n^{-2}\mathcal{I}_{sk}^{-2}}{[q^{-2}s^{-2\alpha}Q_{sk}]^2}$$
$$= o_n^*(1)q^6s^{4\alpha+2}n^{-1/(2\alpha+1)}. \tag{54}$$

Let us stress that in (54) the sequence $o_n^*(1)$ vanishes uniformly over $k\in\{0,1,\ldots,s-1\}$. Using (54), together with (49), in (48), we get

$$\sup_{\hat{\vec{v}}_{sk}\in V_{sk}}\int_{\dot{V}_{sk}^c} E_{\vec{v}_s}\left\{\sum_{j=J_*}^{J}(\hat{v}_{skj}-v_{skj})^2\right\}g(\vec{v}_{sk}|\vec{\tau}_{sk})\mathrm{d}\vec{v}_{sk}$$
$$= o_n^*(1)s^{-1}n^{-2\alpha/(2\alpha+1)},\quad q\in(3,\ln(n)). \tag{55}$$

In the last line, it is stressed that the result is established for q satisfying that specific restriction.

Step 6 is the beginning of evaluation of the first term in the right side of (47), which is the main term in the lower bound. We would like to convert evaluation of the term into evaluation of the risk for a particular parameter v_{skj}. To do this, we use notation (36), definition of the sets $\ddot{V}_{sk}$, V_{-skj} and $V_{skj} := V_{-skj} \times \{v_{skj} : |v_{skj}| < s^2 n^{-1/2}\}$, presented in the Notation, and write

$$\begin{aligned}
&\inf_{\hat{\vec{v}}_{sk} \in V_{sk}} \int_{\ddot{V}_{sk}} E_{\vec{v}_s} \left\{ \sum_{j=J_*}^{J} (\hat{v}_{skj} - v_{skj})^2 \right\} g(\vec{v}_{sk}|\vec{\tau}_{sk}) \mathrm{d}\vec{v}_{sk} \\
&\geq \sum_{j=J_*}^{J} \inf_{\hat{\vec{v}}_{sk} \in V_{sk}} \int_{\ddot{V}_{sk}} E_{\vec{v}_s} \{(\hat{v}_{skj} - v_{skj})^2\} g(\vec{v}_{sk}|\vec{\tau}_{sk}) \mathrm{d}\vec{v}_{sk} \\
&\geq \sum_{j=J_*}^{J} \inf_{\hat{\vec{v}}_{sk} \in V_{sk}} \int_{\vec{v}_{sk} \in V_{skj}} E_{\vec{v}_s} \{(\hat{v}_{skj} - v_{skj})^2\} g(\vec{v}_{sk}|\vec{\tau}_{sk}) \mathrm{d}\vec{v}_{sk} \\
&=: \sum_{j=J_*}^{J} R_{skj}(\vec{v}_{-sk}), \quad \vec{v}_{-sk} \in V_{-sk}.
\end{aligned} \tag{56}$$

Now we are making a number of steps to evaluate $R_{skj}(\vec{v}_{-sk})$ for $\vec{v}_{-sk} \in V_{-sk}$. Note that the expectation in (56) is taken with respect to the joint probability density

$$p(x^n|\vec{v}_s) = \prod_{l=1}^{n} h(x_l|\vec{v}_s) \mathrm{e}^{-\int_0^{x_l} h(y|\vec{v}_s)\mathrm{d}y}, \tag{57}$$

where $x^n = (x_1, \ldots, x_n)$ and $h(x|\vec{v}_s)$ is defined in (36). Let us also remember notation $\vec{v}_{-skj,0}$ for a vector $\vec{v}_s$ with its element v_{skj} replaced by zero, and $U_{sk} := V_{-sk} \times V_{-skj}$.

Step 7 is to replace in $R_{skj}(\vec{v}_{-sk})$ the expectation based on density $p(\vec{x}|\vec{v}_s)$ by the expectation based on density $p(\vec{x}|\vec{v}_{-skj,0})$. Write

$$E_{\vec{v}_s}\{(\hat{v}_{skj} - v_{skj})^2\} = E_{\vec{v}_{-skj,0}} \left\{ \frac{p(X^n|\vec{v}_s)}{p(X^n|\vec{v}_{-skj,0})} (\hat{v}_{skj} - v_{skj})^2 \right\}, \quad \vec{v}_s \in U_{sk}. \tag{58}$$

Step 8 is to find for all $\vec{v}_s \in U_{sk}$ the limiting distribution of the likelihood ratio $p(X^n|\vec{v}_s)/p(X^n|\vec{v}_{-skj,0})$. We would like to show that uniformly over all considered parameters

$$\begin{aligned}
&\frac{p(x^n|\vec{v}_s)}{p(x^n|\vec{v}_{-skj,0})} = \mathrm{e}^{u\zeta_{skj} - (1/2)u^2(1+\mu_{skj})} \text{ where } \zeta_{skj} \xrightarrow{\mathcal{D}} N(0,1), \quad \mu_{skj} \xrightarrow{\mathcal{P}} 0, \\
&\qquad u = (\mathcal{I}_{sk} n)^{1/2} v_{skj}.
\end{aligned} \tag{59}$$

Note that in (59), we are dealing with convergence of a series of underlying distributions (they change with n) and a distribution depends on the parameters. As a result, it is a tedious path to prove (39) directly. Instead we are using handy Theorem 3.1$'$ and Remark 3.2 in Ibragimov and Khasminskii (1981) which establish (39) if the following conditions hold for all $\vec{v}_s \in U_{sk}$, all considered k and j, and all sufficiently large n: (a) The probability density $p(x|\vec{v}_s)$ is continuous in v_{skj} and $\sqrt{p(x|\vec{v}_s)}$ is differentiable with respect to v_{skj}; (b) The following relations hold uniformly over considered k, j and $\vec{v}_s$:

$$E_{\vec{v}_{-skj,0}}\{[p'(X|\vec{v}_{-skj,0})/p(X|\vec{v}_{-skj,0})]^2\} = \mathcal{I}_{sk}(1 + o_n^*(1)), \tag{60}$$

where $p'(x|\vec{v}_s) := \partial p(x|\vec{v}_s)/\partial v_{skj}$,

$$|\ln(\mathcal{I}_{sk})| \le C^* < \infty, \tag{61}$$

and

$$\int_0^\infty [\partial^2 \sqrt{p(x|\vec{v}_s)}/\partial v_{skj}^2]^2 \mathrm{d}x = o_n^*(1)n. \tag{62}$$

We begin verification of the conditions with part (a). Let us calculate derivative of the square root of the density. Write using (36) and (57) for $\vec{v}_s \in U_{sk}$

$$\partial\sqrt{p(x|\vec{v}_s)}/\partial v_{skj} = \frac{\varphi_{skj}(x)\phi_{sk}(x)\mathrm{e}^{-\int_0^x h(y|\vec{v}_s)\mathrm{d}y}}{2\sqrt{p(x|\vec{v}_s)}} - (1/2)\int_0^x h'(y|\vec{v}_s)\mathrm{d}y\sqrt{p(x|\vec{v}_s)}, \tag{63}$$

where $h'(y|\vec{v}_s) := \partial h(y|\vec{v}_s)/\partial v_{skj} = \varphi_{skj}(y)\phi_{sk}(y)$. Note that, for the studied $\vec{v}_s \in U_{sk}$, the density $p(x|\vec{v}_s)$ is uniformly bounded below from zero at the interval $[a + k/s, a + (k+1)/s]$ which is the support for $\phi_{sk}(x)$, and this verifies part (a).

Now let us verify part (b), and we begin with (60). Set $h'(x|\vec{v}_s) := \partial h(x|\vec{v}_s)/\partial v_{skj}$ and write

$$\begin{aligned}
\int_0^\infty \frac{[p'(x|\vec{v}_{-skj,0})]^2}{p(x|\vec{v}_{-skj,0})}\mathrm{d}x &= \int_0^\infty \Bigg[\frac{h'(x|\vec{v}_{-skj,0})}{h(x|\vec{v}_{-skj,0})} \\
&\quad - \int_0^x h'(y|\vec{v}_{-skj,0})\mathrm{d}y\Bigg]^2 h(x|\vec{v}_{-skj},0)\mathrm{e}^{-\int_0^x h(y|\vec{v}_{-skj,0})\mathrm{d}y}\mathrm{d}x \\
&= \int_0^\infty [h'(x|\vec{v}_{-skj,0})]^2 h^{-1}(x|\vec{v}_{-skj,0})\mathrm{e}^{-\int_0^x h(y|\vec{v}_{-skj,0})\mathrm{d}y}\mathrm{d}x \\
&\quad -2\int_0^\infty h'(x|\vec{v}_{-skj,0})\left[\int_0^x h'(y|\vec{v}_{-skj,0})\mathrm{d}y\right] \\
&\quad \times\mathrm{e}^{-\int_0^x h(y|\vec{v}_{-skj,0})\mathrm{d}y}\mathrm{d}x \\
&\quad +\int_0^\infty \left[\int_0^x h'(y|\vec{v}_{-skj,0})\mathrm{d}y\right]^2 h(x|\vec{v}_{-skj,0})\mathrm{e}^{-\int_0^x h(y|\vec{v}_{-skj,0})\mathrm{d}y}\mathrm{d}x \\
&=: A_1 + A_2 + A_3.
\end{aligned} \tag{64}$$

Let us evaluate the three terms in (64). Remember that we are considering $\vec{v}_s \in U_{sk}$ and write for A_1,

$$A_1 = \int_0^\infty [\varphi_{skj}(x)\phi_{sk}(x)]^2 h^{-1}(x|\vec{v}_{-skj,0})\mathrm{e}^{-\int_0^x h(y|\vec{v}_{-skj,0})\mathrm{d}y}\mathrm{d}x$$
$$= [h_0^{-1}(a+k/s)\mathrm{e}^{-\int_0^{a+k/s} h_0(y)\mathrm{d}y} + o_n^*(1)]\int_{a+k/s}^{a+(k+1)/s}[\varphi_{skj}(x)\phi_{sk}(x)]^2\mathrm{d}x. \tag{65}$$

In the last equality, we used continuity of the pivot on $[a, a+1]$, which is the case for all four theorems, and the fact that the flattop kernel $\phi_{sk}(x)$ vanishes beyond $[a+k/s, a+(k+1)/s]$, and that $h(x|\vec{v}_{-skj,0})-h_0(x) = o_n^*(1)$ uniformly over $\vec{v}_s \in U_{sk}$. For the second factor in the right side of (65), we can write using definition of the flattop kernel $\phi_{sk}(x)$ and that φ_{skj} are elements of the cosine basis on $[a+k/s, a+(k+1)/s]$,

$$\int_{a+k/s}^{a+(k+1)/s}[\varphi_{skj}(x)\phi_{sk}(x)]^2\mathrm{d}x = 1$$
$$+\int_{a+k/s}^{a+(k+1)/s}\varphi_{skj}^2(x)(\phi_{sk}^2(x)-1)\mathrm{d}x = 1 + o_n^*(1)/\ln(n). \tag{66}$$

Using (66) in (65), we conclude that

$$A_1 = \mathcal{I}_{sk}(1+o_n^*(1)). \tag{67}$$

To estimate A_2, remember that $h'(y|\vec{v}_{-skj,0}) = \varphi_{skj}(y)\phi_{sk}(y)$, $\phi_{sk}(y) = 0$ and $\mathrm{d}\phi_{sk}(y)/\mathrm{d}y = 0$ beyond the interval $[a+k/s, a+(k+1)/s]$. Also, using integration by parts and that $j \ge J_* > C^* n^{1/(2\alpha+1)}/(s\ln(n))$, $C^* > 0$, we can write

$$\int_0^x h'(y|\vec{v}_s)\mathrm{d}y$$
$$= \int_{a+k/s}^x \varphi_{skj}(y)\phi_{sk}(y)\mathrm{d}y = (\pi sj)^{-1}(2s)^{1/2}\Big[\sin(\pi j[s(x-a)-k])\phi_{sk}(x)$$
$$-\int_{a+k/s}^x \sin(\pi j[s(y-a)-k])[\mathrm{d}\phi_{sk}(y)/\mathrm{d}y]\mathrm{d}y\Big]$$
$$= o_n^*(1)n^{-1/(2\alpha+2)}, \quad x > a+k/s. \tag{68}$$

We conclude that $A_2 = o_n^*(1)$. For A_3, we can write using (57) and (68) that

$$A_3 = \int_0^\infty\left[\int_0^x \varphi_{skj}(y)\phi_{sk}(y)\mathrm{d}y\right]^2 p(x|\vec{v}_{-skj,0})\mathrm{d}x = o_n^*(1). \tag{69}$$

Relation (60) is verified. Now let us remember that by its definition $\mathcal{I}_{sk} = h_0(a+k/s)\mathrm{e}^{\int_0^{a+k/s} h_0(y)\mathrm{d}y}$, and then (61) is the property of a pivot $h_0(x)$ which is valid for the considered memoryless pivot (as well as for pivots considered in other theorems of Sect. 2).

Finally, let us check (62). Using (63), we can write for any $\vec{v}_s \in U_{sk}$,

$$2\partial^2\sqrt{p(x|\vec{v}_s)}/\partial v_{skj}^2 = \varphi_{skj}(x)\phi_{sk}(x)\mathrm{e}^{-\int_0^x h(y|\vec{v}_s)\mathrm{d}y} \left[-\int_0^x h'(y|\vec{v}_s)\mathrm{d}y\sqrt{p(x|\vec{v}_s)} - \partial\sqrt{p(x|\vec{v}_s)}/\partial v_{skj}]p^{-1}(x|\vec{v}_s) - \int_0^x \varphi_{skj}(y)\phi_{sk}(y)\mathrm{d}y[\partial\sqrt{p(x|\vec{v}_s)}/\partial v_{skj}\right]. \tag{70}$$

Then using (68), $h'(x|\vec{v}_s) = \varphi_{skj}(x)\phi_{sk}(x)$, $p(x|\vec{v}_s) > C^* > 0$ for $x \in [a+k/s, a+(k+1)/s]$ and $\vec{v}_s \in U_{sk}$, and $\int_0^\infty p(x|\vec{v}_s)\mathrm{d}x = 1$, we conclude that (62) holds. We checked that all conditions required for the validity of (59) hold.

Step 9 is to evaluate $R_{skj}(\vec{v}_{-sk})$ defined in (56) using (58), (59) and Fatou's lemma. Write

$$R_{skj}(\vec{v}_{-sk}) \geq \int_{V_{-skj}} g(\vec{v}_{-skj}|\vec{\tau}_{-skj})\left[\inf_{\hat{\bar{v}}_{skj}} \int_{|v_{skj}|<s^2n^{-1/2}} E_{\vec{v}_s}\left\{(\hat{v}_{skj} - v_{skj})^2\right\} g(v_{skj}|\tau_{skj})\mathrm{d}v_{skj}\right]\mathrm{d}\vec{v}_{-skj}$$
$$=: \int_{V_{-skj}} g(\vec{v}_{-skj}|\vec{\tau}_{-skj})\dot{R}_{skj}(\vec{v}_{-skj}, \vec{v}_{-sk})\mathrm{d}\vec{v}_{-skj}. \tag{71}$$

Now we are evaluating term $\dot{R}_{skj}(\vec{v}_{-skj})$ in right side of (71) using (59), definition of the Gaussian density $g(v_{skj}|\tau_{skj})$ and making change of variable $u = [\mathcal{I}_{sk}n]^{1/2}v_{skj}$. Write

$$\dot{R}_{skj}(\vec{v}_{-skj}, \vec{v}_{-sk}) = \inf_{\hat{\bar{v}}_{skj}} \int_{|u|\leq s^2 I_{sk}^{1/2}} [2\pi\tau_{skj}^2\mathcal{I}_{sk}n]^{-1/2} E_{\vec{v}_{-skj,0}}\{\mathrm{e}^{u\zeta_{skj}-(1/2)u^2[1+\mu_{skj}+(\tau_{skj}^2 I_{sk}n)^{-1}]}[\hat{v}_{skj} - u(I_{sk}n)^{-1/2}]^2\}\mathrm{d}u. \tag{72}$$

Introduce a constant $b \in (3, (1/2)s^2 I_{sk}^{1/2})$, which exists for all sufficiently large n, and set $t := \tau_{skj}^2\mathcal{I}_{sk}n$. Continue (72),

$$\dot{R}_{skj}(\vec{v}_{-skj}, \vec{v}_{-sk}) \geq [2\pi\tau_{skj}^2\mathcal{I}_{sk}^3 n^3]^{-1/2} E_{\vec{v}_{-skj,0}}\left\{\inf_{\hat{v}_{skj}} \int_{-2b}^{2b} \mathrm{e}^{u\zeta_{skj}-(1/2)u^2(1+t^{-1})} \times \min(b^{1/2}, (u-(\mathcal{I}_{sk}n)^{1/2}\hat{v}_{skj})^2)\mathrm{e}^{-(1/2)u^2\mu_{skj}} I(|\zeta_{skj}|(1+t^{-1})^{-1} < b)\right\}\mathrm{d}u. \tag{73}$$

Now we change the variable of integration from u to $y = u - \zeta_{skj}(1+t^{-1})^{-1}$. We also use the known uniform convergence of μ_{skj} to zero, and a familiar inequality $\inf_{r\in(-\infty,\infty)} \int_{-b}^{b} \mathrm{e}^{-\lambda u^2} \min(C, (u-r)^2)\mathrm{d}u = \int_{-b}^{b} \mathrm{e}^{-\lambda u^2} \min(C, u^2)\mathrm{d}y$ which holds for any $\lambda > 0$ and $C > 0$. We continue (73),

$$\begin{aligned}
\dot{R}_{skj}(\vec{v}_{-skj}, \vec{v}_{-sk}) &\geq [2\pi t]^{-1/2}(\mathcal{I}_{sk}n)^{-1} E_{\vec{v}_{-skj,0}} \left\{ \inf_{\hat{v}_{skj}} \left[\int_{-b}^{b} \mathrm{e}^{-y^2(1+t^{-1})^{-1}/2} \right.\right. \\
&\quad \times \min(b^{1/2}, [y + \zeta_{skj}(1+t^{-1})^{-1} - (\mathcal{I}_{sk}n)^{1/2}\hat{v}_{skj}]^2 \mathrm{d}y \bigg] \mathrm{e}^{\zeta_{skj}^2(1+t^{-1})/2} I(|\zeta_{skj}| \\
&\quad \left. < (1+t^{-1})b)) \right\} (1+o_n^*(1)) \\
&\geq [2\pi t]^{-1/2}(\mathcal{I}_{sk}n)^{-1} \int_{-b}^{b} \mathrm{e}^{-y^2(1+t^{-1})/2} \min(b^{1/2}, y^2)\mathrm{d}y \\
&\quad \times E_{\vec{v}_{-skj,0}}\{\mathrm{e}^{\zeta_{skj}^2(1+t^{-1})^{-1}/2} I(|\zeta_{skj}| < (1+t^{-1})b)\}(1+o_n^*(1)). \qquad (74)
\end{aligned}$$

Here, $o_n^*(1) \to 0$, $n \to \infty$ uniformly over all considered parameters. Remember that

$$t = \tau_{skj}^2 \mathcal{I}_{sk} n = (1 - 3q^{-1}) \max(q^{-1}, \min(q, (J/j)^{\alpha} - 1)). \qquad (75)$$

Also,

$$\begin{aligned}
&[2\pi(1+t^{-1})^{-1}]^{-1/2} \int_{-b}^{b} \mathrm{e}^{-y^2(1+t^{-1})/2} \min(b^{1/2}, y^2)\mathrm{d}y \\
&\quad \geq (1+t^{-1})^{-1} - [2\pi(1+t^{-1})^{-1}]^{-1/2} b \int_{|y|>b^{1/4}} \mathrm{e}^{-y^2(1+t^{-1})/2}\mathrm{d}y \\
&\quad = (1+t^{-1})^{-1} - o_b^*(1),
\end{aligned}$$

where $o_b^*(1) \to 0$, $b \to \infty$ uniformly over all considered parameters. Further, using Fatou's lemma and (59), we get

$$\begin{aligned}
&E_{\vec{v}_{-skj,0}}\{\mathrm{e}^{\zeta_{skj}^2(1+t^{-1})^{-1}/2} I(|\zeta_{skj}| < (1+t^{-1})b)\} \\
&\quad \geq \int_{-\infty}^{\infty} (2\pi)^{-1/2} \mathrm{e}^{-(1/2)z^2[1-(1+t^{-1})^{-1}]} dz - |o_b^*(1)| = (1+t)^{1/2}(1+o_b^*(1)),
\end{aligned}$$

where in the last relation we used $1 - (1+t^{-1})^{-1} = t^{-1}(1+t^{-1})^{-1} = (1+t)^{-1}$. Using these results, together with (75), in (74) we get

$$\begin{aligned}
\dot{R}_{skj}(\vec{v}_{-skj}, \vec{v}_{-sk}) &\geq (\mathcal{I}_{sk}n)^{-1}[(1+t^{-1})t]^{-1/2}(1+t^{-1})^{-1}(1+t)^{1/2}(1+o_b^*(1) \\
&\quad + o_n^*(1)) \\
&= (\mathcal{I}_{sk}n)^{-1}[t/(1+t)](1+o_b^*(1)+o_n^*(1)) = \frac{(\mathcal{I}_{sk}n)^{-1}\tau_{skj}^2}{(\mathcal{I}_{sk}n)^{-1}+\tau_{skj}^2}(1+o_b^*(1)
\end{aligned}$$

$$+o_n^*(1)). \tag{76}$$

To finish step 9, we plug (76) in (71) and get for $b \in (3, (1/2)s^2\mathcal{I}_{sk})$ and all sufficiently large n

$$R_{skj}(\vec{v}_{-sk}) \geq \left[\int_{V_{-skj}} g(\vec{v}_{-skj}|\vec{\tau}_{-skj}) d\vec{v}_{-skj} \right] \frac{(\mathcal{I}_{sk}n)^{-1}\tau_{skj}^2}{(\mathcal{I}_{sk}n)^{-1} + \tau_{skj}^2}(1$$
$$+ o_b^*(1) + o_n^*(1))I(b \in (3, (1/2)s^2\mathcal{I}_{sk})). \tag{77}$$

Step 10 is to evaluate the integral in (77). Using notation $Z_1, Z_2, \ldots$ for independent standard normal random variables, and remembering definitions of J and τ_{skj}, we can write the integral as the probability of the corresponding event, and then we can use Theorem 6.2 in Kahane (1985) to evaluate that probability. Write

$$\begin{aligned}
&\int_{V_{-skj}} g(\vec{v}_{-skj}|\vec{\tau}_{-skj}) d\vec{v}_{-skj} \\
&= \Pr\left(\max_{x \in [a, a+1]} \left| \sum_{j \in \{J_*, \ldots, J\} \setminus j} \tau_{skj} Z_j \varphi_{skj}(x) \right| \leq (1/2)[s^4 \ln(n)]^{1/2} n^{-\alpha/(2\alpha+1)} \right) \\
&\geq \Pr\left(\left| \sum_{j=J_*}^{J} Z_j \varphi_{skj}(x) \right| \leq C^* n^{1/2} q^{-1/2} [s^4 \ln(n)]^{1/2} n^{-\alpha/(2\alpha+1)} \right) \\
&\geq [1 - o_n^*(1) n^{-2/(2\alpha+1)}] I(q \in (3, s)).
\end{aligned} \tag{78}$$

Note that absolutely similarly we establish inequality

$$\int_{\ddot{V}_{sk}} g(\vec{v}_{sk}|\vec{\tau}_{sk}) d\vec{v}_{sk} \geq [1 - o_n^*(1) n^{-2/(2\alpha+1)}] I(q \in (3, s)). \tag{79}$$

Remember that $s \to \infty$ as $n \to \infty$, so for all sufficiently large n and all $k \in \{0, 1, \ldots, s-1\}$ we have $(3, s) \subset (3, (1/2)s^2\mathcal{I}_{sk})$. Keeping this in mind, together with using (78) in (77) we conclude that for all sufficiently large n, $b \in (3, s)$ and $q \in (3, s)$, the following relation holds:

$$R_{skj}(\vec{v}_{-sk}) \geq \frac{(\mathcal{I}_{sk}n)^{-1}\tau_{skj}^2}{(\mathcal{I}_{sk}n)^{-1} + \tau_{skj}^2}[1 + o_n^*(1) + o_q^*(1) + o_b^*(1)]. \tag{80}$$

Now we plug (80) in (56) and get

$$\sum_{j=J_*}^{J} R_{skj}(\vec{v}_{-sk}) \geq \left[\sum_{j=J_*}^{J} \frac{(\mathcal{I}_{sk}n)^{-1}\tau_{skj}^2}{(\mathcal{I}_{sk}n)^{-1} + \tau_{skj}^2} \right] (1 + o_n^*(1) + o_q^*(1) + o_b^*(1)). \tag{81}$$

Step 11 is to evaluate the sum in the right side of (81). Remember definitions of J (see Notation), $P(\alpha, Q)$ [see line (2.3)], and that $\tau^2_{skj} = n^{-1}\mathcal{I}^{-1}_{sk}(1 - 3q^{-1})\max(q^{-1}, \min(q, (J/j)^\alpha - 1))$. Write for $q \in (3, s)$,

$$\begin{aligned}
&\sum_{j=J_*}^{J} \frac{(\mathcal{I}_{sk}n)^{-1}\tau^2_{skj}}{(\mathcal{I}_{sk}n)^{-1} + \tau^2_{skj}} \\
&\quad = (\mathcal{I}_{sk}n)^{-1} \sum_{j=J_*}^{J} \frac{(1-3q^{-1})\max(q^{-1}, \min(q, (J/j)^\alpha - 1))}{1 + (1-3q^{-1})\max(q^{-1}, \min(q, (J/j)^\alpha - 1))} \\
&\quad \geq (\mathcal{I}_{sk}n)^{-1} \sum_{j=J_*}^{J} [1 - (j/J)^\alpha][1 + o^*_q(1) + o^*_n(1)] \\
&\quad = (\mathcal{I}_{sk}n)^{-1} J(1 - (1+\alpha)^{-1})[1 + o^*_q(1) + o^*_n(1)] \\
&\quad = \mathcal{I}^{-1}_{sk}(s^{-2\alpha}\mathcal{I}_s)^{1/(2\alpha+1)} n^{-2\alpha/(2\alpha+1)} P(\alpha, Q)[1 + o^*_q(1) + o^*_n(1)]. \qquad (82)
\end{aligned}$$

Now we use (82) in (81) and then the result in (56). The obtained inequality, together with (55), yields the following lower bound for the right side of (47), which holds for any $b \in (3, s)$ and $q \in (3, s)$,

$$R_{sk}(\vec{v}_{-sk}) \geq \mathcal{I}^{-1}_{sk}(s^{-2\alpha}\mathcal{I}_s)^{1/(2\alpha+1)} n^{-2\alpha/(2\alpha+1)} P(\alpha, Q)[1 + o^*_n(1) + o^*_q(1) + o^*_b(1)]. \qquad (83)$$

Step 12 is to plug (83) in (46). Before doing this, let us note that using inequalities (54), (79) and the classical probability inequality $P(A \cap B) \geq P(A) + P(B) - 1$, we can write for any $b \in (3, s)$ and $q \in (3, s)$,

$$\begin{aligned}
&\int_{V_{-sk}} \left[\prod_{r \in \{0,1,\ldots,s-1\}\setminus k} g(\vec{v}_{sr}|\vec{\tau}_{sr}) \right] \mathrm{d}\vec{v}_{-sk} \\
&\quad \geq (1 - C^* q^6 n^{-1/(2\alpha+2)})^{s-1} = 1 - o^*_n(1)\ln^{-1}(n). \qquad (84)
\end{aligned}$$

Also, a direct calculation, based on definition of $\mathcal{I}_s$ and $\mathcal{I}_{sk}$, shows that

$$\begin{aligned}
&\sum_{k=0}^{s-1} \mathcal{I}^{-1}_{sk}(s^{-2\alpha} I_s)^{1/(2\alpha+1)} \\
&\quad = s^{-2\alpha/(2\alpha+1)}\mathcal{I}_s^{1/(2\alpha+1)} \sum_{k=0}^{s-1} \mathcal{I}^{-1}_{sk} = \left[s^{-1} \sum_{k=0}^{s-1} \mathcal{I}^{-1}_{sk} \right]^{2\alpha/(2\alpha+1)} \\
&\quad = \left[s^{-1} \sum_{k=0}^{s-1} h_0(a + k/s) \mathrm{e}^{\int_0^{a+k/s} h_0(v)\mathrm{d}v} \right]^{2\alpha/(2\alpha+1)} \\
&\quad = \left[\int_a^{a+1} h_0(x) \mathrm{e}^{\int_0^x h_0(v)\mathrm{d}v} \mathrm{d}x \right]^{2\alpha/(2\alpha+1)} (1 + o^*_n(1)), \qquad (85)
\end{aligned}$$

where the last equality is valid for any Riemann integrable on $[a, a+1]$ function $h_0(x)$. Remember that we consider a particular $h_0(x) = Q_0(1-1/s^2)$, $x \in [a, a+1]$. This yields that $\int_a^{a+1} h_0(x)\mathrm{e}^{\int_0^x h_0(v)\mathrm{d}v}\mathrm{d}x = G^{-1}(a)(\mathrm{e}^{Q_0}-1)(1+o_n(1))$. Now, we can finish step 12 by utilizing the last relation in (85), then using the obtained inequality in the right side of (83), and then using the obtained inequality and (84) in (46). As a result, we get

$$R_s \geq P(\alpha, Q)([G^{-1}(a)(\mathrm{e}^{Q_0}-1)]n^{-1})^{2\alpha/(2\alpha+1)}[1+o_n^*(1)+o_b^*(1) + o_q^*(1)]I(b \in (3, s))I(q \in (3, s)). \tag{86}$$

Finally, we note that according to (86) constants b and q can be chosen as large as desired for sufficiently large n. □

Proof of Theorem 2 There are two changes in the proof. The first one is in using a different pivot on the interval $[a, a+1]$. Here, we use $h_0(x) = Q_0^*(1-s^{-2})$, $x \in [a, a+1]$. Note that this pivot is plainly Riemann integrable, bounded below from zero and continuous on that interval. The second change is in replacing parameter Q on $[Q - c_1(Q_0^*)^2]/c_2$. Then, following along lines of the previous proof we verify assertion of Theorem 2. □

Proof of Theorem 3 This proof follows along lines of the proof of Theorem 1 if the specified in the theorem pivot is used. □

Proof of Theorem 4 This proof also follows along lines of the proof of Theorem 1, only here a new class $\mathcal{H}_s$ is used to satisfy restrictions of the class $\mathcal{S}_4$. Namely, hazard rates from $\mathcal{S}_4$ have fixed Fourier coefficients (for the cosine series expansion on $[a, a+1]$) on lower frequencies and this is reflected in the new class $\mathcal{H}_s$. Set

$$\mathcal{H}_s = \Bigg\{h: \ h(x) = h_0(x)I(x \notin [a, a+1]) + \Bigg[\sum_{j=0}^{M_n-1}\int_a^{a+1} h_0(u)\varphi_j(u)\mathrm{d}u\varphi_j(x) + \sum_{k=0}^{s-1} f_{sk}(x)\phi_{sk}(x)\Bigg] I(a \leq x \leq a+1), \ \ f_{sk} \in \mathcal{H}_{sk}, \ h(x) \geq 0, x \geq 0\Bigg\}.$$

Let us also note that $M_n = o_n^*(1)J_*$. Also, because Fourier coefficients of the pivot $h_0(x)$, with respect to cosine basis on $[a, a+1]$, are absolutely summable, the pivot is continuous and Riemann integrable on $[a, a+1]$. Remember that it is also assumed that the pivot is bounded below from zero on that interval. Keeping these remarks in mind, we can follow the proof of Theorem 1 and verify assertion of Theorem 4. □

In what follows $\mathcal{S}_1$–$\mathcal{S}_4$ denote four classes of hazard rates studied in Theorems 1–4, respectively (we may skip parameters of a class to simplify notation), and Cs are generic positive constants that do not depend on an underlying h or an underlying function class.

Proof of Theorem 5 We begin with establishing several properties of statistic $\hat{\theta}_j$ defined in (14). Write

$$\hat{\theta}_j = n^{-1}\sum_{l=1}^{n}\frac{\varphi_j(X_l)I(X_l \in [a,a+1])}{G(X_l)} + n^{-1}\sum_{l=1}^{n}\frac{\varphi_j(X_l)I(X_l \in [a,a+1])[G(X_l)-\eta_l/n]}{G^2(X_l)} + n^{-1}\sum_{l=1}^{n}\frac{\varphi_j(X_l)I(X_l \in [a,a+1])[G(X_l)-\eta_l/n]^2}{G^2(X_l)\eta_l/n}. \quad (87)$$

Now we are evaluating the expectation of $\hat{\theta}_j$ by considering 3 terms in (87) in turn. For expectation of the first term, we can write

$$E_h\left\{n^{-1}\sum_{l=1}^{n}\frac{\varphi_j(X_l)I(X_l \in [a,a+1])}{G(X_l)}\right\} = \int_a^{a+1}\frac{h(x)G(x)\varphi_j(x)}{G(x)}\mathrm{d}x = \theta_j. \quad (88)$$

Evaluation of expectation of the second term in (87), we begin with a conditional expectation,

$$E_h\{(G(X_l)-\eta_l/n)|X_l = x\} = E_h\left\{G(x) - n^{-1}\sum_{s\neq l, s=1}^{n} I(X_s \geq x) - n^{-1}\right\} = G(x) - (n-1)n^{-1}G(x) - n^{-1} = n^{-1}(G(x)-1). \quad (89)$$

Now let us show that

$$\min_{i\in\{1,\ldots,4\}}\inf_{h\in\mathcal{S}_i} G(a+1) > 0. \quad (90)$$

Using Cauchy–Schwarz inequality, we can write

$$G(a+1) = G(a)\mathrm{e}^{-\int_a^{a+1}h(x)\mathrm{d}x} \geq G(a)\mathrm{e}^{-[\int_a^{a+1}h^2(x)\mathrm{d}x]^{1/2}}.$$

Now remember that for all four studied function classes $\mathcal{S}_i$ it is assumed that $G(a) > 0$ and $\max_{i\in\{1,\ldots,4\}}\sup_{h\in\mathcal{S}_i}\int_a^{a+1}h^2(x)\mathrm{d}x < C < \infty$. This verifies (90).

Using (89) and (90) allows us to evaluate expectation of the second term in (87). Write

$$E_h\left\{n^{-1}\sum_{l=1}^{n}\frac{\varphi_j(X_l)I(X_l \in [a,a+1])[G(X_l)-\eta_l/n]}{G^2(X_l)}\right\} = E_h\left\{E_h\left\{n^{-1}\sum_{l=1}^{n}\frac{\varphi_j(X_l)I(X_l \in [a,a+1])[G(X_l)-\eta_l/n]}{G^2(X_l)}|X_l\right\}\right\}$$

$$= n^{-1} E_h \left\{ \frac{\varphi_j(X_l) I(X_l \in [a, a+1])[G(X_l) - 1]}{G^2(X_l)} \right\}$$
$$= n^{-1} \int_a^{a+1} [h(x)(1 - G^{-1}(x))]\varphi_j(x)\mathrm{d}x =: n^{-1}\kappa_j, \tag{91}$$

where due to the Parseval identity and (90) we have the following relation for Fourier coefficients κ_j,

$$\max_{i\in\{1,\ldots,4\}} \sup_{h\in\mathcal{S}_i} \sum_{j=0}^{\infty} \kappa_j^2 = \max_{i\in\{1,\ldots,4\}} \sup_{h\in\mathcal{S}_i} \int_a^{a+1} [h(x)(1-G^{-1}(x))]^2 \mathrm{d}x < C < \infty. \tag{92}$$

Now let us evaluate the expectation of the third term in (87). We begin with the following conditional expectation:

$$E_h\{(G(X_l) - \eta_l/n)^2 | X_l = x\}$$
$$= E_h \left\{ \left[n^{-1} \sum_{s\in\{1,\ldots,n\}\setminus\{l\}} (I(X_s \geq x) - G(x)) + n^{-1}(1 - G(x)) \right]^2 \right\}$$
$$\leq n^{-1} G(x)(1 - G(x)) + n^{-2}(1 - G(x))^2. \tag{93}$$

The familiar Hoeffding inequality (see Petrov 1975) states that if $V_1, V_2, \ldots, V_m$ are independent mean zero random variables with bounded ranges, that is, $\Pr(V_i \in [a_i, b_i]) = 1$, $-\infty < a_i < b_i < \infty$, $i = 1, 2, \ldots, m$ then for each $\epsilon > 0$

$$\Pr\left(\sum_{i=1}^{m} V_i \geq \epsilon \right) \leq \mathrm{e}^{-2\epsilon^2 / \sum_{i=1}^{m}(b_i - a_i)^2}. \tag{94}$$

Note that for Bernoulli random variables we have $b_i - a_i = 1$. Consider some positive ϵ such that $\epsilon - n^{-1} > (1/2)\epsilon[(n-1)/n]^{1/2}$. Then using (94), we get

$$\Pr_h(|G(X_l) - \eta_l/n| > \epsilon | X_l = x)$$
$$= \Pr_h(|n^{-1} \sum_{s\in\{1,\ldots,n\}\setminus\{l\}} [G(x) - I(X_s \geq x)] - n^{-1}(1 - G(x))| > \epsilon)$$
$$\leq \Pr_h(|n^{-1} \sum_{s\neq l, s=1}^{n} [G(x) - I(X_s \geq x)]|$$
$$> (\epsilon - n^{-1}(1 - G(x)))) \leq 2\mathrm{e}^{-n\epsilon^2/2}. \tag{95}$$

In its turn, (95) implies the following probability inequality,

$$\sum_{l=1}^{n} \Pr_h(\{|G(X_l) - \eta_l/n| > \epsilon\}) \leq 2n\mathrm{e}^{-n\epsilon^2/2}. \tag{96}$$

Note that for $x \in [a, a+1]$, we have $G(x) \geq G(a+1) > 0$, and then we have the following chain of relations for events,

$$\left\{\frac{I(X_l \in [a, a+1])}{\eta_l/n} > \frac{2}{G(a+1)}\right\} = \left\{\frac{\eta_l}{n} < \frac{G(a+1)}{2}, X_l \in [a, a+1]\right\}$$
$$\subset \{G(X_l) - \eta_l/n > G(a+1)/2, X_l \in [a, a+1]\}$$
$$\subset \{|G(X_l) - \eta_l/n| > G(a+1)/2\}.$$

This together with (96) yields

$$\max_{i\in\{1,\ldots,4\}} \sup_{h\in\mathcal{S}_i} \sum_{l=1}^{n} \Pr_h\left(\frac{I(X_l \in [a, a+1])}{\eta_l/n} > 2/G(a+1)\right) \leq 2n\mathrm{e}^{-nG^2(a+1)/8}. \tag{97}$$

Now we can evaluate the third term in (87). Note that $\eta_l \geq 1$, $|G(X_l) - \eta_l/n| \leq 1$ and then with the help of (93) and (97) we can write for any h from the four considered function classes,

$$E_h\left\{\left|n^{-1}\sum_{l=1}^{n}\frac{\varphi_j(X_l)I(X_l \in [a, a+1])[G(X_l) - \eta_l/n]^2}{G^2(X_l)\eta_l/n}\right|\right\}$$
$$\leq CE_h\left\{I\left(\frac{I(X_l \in [a, a+1])}{\eta_l/n} \leq \frac{2}{G(a+1)}\right) \times (G(X_l) - \eta_l/n)^2\right\}$$
$$+CnE_h\left\{I\left(\frac{I(X_l \in [a, a+1])}{\eta_l/n} > \frac{2}{G(a+1)}\right)\right\} \leq Cn^{-1}, \tag{98}$$

where all Cs are uniformly bounded over h from the four functional classes.

Using (88), (91) and (98) in the right side of (87), we get

$$|E_h\{\hat{\theta}_j\} - \theta_j| \leq Cn^{-1}. \tag{99}$$

Now we are evaluating the mean squared error of $\hat{\theta}_j$. Using (87), we can write

$$E_h\{(\hat{\theta}_j - \theta_j)^2\} = E_h\left\{\left[n^{-1}\sum_{l=1}^{n}\left(\frac{\varphi_j(X_l)I(X_l \in [a, a+1])}{G(X_l)} - \theta_j\right)\right.\right.$$
$$+n^{-1}\sum_{l=1}^{n}\frac{\varphi_j(X_l)I(X_l \in [a, a+1])[G(X_l) - \eta_l/n]}{G^2(X_l)}$$
$$\left.\left.+n^{-1}\sum_{l=1}^{n}\frac{\varphi_j(X_l)I(X_l \in [a, a+1])[G(X_l) - \eta_l/n]^2}{G^2(X_l)\eta_l/n}\right]^2\right\}. \tag{100}$$

It is convenient to consider the three sums in (100) in turn. Using (88) and $\varphi_j^2(x) = 1 + 2^{-1/2}\varphi_{2j}(x)$, we get

$$E_h\left\{\left[n^{-1}\sum_{l=1}^{n}\left(\frac{\varphi_j(X_l)I(X_l \in [a, a+1])}{G(X_l)} - \theta_j\right)\right]^2\right\}$$
$$= n^{-1}\left[E_h\left\{\left(\frac{\varphi_j(X_l)I(X_l \in [a, a+1])}{G(X_l)}\right)^2\right\} - \theta_j^2\right]$$
$$= n^{-1}\left[\int_a^{a+1} h(x)G^{-1}(x)\mathrm{d}x + \nu_j - \theta_j^2\right],$$
$$\nu_j := 2^{-1/2}\int_a^{a+1} h(x)G^{-1}(x)\varphi_{2j}(x)\mathrm{d}x, \tag{101}$$

where according to the Bessel inequality

$$\max_{i=1,\ldots,5}\sup_{h\in\mathcal{S}_i}\sum_{j=0}^{\infty}\nu_j^2 \le \max_{i=1,\ldots,5}\sup_{h\in\mathcal{S}_i}\int_a^{a+1} h^2(x)G^{-2}(x)\mathrm{d}x < C < \infty. \tag{102}$$

Before proceeding to the analysis of other two sums, let us establish the following useful inequality. For any positive integer k

$$E_h\{[G(X_l) - \eta_l/n]^{2k}\} \le C_k n^{-k}, \;\; C_k < \infty, \tag{103}$$

where C_k depends only on k. Indeed, similarly to (93), we can write

$$E_h\{[G(X_l) - \eta_l/n]^{2k}\}$$
$$= E_h\{E_h\{[n^{-1}\sum_{s\in\{1,\ldots,n\}\setminus\{l\}}(I(X_s \ge X_l) - G(X_l)) + n^{-1}(1 - G(X_l))]^{2k}|X_l\}\}.$$

Then (103) is verified via applying two familiar inequalities (see Petrov 1975): (i) For independent and mean zero random variables $V_1, \ldots, V_m$

$$E\left\{|m^{-1}\sum_{i=1}^{m}V_i|^p\right\} \le C_p^* m^{-p/2-1}\sum_{i=1}^{m}E\{|V_i|^p\}, \;\; p \ge 2, \tag{104}$$

where C_p^* is a finite absolute constant depending only on p; (ii) For any two constants a and b

$$|a + b|^p \le 2^{p-1}(|a|^p + |b|^p), \;\; p \ge 1. \tag{105}$$

Now we can continue evaluation of $E_h\{(\hat{\theta}_j - \theta_j)^2\}$. We explore the second sum in (100),

$$E_h\left\{\left[n^{-1}\sum_{l=1}^{n}\frac{\varphi_j(X_l)I(X_l\in[a,a+1])[G(X_l)-\eta_l/n]}{G^2(X_l)}\right]^2\right\}$$
$$=E_h\left\{n^{-2}\sum_{l,k=1}^{n}\frac{\varphi_j(X_l)\varphi_j(X_k)I(X_l,X_k\in[a,a+1]^2)[G(X_l)-\eta_l/n][G(X_k)-\eta_k/n]}{G^2(X_l)G^2(X_k)}\right\}. \tag{106}$$

To continue the evaluation, we are considering the conditional expectation of a particular factor in (106). For any pair $(x,y)\in[a,a+1]^2$ and $k\neq l$, we can write

$$E_h\{[G(X_l)-\eta_l/n][G(X_k)-\eta_k/n]|X_l=x,X_k=y\}$$
$$=E_h\left\{\left[n^{-1}\sum_{s\in\{1,\dots,n\}\setminus\{k,l\}}(I(X_s\geq x)-G(x))+n^{-1}(1+I(y\geq x)-2G(x))\right]\right.$$
$$\left.\times\left[n^{-1}\sum_{s\in\{1,\dots,n\}\setminus\{k,l\}}(I(X_s\geq y)-G(y))+n^{-1}(1+I(x\geq y)-2G(y))\right]\right\}$$
$$=n^{-2}(1+I(y\geq x)-2G(x))$$
$$\times(1+I(x\geq y)-2G(y))+(n-2)n^{-2}[E_h\{I(X\geq x)I(X\geq y)\}-G(x)G(y)]. \tag{107}$$

The last expectation in (107) can be simplified via using the following equality,

$$E_h\{I(X\geq x)I(X\geq y)\}=\Pr_h(X\geq\max(x,y))=G(\max(x,y)).$$

Using this equality in (107) and then the obtained result in (106) we get with the help of (103)

$$\left|E_h\left\{\left[n^{-1}\sum_{l=1}^{n}\frac{\varphi_j(X_l)I(X_l\in[a,a+1])[G(X_l)-\eta_l/n]}{G^2(X_l)}\right]^2\right\}\right|$$
$$\leq Cn^{-2}+\frac{n-2}{n^4}\sum_{l=1}^{n}\sum_{k\in\{1,\dots,n\}\setminus\{l\}}E_h\left\{\frac{\varphi_j(X_l)\varphi_j(X_k)I((X_l,X_k)\in[a,a+1]^2)}{G^2(X_l)G^2(X_k)}\right.$$
$$\left.\times[G(\max(X_l,X_k))-G(X_l)G(X_k)]\right\}=Cn^{-2}+\frac{(n-1)(n-2)}{n^3}$$
$$\times E_h\left\{\frac{\varphi_j(X_1)\varphi_j(X_2)I((X_1,X_2)\in[a,a+1]^2)[G(\max(X_1,X_2))-G(X_1)G(X_2)]}{G^2(X_1)G^2(X_2)}\right\}$$
$$\leq Cn^{-1}[b'_j+n^{-1}],\ \text{where}\ \max_{i\in\{1,\dots,4\}}\sup_{h\in\mathcal{S}_i}\sum_{j=0}^{\infty}[b'_j]^2<C<\infty, \tag{108}$$

and

$$b'_j := E_h\{\varphi_j(X_1)\varphi_j(X_2)I((X_1, X_2) \in [a, a+1]^2)[G(\max(X_1, X_2)) \\ -G(X_1)G(X_2)]G^{-2}(X_1)G^{-2}(X_2)\}.$$

Note that b'_j are particular Fourier coefficients in the tensor-product cosine basis $\varphi_i(x)\varphi_j(y)$ on $[a, a+1]^2$ and, therefore, according to the Bessel inequality and (90) we have $\sum_{j=0}^{\infty}(b'_j)^2 \le C[\int_a^{a+1} h^2(x)\mathrm{d}x]^{1/2}$, and this verifies the last inequality in (108). Finally, inequality (103) allows us to conclude that the expectation of the squared third term in (100) is at most Cn^{-2}. Combining the obtained results in (100), together with the Cauchy inequality $(c+d)^2 \le (1+\rho)c^2 + (1+\rho^{-1})d^2$, which is valid for any $\rho > 0$, we conclude that

$$E_h\{(\hat{\theta}_j - \theta_j)^2\} = n^{-1}\int_a^{a+1} h(x)G^{-1}(x)\mathrm{d}x[1 + o_j^*(1) + o_n^*(1)], \tag{109}$$

where $o_j^*(1)$ and $o_n^*(1)$ are bounded and vanish (as j and n increase) uniformly over $h \in \mathcal{S}_i$, $i = 1, \dots, 4$.

We need one more preliminary result. Let $0 \le J_* < J^*$ be two deterministic integer sequences in n and $\gamma \ge 0$ is a constant. These 3 parameters may depend on information known to the dealer. Introduce a dealer-estimator

$$\tilde{h}(x, \gamma, J_*, J^*) = \sum_{j=0}^{J_*}\hat{\theta}_j\varphi_j(x) + \sum_{j=J_*+1}^{J^*}[1 - (j/J^*)^{\gamma}]\hat{\theta}_j\varphi_j(x), \quad x \in [a, a+1]. \tag{110}$$

Note that our four estimates $\check{h}_i$, considered in (21), are particular cases of the above-defined $\tilde{h}$. This will allow us to study that estimator and then verify (21) as a corollary. Let us calculate the MISE of this estimator. Parseval's identity yields

$$\begin{aligned} &E_h\left\{\int_a^{a+1}(\tilde{h}(x, \gamma, J_*, J^*) - h(x))^2\mathrm{d}x\right\} \\ &= \sum_{j=0}^{J_*}E_h\{(\hat{\theta}_j - \theta_j)^2\} + \sum_{j=J_*+1}^{J^*}E_h\{[(1 - (j/J^*)^{\gamma})(\hat{\theta}_j - \theta_j) \\ &\quad -(j/J^*)^{\gamma}\theta_j]^2\} + \sum_{j>J^*}\theta_j^2 \\ &= \sum_{j=0}^{J_*}E_h\{(\hat{\theta}_j - \theta_j)^2\} + \sum_{j=J_*+1}^{J^*}(1 - (j/J^*)^{\gamma})^2E_h\{(\hat{\theta}_j - \theta_j)^2\} \\ &\quad -2\sum_{j=J_*+1}^{J^*}(1 - (j/J^*)^{\gamma})(j/J^*)^{\gamma}E_h\{\hat{\theta}_j - \theta_j\}\theta_j + \sum_{j=J_*+1}^{J^*}(j/J^*)^{2\gamma}\theta_j^2 + \sum_{j>J^*}\theta_j^2. \end{aligned} \tag{111}$$

Note that we may write

$$\sum_{j=J_*+1}^{J^*}(j/J^*)^{2\gamma}\theta_j^2+\sum_{j>J^*}\theta_j^2$$
$$=(\pi J^*)^{-2\gamma}\sum_{j>J_*}(\pi j)^{2\gamma}\theta_j^2-(J^*)^{-2\gamma}\sum_{j>J^*}[j^{2\gamma}-(J^*)^{2\gamma}]\theta_j^2.$$

Using this, (99) and (109) in (111), we get

$$\Bigg|E_h\left\{\int_a^{a+1}(\tilde{h}(x,\gamma,J_*,J^*)-h(x))^2\mathrm{d}x\right\}$$
$$-\left[\left[n^{-1}(J_*+1)+n^{-1}\sum_{j=J_*+1}^{J^*}(1-(j/J^*)^{\gamma})^2\right]\int_a^{a+1}h(x)G^{-1}(x)\mathrm{d}x\right.$$
$$\left.+(\pi J^*)^{-2\gamma}\sum_{j>J_*}(\pi j)^{2\gamma}\theta_j^2\right]\Bigg|$$
$$\le C^*n^{-1}\left[\sum_{j=0}^{J^*}[o_j^*(1)+o_n^*(1)]+\sum_{j=J_*+1}^{J^*}(j/J^*)^{\gamma}(1-(j/J^*)^{\gamma})|\theta_j|\right.$$
$$\left.+(J^*)^{-2\gamma}\sum_{j>J^*}[j^{2\gamma}-(J^*)^{2\gamma}]\theta_j^2\right]. \tag{112}$$

Note that in (112) the constant C^* and sequences $o_j^*(1)$ and $o_n^*(1)$ do not depend on h. We have established all needed preliminary results to prove Theorem 4.

Now we are considering cases of the four function classes $\mathcal{S}_i$ in turn, and begin with $\mathcal{S}_1=\mathcal{S}_1(\alpha,Q_0,Q,G(a))$. Set $\gamma=\alpha$, $J_*=0$ and $J^*=J_1$ as defined in (17). Note that with these parameters the estimator (110) is the Pinsker's dealer-estimator. First, let us verify the following inequality

$$\sup_{h\in\mathcal{S}_1(\alpha,Q_0,Q,G(a))}\int_a^{a+1}h(x)G^{-1}(x)\mathrm{d}x\le G^{-1}(a)(\mathrm{e}^{Q_0}-1)=:D_1, \tag{113}$$

and note that $J_1=\lfloor(Q/D_1)^{1/(2\alpha+1)}b_n\rfloor$. To verify (113) we use two identities which hold for any $x\ge a$. The former is $d\mathrm{e}^{\int_a^x h(v)\mathrm{d}v}/\mathrm{d}x=h(x)\mathrm{e}^{\int_a^x h(v)\mathrm{d}v}$ and the latter is $G^{-1}(x)=G^{-1}(a)\mathrm{e}^{\int_a^x h(v)\mathrm{d}v}$. Using them we can write

$$\int_a^{a+1}h(x)G^{-1}(x)\mathrm{d}x$$
$$=G^{-1}(a)\int_a^{a+1}h(x)\mathrm{e}^{\int_a^x h(v)\mathrm{d}v}\mathrm{d}x$$

$$= G^{-1}(a)\left[e^{\int_a^{a+1} h(v)dv} - e^{\int_a^a h(v)dv}\right] = G^{-1}(a)\left[e^{\int_a^{a+1} h(v)dv} - 1\right]. \quad (114)$$

For $h \in \mathcal{S}_1(\alpha, Q_0, Q, G(a))$ we have $e^{\int_a^{a+1} h(v)dv} \leq e^{Q_0}$, and this, together with (114), verifies (113).

Using (113) in (112), together with the Cauchy–Schwarz inequality

$$\sum_{j=0}^{\infty} |\theta_j| \leq |\theta_0| + \left[\sum_{j=1}^{\infty} j^{-2\alpha}\right]^{1/2} \left[\sum_{j=1}^{\infty} j^{2\alpha}\theta_j^2\right]^{1/2}, \quad (115)$$

implies (also remember that D_1 is defined in (113) and Cs are generic positive constants)

$$\sup_{h\in\mathcal{S}_1(\alpha,Q_0,Q,G(a))} E_h\left\{\int_a^{a+1} (\tilde{h}(x,\alpha,0,J^*) - h(x))^2\right\}$$
$$\leq \left[n^{-1}D_1\sum_{j=0}^{J^*}(1-(j/J^*)^\alpha)^2 + (\pi J^*)^{-2\alpha}Q\right] + Cn^{-1}J^* o_n(1). \quad (116)$$

Let us simplify the first term in (116),

$$n^{-1}D_1\sum_{j=0}^{J^*}(1-(j/J^*)^\alpha)^2 + (\pi J^*)^{-2\alpha}Q$$
$$= n^{-1}D_1\left[(J^*+1) - 2(J^*)^{-\alpha}\sum_{j=0}^{J^*} j^\alpha + (J^*)^{-2\alpha}\sum_{j=0}^{J^*} j^{2\alpha}\right]$$
$$+ (\pi J^*)^{-2\alpha}Q = n^{-1}D_1 J^*[1 - 2(\alpha+1)^{-1} + (2\alpha+1)^{-1} + o_n(1)]$$
$$+ (\pi J^*)^{-2\alpha}Q$$
$$= n^{-1}D_1 J^* \frac{2\alpha^2}{(2\alpha+1)(\alpha+1)} + \frac{Q}{(\pi J^*)^{2\alpha}} + o_n(1)n^{-2\alpha/(2\alpha+1)}. \quad (117)$$

Now plug in $J^* = J_1$, using the expression for J_1 given below line (113), b_n defined in (16), and remember formula (6) for $P(\alpha, Q)$. We continue (117),

$$n^{-1}D_1\sum_{j=0}^{J^*}(1-(j/J^*)^\alpha)^2 + (\pi J^*)^{-2\alpha}Q$$
$$= n^{-1}D_1(Q/D_1)^{1/(2\alpha+1)}\left[\frac{n(2\alpha+1)(\alpha+1)}{\alpha\pi^{2\alpha}}\right]^{1/(2\alpha+1)}\frac{2\alpha^2}{(2\alpha+1)(\alpha+1)}$$
$$+\pi^{-2\alpha}(Q/D_1)^{-2\alpha/(2\alpha+1)}\left[\frac{n(2\alpha+1)(\alpha+1)}{\alpha\pi^{2\alpha}}\right]^{-2\alpha/(2\alpha+1)} Q + o_n(1)n^{-2\alpha/(2\alpha+1)}$$

$$= (n^{-1}D_1)^{2\alpha/(2\alpha+1)} Q^{1/(2\alpha+1)}[(\alpha+1)(2\alpha+1)\pi]^{-2\alpha/(2\alpha+1)}[2\alpha^2\alpha^{-1/(2\alpha+1)} + \alpha^{2\alpha/(2\alpha+1)}] + o_n(1)n^{-2\alpha/(2\alpha+1)} = P(\alpha, Q)(D_1 n^{-1})^{2\alpha/(2\alpha+1)}(1+o_n(1)). \tag{118}$$

Using (118) in (116) verifies Theorem 5 for the global Sobolev class $\mathcal{S}_1$.

Now we are considering class $\mathcal{S}_2 = \mathcal{S}_2(\alpha, c_1, c_2, Q, G(a))$. Consider a particular $h \in \mathcal{S}_2$ and set $\theta := \int_a^{a+1} h(x)\mathrm{d}x$, $\gamma = \alpha$, $J_* := 0$, $J_\theta := \lfloor [(Q - c_1\theta^2)/(c_2 d_\theta)]^{1/(2\alpha+1)} b_n \rfloor$ where $d_\theta := G^{-1}(a)(\mathrm{e}^\theta - 1)$. Note that J_θ and d_θ depend on $h(x)$, $x \in [a, a+1]$ only via θ, and this is what the subscript stresses. Repeating steps leading to (116), we get

$$E_h\left\{\int_a^{a+1} (\tilde{h}(x, \alpha, 0, J_\theta) - h(x))^2 \mathrm{d}x\right\} \le n^{-1} d_\theta \sum_{j=0}^{J_\theta} (1 - (j/J_\theta)^\alpha)^2 + (\pi J_\theta)^{-2\alpha} \sum_{j > J_\theta} (\pi j)^{2\alpha}\theta_j^2 + o_n(1)n^{-2\alpha/(2\alpha+1)}. \tag{119}$$

Note that $o_n(1) \to 0$ as $n \to \infty$ uniformly over $\theta \in (0, \sqrt{Q/c_1})$. For $h \in \mathcal{S}_2$ we have $\sum_{j>J_\theta}(c_1 + c_2(\pi j)^{2\alpha})\theta_j^2 \le Q - c_1\theta^2$, and this implies the inequality $\sum_{j>J_\theta}(\pi j)^{2\alpha}\theta_j^2 \le (Q - c_1\theta^2)/c_2$. Using this we can continue (119)

$$E_h\left\{\int_a^{a+1} (\tilde{h}(x, \alpha, 0, J_\theta) - h(x))^2 \mathrm{d}x\right\} \le n^{-1} d_\theta \sum_{j=0}^{J_\theta} (1 - (j/J_\theta)^\alpha)^2 + (\pi J_\theta)^{-2\alpha}[(Q - c_1\theta^2)/c_2] + o_n(1)n^{-2\alpha/(2\alpha+1)}. \tag{120}$$

To simplify the right side of (120) we are using (118). Note that relation (118) holds for an arbitrary pair (D_1, Q) as long as D_1 and Q are positive constants, and here we will use (118) with $(D_1, Q) = (d_\theta, (Q - c_1\theta^2)/c_2)$. We are continuing evaluation of (120),

$$\begin{aligned} &E_h\left\{\int_a^{a+1} (\tilde{h}(x, \alpha, 0, J_\theta) - h(x))^2 \mathrm{d}x\right\} \\ &\le P(\alpha, (Q - c_1\theta^2)/c_2)(d_\theta n^{-1})^{2\alpha/(2\alpha+1)}(1 + o_n(1)) \\ &= P(\alpha, 1/c_2)(G(a))^{-2\alpha/(2\alpha+1)}[(Q - c_1\theta^2)^{1/(2\alpha+1)}(\mathrm{e}^\theta - 1)^{2\alpha/(2\alpha+1)}] n^{-2\alpha/(2\alpha+1)}(1 + o_n(1)). \end{aligned} \tag{121}$$

In the right side of (121), the only factor that depends on θ (that is on h) is the one in the square brackets. Let us show that this factor takes on its maximum when $\theta = Q_0^*$, that is,

$$[(Q-c_1\theta^2)^{1/(2\alpha+1)}(e^\theta-1)^{2\alpha/(2\alpha+1)}]$$
$$\leq [(Q-c_1(Q_0^*)^2)^{1/(2\alpha+1)}(e^{Q_0^*}-1)^{2\alpha/(2\alpha+1)}],\ \ \theta\geq 0, \tag{122}$$

where Q_0^*, defined in Sect. 2, is the positive solution of $c_1(z^2-\alpha^{-1}z+\alpha^{-1}ze^z)=Q$. To check (122), set $\psi(z):=(Q-c_1z^2)(e^z-1)^{2\alpha}$, $z\geq 0$. Then, the derivative of $\psi(z)$ is $\psi'(z)=-2c_1z(e^z-1)^{2\alpha}+(Q-c_1z^2)2\alpha(e^z-1)^{2\alpha-1}$. We equate the derivative to zero and get equation $g(z)=Q$, $z\geq 0$, whose roots are extreme points of $\psi(z)$, where $g(z):=c_1(z^2-\alpha^{-1}z+\alpha^{-1}ze^z)$. As we can see, $g(z)$ is increasing for nonnegative z and this yields that the extreme point is unique. We have the extreme point as Q_0^*, and furthermore this is the point of maximum of $\psi(z)$. This verifies (122).

Using (122), we set $\theta=Q_0^*$, note that this particular θ does not depend on an underlying $h\in\mathcal{S}_2$, and conclude with the help of (121) that

$$\sup_{h\in\mathcal{S}_2} E_h\left\{\int_a^{a+1}(\tilde{h}(x,\alpha,0,J_{Q_0^*})-h(x))^2dx\right\}$$
$$\leq P(\alpha,(Q-c_1(Q_0^*)^2)/c_2)(d_{Q_0^*}n^{-1})^{2\alpha/(2\alpha+1)}(1+o_n(1)). \tag{123}$$

This result proves (21) for the class $\mathcal{S}_2$.

Estimate $\tilde{h}(x,\alpha,J_*,J^*)$, with $J_*=0$ and $J^*=J_3$ attains the right side of the lower bound of Theorem 3. This assertion is verified identically to (113)–(118). This proves (21) for the Golubev's class S_3.

For the weakly restricted local class S_4, we use estimate $\tilde{h}(x,\alpha,J_*,J^*)$ with $J_*=M_n-1$ and $J^*=J_3$. The MISE of this estimate attains the lower bound of Theorem 4, and this again is verified following (113)–(118). Relation (21) is verified for all four function classes.

Now let us verify (23) given (22). The MISE of a Pinsker's dealer-estimator is given in (111) where we should set $J_*=0$ and $J^*:=J_n:=J(n)$. Remember that J_n, used by a dealer-estimator, may depend on the class $\mathcal{S}_4$. There are five sums in (111). The first one is of order n^{-1} according to (109). Further, only the third sum may take on negative values, and according to (99) and relation $\sup_{h\in\mathcal{S}_4}\sum_{j=0}^\infty|\theta_j|\leq \sup_{h\in\mathcal{S}_4}\sum_{j=0}^{M_n-1}|\theta_j|+\sup_{h\in\mathcal{S}_4}\sum_{j=M_n}^\infty|\theta_j|<\infty$, the third term is of order n^{-1}. As a result, the first and third terms are negligibly small with respect to $n^{-2\alpha/(2\alpha+1)}$. We are left with the analysis of three positive sums in (111). Suppose that (23) is incorrect. Then, there exists $J_n\in\{0,1,\ldots,n\}$ such that

$$\max\left(\sup_{h\in\mathcal{S}_4}\sum_{j=1}^{J_n}[1-(j/J_n)^\alpha]^2E_h\{(\hat{\theta}_j-\theta_j)^2\},\ \sup_{h\in\mathcal{S}_4}\sum_{j=1}^{J_n}(j/J_n)^{2\alpha}\theta_j^2,\ \sup_{h\in\mathcal{S}_4}\sum_{j>J_n}\theta_j^2\right)$$
$$\leq Cn^{-2\alpha/(2\alpha+1)}. \tag{124}$$

The third sum in (124) is decreasing in J_n, and this yields that $J_n>Cn^{1/(2\alpha+1)}$ (remember that Cs denote generic positive constants that may be different even in the same line). Using (124), (109) and (50) we get for the first sum that

$$Cn^{-2\alpha/(2\alpha+1)} \geq \sup_{h\in\mathcal{S}_4}\sum_{j=1}^{J_n}[1-(j/J_n)^{\alpha}]^2 E_h\{(\hat{\theta}_j-\theta_j)^2\} > Cn^{-1}J_n\frac{2\alpha^2}{(2\alpha+1)(\alpha+1)}. \tag{125}$$

This yields that $J_n < Cn^{1/(2\alpha+1)}$, and as a result we conclude that J_n should be of order $n^{1/(2\alpha+1)}$. Then for the second sum in (124) we get $M_n = o_n(1)J_n$ and write for all large n,

$$Cn^{-2\alpha/(2\alpha+1)} \geq \sup_{h\in\mathcal{S}_4}\sum_{j=1}^{J_n}(j/J_n)^{2\alpha}\theta_j^2 > J_n^{-2\alpha}\sum_{j=1}^{M_n-1} j^{2\alpha}\left[\int_a^{a+1} h_0(x)\varphi_j(x)\mathrm{d}x\right]^2. \tag{126}$$

Now note that $J_n^{-2\alpha} \geq Cn^{-2\alpha/(2\alpha+1)}$ and according to (22) $\sum_{j=1}^{M_n-1} j^{2\alpha}[\int_a^{a+1} h_0(x)\varphi_j(x)\mathrm{d}x]^2 \to \infty$ as $n \to \infty$. This contradiction in (126) establishes validity of (23). □

Proof of Theorem 6 Consider the mean squared error of a shrinkage estimator $\lambda_j\hat{\theta}_j$ of Fourier coefficients $\theta_j := \int_a^{a+1} h(x)\varphi_j(x)\mathrm{d}x$ where $\lambda_0 = 1$, $\lambda_j \in [0, 1]$ for $j \geq 1$, $\sum_{j=0}^{J^*}\lambda_j^2 \to \infty$ as $n \to \infty$, and $J^* := \sum_{k=1}^{K_n} L_k$. Note that J^* is of order $n^{1/3}\ln\ln(n)$ due to definition of K_n. Write

$$E_h\{(\lambda_j\hat{\theta}_j-\theta_j)^2\} = \lambda_j^2 E_h\{(\hat{\theta}_j-\theta_j)^2\} + (1-\lambda_j)^2\theta_j^2 - 2(1-\lambda_j)\lambda_j E_h\{\hat{\theta}_j-\theta_j\}\theta_j. \tag{127}$$

Using (99) and (109), we continue (127),

$$E_h\{(\lambda_j\hat{\theta}_j-\theta_j)^2\} = [\lambda_j^2 n^{-1}d^* + (1-\lambda_j)^2\theta_j^2] + n^{-1}\rho_j(h, n, \lambda_j), \tag{128}$$

where $|\rho_j(h, n, \lambda_j)| \leq C^*[\lambda_j^2(o_j^*(1) + o_n^*(1)) + |\theta_j|]$ and $d^* = \int_a^{a+1} h(x)G^{-1}(x)\mathrm{d}x$ (remember that d^* is defined in (26)). Introduce an estimator of $h(x)$ for $x \in [a, a+1]$,

$$\bar{h}(x, \{\lambda_j\}) := \sum_{j=0}^{J^*}\lambda_j\hat{\theta}_j\varphi_j(x). \tag{129}$$

Our next step is to show that the oracle-estimator asymptotically dominates a class of such estimators. Using (128) and the Parseval identity, we can evaluate the MISE of estimator (129),

$$\begin{aligned}
&E_h\left\{\int_a^{a+1}(\bar{h}(x, \{\lambda_j\}) - h(x))^2\mathrm{d}x\right\}\\
&= \sum_{j=0}^{J^*}[n^{-1}d^*\lambda_j^2 + (1-\lambda_j)^2\theta_j^2] + \sum_{j>J^*}\theta_j^2 + n^{-1}\sum_{j=0}^{J^*}\rho_j(h, n, \lambda_j)\\
&\leq \sum_{j=0}^{J^*}[n^{-1}d^*\lambda_j^2 + (1-\lambda_j)^2\theta_j^2] + \sum_{j>J^*}\theta_j^2 + \left[o_n^*(1)n^{-1}\sum_{j=0}^{J^*}\lambda_j^2\right.
\end{aligned}$$

$$+C^* n^{-1} I\left(\sum_{j=0}^{J^*} \lambda_j^2 < \ln\ln(n)\right)\Bigg], \tag{130}$$

where $o_n^*(1) \to 0$ as $n \to \infty$ uniformly over hazard rates h from the four functional classes S_i. Note that the inequality in (130) is due to the upper bound for the third sum in the right part of the equality in the top line of (130). Also, the term with the indicator in (130) is used to make the inequality valid for the case when $\sum_{j=0}^{J^*} \lambda_j^2$ does not increase to infinity as n increases. We do assume that the latter does not occur, but the general inequality (130) holds for all cases and this will allow us to use it later for any sequence λ_j. We will be reminded about this remark later in the proof of Theorem 7.

Remember our notation for blocks, introduced above line (24), and write for the first term in (130),

$$\begin{aligned}
&\sum_{j=0}^{J^*}[n^{-1}d^*\lambda_j^2 + (1-\lambda_j)^2\theta_j^2] = \sum_{k=1}^{K_n}\sum_{j\in B_k}[n^{-1}d^*\lambda_j^2 + (1-\lambda_j)^2\theta_j^2] \\
&\quad = \sum_{k=1}^{K_n} L_k[n^{-1}d^*\mu_k^2 + (1-\mu_k)^2\Theta_k] + \left[n^{-1}d^*\sum_{k=1}^{K_n}\sum_{j\in B_k}(\lambda_j^2 - \mu_k^2)\right. \\
&\qquad \left. + \sum_{k=1}^{K_n}\sum_{j\in B_k}(2-\mu_k-\lambda_j)(\mu_k-\lambda_j)\theta_j^2\right] \\
&\quad =: A + B, \quad \text{ where } \mu_k := \max_{j\in B_k}\lambda_j,
\end{aligned} \tag{131}$$

and remember that $\Theta_k = L_k^{-1}\sum_{\mathrm{J}\in B_k}\theta_j^2$. Let us find lower bounds for the two terms in (131). To evaluate A, we note that the minimum of $\psi(z) := L[n^{-1}d^*z^2 + (1-z)^2\Theta]$ over $z \in [0, 1]$ is attained at $z^* = \Theta/[\Theta + n^{-1}d^*]$, and then $\psi(z^*) = n^{-1}dLz^* = n^{-1}d^*L\Theta/[\Theta + n^{-1}d^*]$. Using this fact, we conclude that

$$A \geq n^{-1}d^*\sum_{k=1}^{K_n} L_k\Theta_k[\Theta_k + n^{-1}d^*]^{-1}. \tag{132}$$

To evaluate term B in (131) from below we note that in B the second sum is nonnegative. To evaluate the first sum, which is nonpositive, let us in addition to that made in the beginning of the proof assumptions about $\{\lambda_j\}$ assume that

$$\mu_{k+1} \leq \min_{j\in B_k}\lambda_j, \quad k = 1, \ldots, K_n - 1. \tag{133}$$

Note that (133), as well as the earlier made assumptions about $\{\lambda_j\}$, holds for the shrinkage coefficients used by the four dealer-estimators. Then, we can write

$$\sum_{k=1}^{K_n} L_k\mu_k^2 \le L_1 + \sum_{k=2}^{K_n} L_k\mu_k^2 \le L_1 + \sum_{k=2}^{K_n}[L_k/L_{k-1}] \sum_{j\in B_{k-1}} \lambda_j^2$$

$$= L_1 + \sum_{k=1}^{K_n-1}[L_{k+1}/L_k] \sum_{j\in B_k} \lambda_j^2$$

$$\le C^* \ln(n) + \frac{[1+1/(\ln(n)\ln(\ln(n)))]^{\lfloor \ln(n)\rfloor+1}+1}{[1+1/(\ln(n)\ln(\ln(n)))]^{\lfloor \ln(n)\rfloor}} \sum_{k=\lfloor \ln(n)\rfloor+2}^{K_n-1} \sum_{j\in B_k} \lambda_j^2$$

$$\le (1+o_n^*(1)) \sum_{j=0}^{J^*} \lambda_j^2 + C^* \ln(n). \tag{134}$$

Using this result, we conclude that given (133) the term B in (131) can be bounded from below as follows:

$$B \ge -|o_n^*(1)| n^{-1} d^* \sum_{j=0}^{J^*} \lambda_j^2 - C^* n^{-1} \ln(n). \tag{135}$$

Using (132) and (135) in (131), and then using the obtained inequality in (130) we conclude that given (133) the following inequality holds:

$$n^{-1} d^* \sum_{k=1}^{K_n} L_k \frac{\Theta_k}{\Theta_k + d^* n^{-1}} \le (1+o_n^*(1)) E_h \left\{ \int_a^{a+1} (\bar{h}(x,\{\lambda_j\}) - h(x))^2 \mathrm{d}x \right\} + C^* n^{-1} \ln(n). \tag{136}$$

Next step is to evaluate the MISE of the oracle-estimator (24). Set $\lambda_j = \Theta_k[\Theta_k + d^* n^{-1}]^{-1}$ for $j \in B_k$, $k \in \{1, 2, \dots, K_n\}$. It is sufficient to consider the case $\sum_{j=0}^{J^*} \lambda_j^2 \to \infty$ as $n \to \infty$ because otherwise the hazard rate on $[a, a+1]$ is parametric and it is estimated by the oracle-estimator with the MISE proportional to n^{-1}. Using (130), together with the remark made below that line, we can write

$$\left| E_h \left\{ \int_a^{a+1} (\tilde{h}^*(x,h) - h(x))^2 \mathrm{d}x \right\} - \left[n^{-1} d^* \sum_{k=1}^{K_n} \frac{L_k \Theta_k}{\Theta_k + d^* n^{-1}} + \sum_{j>J^*} \theta_j^2 \right] \right|$$

$$\le o_n^*(1) n^{-1} d^* \sum_{k=1}^{K_n} L_k \left[\frac{\Theta_k}{\Theta_k + d^* n^{-1}} \right]^2$$

$$+ C n^{-1} I \left(\sum_{k=1}^{K_n} L_k \left[\frac{\Theta_k}{\Theta_k + d^* n^{-1}} \right]^2 < \ln(\ln(n)) \right). \tag{137}$$

Now we note that smoothing coefficients of the four dealer-estimators $\check{h}_i(x)$ satisfy the relation (133) and the assumptions made in the beginning of the proof, we also have $\max_{i\in\{1,2,3\}} J_i < J^*$, and then with the help of (136) and (137) we conclude that for all four considered dealer-estimators $\check{h}_i(x)$ we can write

$$E_h\left\{\int_a^{a+1}(\tilde{h}^*(x,h)-h(x))^2\mathrm{d}x\right\}$$
$$\leq (1+o_n^*(1))\min_{i\in\{1,2,3,4\}} E_h\left\{\int_a^{a+1}(\check{h}_i(x)-h(x))^2\mathrm{d}x\right\}+C^*n^{-1}\ln(n), \quad (138)$$

where $o_n^*(1)\to 0$, $n\to\infty$ uniformly over h from the four classes $\mathcal{S}_i$. □

Proof of Theorem 7 Using the Cauchy inequality, we can bound the estimator's MISE,

$$E_h\left\{\int_a^{a+1}(\hat{h}(x)-h(x))^2\mathrm{d}x\right\}$$
$$\leq E_h\left\{\int_a^{a+1}(\tilde{h}^*(x,h)-h(x))^2\mathrm{d}x\right\}(1+\rho)$$
$$+E_h\left\{\int_a^{a+1}(\tilde{h}^*(x,h)-\hat{h}(x))^2\mathrm{d}x\right\}(1+\rho^{-1}),\ \rho>0. \quad (139)$$

The MISE of oracle-estimator $\tilde{h}^*(x,h)$ is evaluated in Theorem 6, and we are considering the second expectation in (139) using the Parseval identity,

$$E_h\left\{\int_a^{a+1}(\tilde{h}^*(x,h)-\hat{h}(x))^2\mathrm{d}x\right\}$$
$$=E_h\left\{\sum_{k=1}^{K_n}\left[\frac{\Theta_k}{\Theta_k+d^*n^{-1}}-\frac{L_k^{-1}\sum_{j\in B_k}\hat{\theta}_j^2-\hat{d}n^{-1}}{L_k^{-1}\sum_{j\in B_k}\hat{\theta}_j^2}I\left(L_k^{-1}\sum_{j\in B_k}\hat{\theta}_j^2>(\hat{d}\right.\right.\right.$$
$$\left.\left.\left.+1/\ln(n))n^{-1}\right)\right]^2\sum_{j\in B_k}\hat{\theta}_j^2\right\}. \quad (140)$$

Let us consider a particular $k\in\{1,\ldots,K_n\}$ in the sum, set $\hat{\Theta}_k := L_k^{-1}\sum_{j\in B_k}\hat{\theta}_j^2 - d^*n^{-1}$, where d^* is defined in (3.14), and write

$$\left[\frac{\Theta_k}{\Theta_k+d^*n^{-1}}-\frac{L_k^{-1}\sum_{j\in B_k}\hat{\theta}_j^2-\hat{d}n^{-1}}{L_k^{-1}\sum_{j\in B_k}\hat{\theta}_j^2}I\left(L_k^{-1}\sum_{j\in B_k}\hat{\theta}_j^2>(\hat{d}\right.\right.$$
$$\left.\left.+1/\ln(n))n^{-1}\right)\right]^2\sum_{j\in B_k}\hat{\theta}_j^2$$

$$= \frac{n^{-2}L_k[d^*(\Theta_k - \hat{\Theta}_k) + (\hat{d} - d^*)(\Theta_k + d^*n^{-1})]^2}{(\Theta_k + d^*n^{-1})^2(\hat{\Theta}_k + d^*n^{-1})} I(\hat{\Theta}_k$$
$$> (\hat{d} - d^* + 1/\ln(n))n^{-1})$$
$$+ \frac{\Theta_k^2 L_k(\hat{\Theta}_k + d^*n^{-1})}{(\Theta_k + d^*n^{-1})^2} I(\hat{\Theta}_k \le (\hat{d} - d^* + 1/\ln(n))n^{-1}) := A_{k1} + A_{k2}. \tag{141}$$

We begin with the analysis of A_{k1}. Using the Cauchy inequality, we get

$$A_{k1} \le \frac{2n^{-2}L_k[(d^*(\hat{\Theta}_k - \Theta_k))^2 + (\hat{d} - d^*)^2(\Theta_k + d^*n^{-1})^2]}{(\Theta_k + d^*n^{-1})^2(\hat{\Theta}_k + d^*n^{-1})} \times I(\hat{\Theta}_k$$
$$> (\hat{d} - d^* + 1/\ln(n))n^{-1}). \tag{142}$$

Now we need two directly verified inequalities which are established similarly to (93) and (99),

$$E_h\{(\hat{d} - d^*)^2\} \le Cn^{-1}, \;\; E\{(\hat{\Theta}_k - \Theta_k)^2\} \le CL_k^{-1}n^{-1}(\Theta_k + n^{-1}), \;\; k \in \{1, \dots, K_n\}, \tag{143}$$

where here and in what follows generic constants Cs are uniformly bounded for all considered hazard rates h and n. Also note that by its definition $\hat{d} > 0$, and that $[\hat{\Theta}_k + d^*n^{-1}]^{-1} I(\hat{\Theta}_k > (\hat{d} - d^* + 1/\ln(n))n^{-1}) < \ln(n)n$. Using these results, we establish that

$$E\left\{\frac{n^{-2}L_k(d^*)^2(\hat{\Theta}_k - \Theta_k)^2}{(\Theta_k + d^*n^{-1})^2(\hat{\Theta}_k + d^*n^{-1})} I(\hat{\Theta}_k > (\hat{d} - d^* + 1/\ln(n))n^{-1})\right\} \le C\ln(n)n^{-1},$$

and

$$E\left\{\frac{n^{-2}L_k(\hat{d} - d^*)^2}{(\hat{\Theta}_k + d^*n^{-1})} I(\hat{\Theta}_k > (\hat{d} - d^* + 1/\ln(n))n^{-1})\right\} \le C\ln(n)n^{-2}L_k.$$

Using these inequalities in (142), we conclude that for $k \in \{1, \dots, K_n\}$

$$E\{A_{k1}\} \le C\ln(n)n^{-1}. \tag{144}$$

Now let us evaluate $E_h\{A_{k2}\}$. Write

$$E_h\{A_{k2}\} = \frac{L_k\Theta_k^2}{(\Theta_k + d^*n^{-1})^2} E_h\{(\hat{\Theta}_k + d^*n^{-1})I(\hat{\Theta}_k \le (\hat{d} - d^* + 1/\ln(n))n^{-1})\}$$
$$=: \frac{L_k\Theta_k^2}{(\Theta_k + d^*n^{-1})^2} B_k. \tag{145}$$

To evaluate B_k, we can write

$$\begin{aligned}B_k &= E_h\{(\hat{\Theta}_k + d^* n^{-1}) I(\hat{\Theta}_k \le (\hat{d} - d^* + 1/\ln(n)) n^{-1})\}\\ &= E_h\{(\hat{\Theta}_k + d^* n^{-1})[I(|\hat{d} - d^*| < 1/\ln(n)) + I(|\hat{d} - d^*| \ge 1/\ln(n))] I(\hat{\Theta}_k\\ &\le (\hat{d} - d^* + 1/\ln(n)) n^{-1})\}.\end{aligned}$$

Using (143) and the Chebyshev inequality, we can continue

$$\begin{aligned}B_k &\le C n^{-1} E_h\{I(\hat{\Theta}_k \le 2n^{-1}/\ln(n))\} + E_h\{(|\hat{d} - d^*|\\ &\quad +1/\ln(n) + d^*) n^{-1} I(|\hat{d} - d^*| \ge 1/\ln(n))\}\\ &\le C n^{-1} E_h\{I(\hat{\Theta}_k \le 2n^{-1}/\ln(n))\}[I(\Theta_k < 4n^{-1}/\ln(n))\\ &\quad +I(\Theta_k \ge 4n^{-1}/\ln(n))] + C n^{-2} \ln^2(n).\end{aligned}$$

Note that

$$\begin{aligned}I(\Theta_k \ge 4n^{-1}\ln(n)) I(\hat{\Theta}_k \le 2n^{-1}/\ln(n)) &\le I(\Theta_k - \hat{\Theta}_k\\ &\ge \Theta_k/2) I(\Theta_k \ge 4n^{-1}/\ln(n)).\end{aligned}$$

Using this, with the help of (143) and the Chebyshev inequality, we continue evaluation of B_k,

$$\begin{aligned}B_k &\le C n^{-1} I(\Theta_k < 4n^{-1}/\ln(n)) + C n^{-1} L_k^{-1/2} n^{-1/2} (\Theta_k + n^{-1})^{1/2} \Theta_k^{-1} I(\Theta_k\\ &\ge 4n^{-1}/\ln(n)) + C n^{-2} \ln^2(n).\end{aligned}$$

Using the last inequality in (145), we get

$$\begin{aligned}E_h\{A_{k2}\} &\le \frac{C L_k \Theta_k n^{-1}}{\Theta_k + d^* n^{-1}} \Bigg[\frac{\Theta_k}{\Theta_k + d^* n^{-1}} I(\Theta_k < 4n^{-1}/\ln(n))\\ &\quad + \frac{\Theta_k L_k^{-1/2} n^{-1/2} (\Theta_k + n^{-1})^{1/2}}{(\Theta_k + d^* n^{-1}) \Theta_k} I(\Theta_k \ge 4n^{-1}/\ln(n)) + C n^{-1} \ln^2(n) \Bigg]\\ &\le \frac{C L_k \Theta_k n^{-1}}{\Theta_k + d^* n^{-1}} \Big[\ln^{-1}(n) I(\Theta_k < 4n^{-1}/\ln(n))\\ &\quad + L_k^{-1/2} I(\Theta_k > 4n^{-1}/\ln(n)) + C n^{-1} \ln^2(n) \Big].\end{aligned} \tag{146}$$

In the square brackets one term depends on $L_k^{-1/2}$; we know that this sequence vanishes as k increases but we need to understand how small a corresponding sum is. Set $t := \lfloor 3\ln(n)(\ln\ln(n))^2 \rfloor$, note that for all large n we have $\ln(n) < t < K_n$, and write

$$n^{-1} \sum_{k=1}^{K_n} \frac{L_k \Theta_k L_k^{-1/2} I(\Theta_k > 4n^{-1}/\ln(n))}{\Theta_k + d^* n^{-1}}$$

$$\leq n^{-1}\sum_{k=1}^{t} L_k + L_t^{-1/2} n^{-1} \sum_{k=t+1}^{K_n} \frac{L_k \Theta_k I(\Theta_k > 4n^{-1}/\ln(n))}{\Theta_k + d^* n^{-1}}.$$

For the first sum, we have

$$\sum_{k=1}^{t} L_k \leq t + \sum_{k=\lfloor \ln(n) \rfloor +1}^{t} [1 + 1/(\ln(n)\ln(\ln(n)))]^k \leq C\ln(n)[\ln(\ln(n))]e^{3\ln(\ln(n))}$$
$$\leq C\ln^{9/2}(n).$$

Furthermore, $L_t > [1+\ln^{-1}(n)/\ln(\ln(n))]^t > Ce^{3\ln(\ln(n))} = C\ln^3(n)$. We conclude that

$$n^{-1}\sum_{k=1}^{K_n} \frac{L_k \Theta_k L_k^{-1/2} I(\Theta_k > 4n^{-1}/\ln(n))}{\Theta_k + d^* n^{-1}}$$
$$\leq n^{-1} C \ln^{9/2}(n) + C\ln^{-3/2}(n)[n^{-1}\sum_{k=1}^{K_n} \frac{L_k \Theta_k}{\Theta_k + d^* n^{-1}} I(\Theta_k > 4n^{-1}/\ln(n)]. \tag{147}$$

Using (147) and (146), we get

$$E\left\{\sum_{k=1}^{K_n} A_{k2}\right\} \leq C\sum_{k=1}^{K_n} \frac{L_k n^{-1}\Theta_k}{\Theta_k + d^* n^{-1}} \ln^{-1}(n) + Cn^{-1}\ln^{9/2}(n). \tag{148}$$

Now we can return to (139). With the help of (144), (148), (141) and (140), we conclude that

$$E_h\left\{\int_a^{a+1} (\hat{h}(x) - h(x))^2 dx\right\}$$
$$\leq E_h\left\{\int_a^{a+1} (\tilde{h}^*(x,h) - h(x))^2 dx\right\}(1+\rho)$$
$$+\left[C\sum_{k=1}^{K_n} \frac{L_k n^{-1}\Theta_k}{\Theta_k + d^* n^{-1}} \ln^{-1}(n) + Cn^{-1}\ln^{9/2}(n)\right](1+\rho^{-1}). \tag{149}$$

Set $\rho = \ln^{-1/2}(n)$, use (137) and continue (149)

$$E_h\left\{\int_a^{a+1} (\hat{h}(x) - h(x))^2 dx\right\} \leq E_h\left\{\int_a^{a+1} (\tilde{h}^*(x,h) - h(x))^2 dx\right\}[1 + C\ln^{-1/2}(n)]$$
$$+Cn^{-1}\ln^5(n). \tag{150}$$

What we see in (150) is the so-called oracle-inequality which relates the MISEs of the estimator and the oracle-estimator.

Now remember that constants Cs do not depend on h and n, and then Theorem 6, together with (50), proves Theorem 7. $\square$

Acknowledgments The research is supported by the NSF Grant DMS–0906790, NSA Grant H982301 310212 and TAF/CAS Grant 11-2341. Comments of referees are greatly appreciated.

References

Antoniadis, A., Grégoire, G., Nason, G. (1999). Density and hazard rate estimation for right-censored data using wavelet methods. *Journal of the Royal Statistical Society, Series B Statistical Methodology*, *61*, 63–84.

Brunel, E., Comte, F. (2005). Penalized contrast estimation of density and hazard rate with censored data. *Sankhya*, *67*, 441–475.

Comte, F., Gaffas, S., Guilloux, A. (2011). Adaptive estimation of the conditional intensity of marker-dependent counting processes. *Annals of Institute of Henri Poincare, Probability and Statistics*, *47*, 1171–1196.

Cox, D. R., Oakes, D. (1984). *Analysis of survival data*. London: Chapman & Hall.

Dögler, S., Rüschendorf, L. (2002). Adaptive estimation of hazard functions. *Probability and Nathematical Statistics*, *22*, 355–379.

Efromovich, S. (1985). Nonparametric estimation of a density with unknown smoothness. *Theory of Probability and its Applications*, *30*, 557–568.

Efromovich, S. (1989). On sequential nonparametric estimation of a density. *Theory of Probability and its Applications*, *34*, 228–239.

Efromovich, S. (1999). *Nonparametric curve estimation: methods, theory and applications*. New York: Springer.

Efromovich, S., Pinsker, M. (1982). Estimation of a square-integrable probability density of a random variable. *Problems of Information Transmission*, *18*, 19–38.

Golubev, G. K. (1991). LAN in problems of nonparametric estimation of functions and lower bounds for quadratic risks. *Theory of Probability and its Applications*, *36*, 152–157.

Gonzalez-Manteiga, W., Cao, R., Marron, J. S. (1996). Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. *Journal of American Statistical Association*, *91*, 1130–1140.

Huber, C., MacGibbon, B. (2004). Lower bounds for estimating a hazard. Advances in Survival Analysis. Handbook of Statistics (Vol. 23, pp. 209–226). Amsterdam: Elsevier.

Ibragimov, I. A., Khasminskii, R. Z. (1981). *Statistical estimation: asymptotic theory*. New York: Springer.

Jankowski, H., Wellner, J. (2009). Nonparametric estimation of a convex bathtub-shaped hazard function. *Bernoulli*, *15*, 1010–1035.

Johnstone, I. (2011). *Gaussian estimation: sequence and wavelet models, manuscript*. Stanford: University of Stanford.

Kahane, J.-P. (1985). *Some random series of functions*. Cambridge: Cambridge University Press.

Lehmann, E. (1983). *Theory of point estimation*. New York: Springer.

Müller, H., Wang, J. (2007). Density and hazard rate estimation. In F. Ruggeri, R. Kenett, F. W. Faltin (Eds.), *Encyclopedia of statistics in quality and reliability* (pp. 517–522). Chichester: Wiley.

Nielsen, J., Linton, O. (1995). Kernel estimation in a nonparametric marker dependent hazard model. *Annals of Statistics*, *23*, 1735–1748.

Patil, P. (1997). Nonparametric hazard rate estimation by orthogonal wavelet method. *Journal of Statistical Planning and Inference*, *60*, 153–168.

Patil, P. N. (1993). On the least squares cross-validation bandwidth in hazard rate estimation. *Annals of Statistics*, *21*, 1792–1810.

Petrov, V. (1975). *Sums of independent random variables*. New York: Springer.

Pinsker, M. S. (1980). Optimal filtering a square integrable signal in Gaussian white noise. *Problems of Information Transmission*, *16*, 52–68.

Prakasa Rao, B. L. S. (1983). *Nonparametric function esimation*. New York: Academic Press.

Rice, J., Rosenblatt, M. (1976). Estimation of the log of survivor function and hazard function. *Sankhya Series A*, *38*, 60–78.

Silverman, B. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Spierdijk, L. (2008). Nonparametric conditional hazard rate estimation: a local linear approach. *Computational Statistics and Data Analysis*, *52*, 2419–2434.

Wang, J.-L. (2005). Smoothing hazard rate. In P. Armitage, T. Colton (Eds.), *Encyclopedia of biostatistics* (2nd ed., Vol. 7, pp. 4486–4497). Chichester: Wiley.

Wasserman, L. (2005). *All of nonparametric statistics*. New York: Springer.

Watson, G., Leadbetter, M. (1964). Hazard rate analysis. I. *Biometrika*, *51*, 175–184.

Wu, S., Wells, M. (2003). Nonparametric estimation of hazard functions by wavelet methods. *Journal of Nonparametric Statistics*, *15*, 187–203.

Zhao, L. (2000). Bayesian aspects of some nonparametric problems. *Annals of Statistics*, *28*, 532–552.