Ann Inst Stat Math (2018) 70:889–917
https://doi.org/10.1007/s10463-017-0617-x

Hazard rate estimation for left truncated and right censored data

Sam Efromovich[1] · **Jufen Chu**[2]

Received: 18 September 2016 / Revised: 15 March 2017 / Published online: 20 September 2017

Abstract Left truncation and right censoring (LTRC) presents a unique challenge for nonparametric estimation of the hazard rate of a continuous lifetime because consistent estimation over the support of the lifetime is impossible. To understand the problem and make practical recommendations, the paper explores how the LTRC affects a minimal (called sharp) constant of a minimax MISE convergence over a fixed interval. The corresponding theory of sharp minimax estimation of the hazard rate is presented, and it shows how right censoring, left truncation and interval of estimation affect the MISE. Obtained results are also new for classical cases of censoring or truncation and some even for the case of direct observations of the lifetime of interest. The theory allows us to propose a relatively simple data-driven estimator for small samples as well as the methodology of choosing an interval of estimation. The estimation methodology is tested numerically and on real data.

Keywords Adaptation · Breast cancer · Longevity · MISE · Nonparametric estimation · Sharp minimax · Survival analysis

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-017-0617-x) contains supplementary material, which is available to authorized users.

✉ Sam Efromovich
efrom@utdallas.edu

Jufen Chu
jufen.chu@novartis.com

1 Department of Mathematical Sciences, The University of Texas at Dallas, Richardson, TX 75080, USA

2 Novartis Oncology, One Health Plaza 315/3420C, East Hanover, NJ 07936-1080, USA

1 Introduction

Consider a continuous nonnegative random variable of interest X^* which can be a lifetime, or a time to event of interest, or an insurance loss, or a commodity price. In all these examples, it is rather typical that a sample from X^* is not observed directly due to the left truncation and right censoring (LTRC). The problem of interest is to estimate, under the mean integrated squared error (MISE) criterion, the hazard rate function $h^{X^*}(x)$ of X^*,

$$h^{X^*}(x) := \frac{f^{X^*}(x)}{G^{X^*}(x)}, \quad G^{X^*}(x) > 0, \quad x \geq 0, \tag{1}$$

where $f^{X^*}(x)$ is the probability density of X^*, $G^{X^*}(x) := \int_x^\infty f^{X^*}(u)\mathrm{d}u = 1 - F^{X^*}(x)$ is the survivor function, and $F^{X^*}(x)$ is the cumulative distribution function of X^*. In what follows, we are using names "hazard rate function," "hazard rate" and "hazard" interchangeably. Recall that the hazard 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. Further, similarly to the density or survivor function, the hazard is a unique characteristic of X^*. See a thorough discussion of the hazard and its pivotal role for the analysis of truncated and censored data in Cox and Oakes (1984), Gill (2006) and Fleming and Harrington (2011).

Let us review known results about nonparametric hazard estimation. On the one hand, the problem appears to be similar (if not almost identical) to the density estimation, and this is also the impression that one can get from the literature. Indeed, the literature traditionally considers estimation over a fixed interval, typically the unit interval [0, 1] given that $G^{X^*}(1) > 0$. This approach is justified by the fact that the hazard is not integrable over its support, to see this note that $G^{X^*}(x) = e^{-\int_0^x h^{X^*}(v)\mathrm{d}v}$. Further, according to (1) the hazard inherits smoothness of the density because the survival function is smoother than the density. In particular, if the density is α-fold differentiable, then the hazard rate is also α-fold differentiable. As a result, estimation procedures suggested for the density can be used for the hazard and, based on a sample of size n, imply the familiar density's rate $n^{-2\alpha/(2\alpha+1)}$ of the MISE convergence. The interested reader can find a number of smart estimation procedures and a discussion of the problem in Silverman (1986), Döhler and Rüschendorf (2002), Wang (2005), Müller and Wang (2007), Jankowski and Wellner (2009), Fleming and Harrington (2011), Bagkavos and Patil (2012), Brody (2012), Klugman et al. (2012), Patil and Bagkavos (2012), Lee and Wang (2013) and Daepp et al. (2015).

On the other hand, the problem of hazard estimation is more complicated than its density counterpart even for the case of a direct sampling from X^*. Indeed, consider estimation of the hazard over an interval $[0, b]$. Efromovich (2016) has established that the sharp constant in the MISE convergence is proportional to $[1 - G^{X^*}(b)]/G^{X^*}(b)$, while for the density estimation the sharp constant is proportional to $F^{X^*}(b)$. As a result, while for the density estimation the constant is bounded, for the hazard rate it is unbounded. This is what makes estimation of the hazard more complicated and, at the same time, more interesting and challenging than its density counterpart.

As we shall see shortly, left truncation (LT) and right censoring (RC) make non-parametric estimation of the hazard even more difficult. Statistical model of the LTRC will be defined shortly in Sect. 2, and here let us just recall that the LT means that only a lifetime exceeding a truncating variable is available, and the RC means that only a lifetime exceeding a censoring random variable is available and otherwise the censoring variable is observed, and this fact is known. Even from these heuristic definitions, it becomes clear that the LT may preclude us from estimation of the hazard (and correspondingly the density) near zero, and the RC may dramatically reduce the maximum value of a possible interval of estimation.

There is a vast statistical literature devoted to the case of the LTRC with a majority of those publications devoted to the RC. Let us mention books by Cox and Oakes (1984), Andersen et al. (1993), Efromovich (1999), Klein and Moeschberger (2003), Fleming and Harrington (2011) and Lee and Wang (2013), as well as recent papers Su and Wang (2012), Huang and Qin (2013), Zhang and Zhou (2013), Bindera and Schumache (2014), Lu and Min (2014), Qian and Betensky (2014), Shi et al. (2015), Wang et al. (2015), Bremhorsta and Lamberta (2016) and Talamakrouni et al. (2016). A number of smart and very elaborate procedures of nonparametric estimation of the hazard over a fixed interval are suggested. For instance, more than 20 years ago Uzunogullari and Wang (1992) proposed a kernel estimator which is still highly regarded in the literature (we will use its oracle version as a benchmark in the numerical study). Antoniadis et al. (1999) proposed a two-step wavelet estimator which estimates the density and then divides it by an estimator of the survival function. Wu and Wells (2003) proposed an estimator based on the idea of a wavelet transform of the Nelson–Aalen cumulative hazard. An adaptive sieved maximum-likelihood estimator is proposed in Döhler and Rüschendorf (2002). For the RC, a thorough discussion of projection estimators, based on the Kaplan–Meyer methodology and a penalized minimum contrast, can be found in Brunel and Comte (2008). Cao et al. (2005) considered the problem of estimation of the relative hazard rate estimation for the LTRC. Review of available R-packages can be found in Hagar and Dukic (2015).

There are several important conclusions from the literature. First, under reasonable smoothness assumptions and considering a sufficiently small fixed interval of estimation, the LTRC does not slow down the rate of the MISE convergence. Second, proposed estimators for small samples are rather complicated, especially if smoothness of the hazard is unknown. Third, it is unknown how the LT, RC and an interval of estimation affect the MISE convergence. Finally, the literature on how to choose a feasible interval of estimation for LTRC data is next to none and this task is typically left to practitioners (see, for instance, a review in Hagar and Dukic 2015). Of course, in practice this interval may be dictated by substantive issues stemming from the motivating problem, but then it is up to the statistician to check if the interval is feasible for data at hand and, if the conclusion is negative, to explain this to practitioners.

All these issues are explored in the paper, and this makes it large; hence, some results are placed in the online Supplementary Materials (SM). The context is as follows. The studied model of the LTRC is introduced in Sect. 2. Asymptotic theory is presented in Sect. 3. Theorems 1 and 2 present sharp minimax lower bounds for two functional classes, while Theorems 3 and 4 are devoted to properties of a data-driven estimator for the cases of independent and dependent truncating and censoring random variables, respectively. The recommended estimator for small samples, its numerical analysis and the methodology of choosing a feasibly large interval of estimation can be found in the SM. Section 4 presents main results of the analysis of the NIH WHEL breast cancer data, based on the developed methodology of estimation for the LTRC data, and more can be found in the SM. The SM also contains colored figures. Section 5 contains some proofs, and all remaining proofs can be found in the SM.

2 The LTRC model

We begin with a classical general model of the LTRC following Gill (2006). There is a hidden sequential sampling from a triplet of nonnegative random variables (T^*, X^*, Z^*) whose joint distribution is unknown. In the triplet T^* is the truncating random variable, X^* is the random variable of interest (lifetime), and Z^* is the censoring random variable. While all realizations of the triplet are not available (they are hidden), some may be observed. Namely, we observe the triplet $(T, Y, \Delta) := (T^*, \min(X^*, Z^*), I(X^* \leq Z^*))$ if $\min(X^*, Z^*) \geq T^*$ and otherwise we do not even know that the hidden realization occurred. The hidden sequential sampling from the triplet (T^*, X^*, Z^*) stops as soon as nth observation is obtained. Note that the hidden mechanism of collecting data can be described via a negative binomial experiment such that the experiment stops as soon as nth "success" occurs, data is collected only when a "success" occurs, there is no information on how many "failures" occurred between "successes", and the probability of "success" is

$$p := \mathbb{P}(T^* \leq \min(X^*, Z^*)) > 0. \tag{2}$$

Let us explain the model via a particular example of the NIH WHEL breast cancer project discussed shortly in Sect. 4. Variable X^* is the time from a breast surgery to the cancer recurrence (relapse). Censoring variable Z^* is the time from surgery until a patient stops participation in the project (due to health issues, death, end of the project, moving from the area, etc.) Truncation variable T^* is the time between surgery and the baseline (beginning) of the project. Now let us comment upon the sequence of events. A potential participant will not be available for the study (truncated) if $\min(X^*, Z^*) \leq T^*$. Indeed, if the relapse occurs before the baseline or the patient is not available/qualified/motivated at the baseline, no information about that patient is available in the data (it will be hidden). Note that in the project the truncation may occur after or before censoring. This is what makes the example so interesting.

In what follows, the asymptotic sharp minimax theory will be developed for the case when the three variables in the hidden triplet (T^*, X^*, Z^*) are independent and continuous. At the same time, it will be shown that if the lifetime of interest X^* is independent from the pair (T^*, Z^*), where T^* and Z^* may be dependent and have a mixed joint distribution, then the proposed estimator is still rate minimax. Example of the last setting is when $Z^* := T^* + R^*$, $\mathbb{P}(R^* \geq 0) = 1$ and R^* may have a mixed distribution. More examples and a thorough discussion of models can be found in Klein and Moeschberger (2003), Gill (2006) and Brody (2012).

3 Asymptotic theory

What is known in the hazard estimation literature about the MISE convergence for the LTRC? Surprisingly, not much. First, it is known that neither the LTRC nor a "reasonable" interval of estimation affects the rate of a minimax MISE convergence. Interestingly, this conclusion is based on results established for the density rather than the hazard. Hence, to understand how the LTRC and an interval of estimation $[a, a+b]$ may affect the MISE, a sharp constant in the MISE convergence should be studied. In its turn, this requires to develop a sharp lower bound for the MISE. Second, for the case of direct observations Efromovich (2016) established that the shape of an estimated hazard does not affect the MISE convergence, and this is why in that article both global and local hazard classes have been studied. As we shall see shortly, this is no longer the case for the LTRC. Hence, a local minimax approach should be used to understand how a hazard affects the MISE convergence.

Now we are in a position to explain the methodology of developing the asymptotic minimax theory of hazard estimation based on LTRC data. First, let us formulate a minimax approach based on the game theory considering a play between the dealer and nature [it is shown in Efromovich (2016) that this approach is instrumental in obtaining a sharp constant for the case of direct observations, and as we shall see shortly it becomes even more pivotal when we are dealing with the LTRC]. Under the approach, the dealer and nature are participating in a minimax game whose rules are defined by a class $\mathcal{S}$ of hazard functions, a chosen risk, and a generator of LTRC observations. The rules are known to the dealer and nature. Nature begins the game by choosing a hazard from the class $\mathcal{S}$ whose estimation maximizes the risk, generates a LTRC sample and then presents it to the dealer. The dealer proposes an estimator which minimizes the MISE and then that minimum defines the dealer's minimax lower bound.

Any dealer's lower bound is of a theoretical interest, but if it can be attained by an estimator (statistic) then the lower bound becomes of a practical interest and that estimator is called efficient. In what follows, we first propose a sharp dealer's lower minimax bound and then introduce an efficient estimator that attains the dealer's bound.

In what follows, in addition to notation of the previous sections, $(0, c_h)$ denotes the largest interval for which we estimate the hazard. An interval $[a, a+b]$, with $0 < a < a+b < c_h$ and $b > 2[3\ln(\ln(n))]^{-2}$, will be referred to as the interval of interest over which an underlying hazard rate $h(x) := h^{X^*}(x)$ should be estimated based on a LTRC sample of size $n > 20$. For the interval of interest, $\varphi_0(x) := b^{-1/2}$, $\varphi_j(x) := (2/b)^{1/2}\cos(\pi j(x-a)/b)$, $j = 1, 2, \dots$ is the classical cosine basis. Also, in what follows α is a nonnegative integer and $q^{(\alpha)}(x) := d^\alpha q(x)/dx^\alpha$ denotes the αth derivative of $q(x)$.

3.1 Sharp lower bound

To consider a local minimax, introduce an anchor hazard $h_0(x)$, $x \geq 0$ with $G_0(x) := e^{-\int_0^x h_0(v)\mathrm{d}v}$ being the corresponding anchor survival function. Introduce a class of anchor hazards,

$$\mathcal{H} := \left\{ h_0 : \int_0^{a+b} h_0(v)\mathrm{d}v < \infty; \quad \inf_{x\in[a,a+b]} h_0(x) > 0; \right.$$
$$\left. = \int_0^{\infty} h_0(x)\infty; \quad h_0(x) \geq 0, x \geq 0 \right\}. \tag{3}$$

Note that this class includes any hazard such that $G_0(a+b) > 0$ and $h_0(x)$ is separated from zero over the interval of interest. Then for a given anchor $h_0 \in \mathcal{H}$, we can define a local (homogeneous) Sobolev function class of hazards which is created by additive perturbations of the anchor over the interval of interest,

$$\mathcal{S}(\alpha, Q, h_0, a, b, n) := \left\{ h : \; h(x) = h_0(x) + \sum_{j=0}^{\infty} \theta_j \varphi_j(x) I(x \in [a, a+b]), \;\; x \geq 0 \; ; \right. \tag{4}$$

$$\left. \sum_{j=0}^{\infty} (\pi j/b)^{2\alpha} \theta_j^2 \leq bQ < \infty, \quad \sup_{x\in[a,a+b]} \left| \sum_{j=0}^{\infty} \theta_j \varphi_j(x) \right| < 1/\ln(n); \;\; h(x) \geq 0, x \geq 0 \right\}. \tag{5}$$

Line (4) implies that considered hazards are additive perturbations of the anchor h_0 over the interval of interest $[a, a+b]$. The first inequality in (5) implies that additive perturbations are from a Sobolev ellipsoid, and the second inequality shows that the magnitude of perturbations decreases as the sample size n increases. Note that for all sufficiently large n such that $1/\ln(n) < \inf_{x\in[a,a+b]} h_0(x)$ we have $\inf_{x\in[a,a+b]} h(x) > 0$, and hence, the last relation in (5) holds. Let us also note that: (i) (3) and (5) yield that for all considered hazards we have $G^{X^*}(a+b) > 0$; (ii) the perturbations take both positive and negative values; (iii) all functions from the class (4) are nonnegative and $\int_0^{\infty} h(x)\mathrm{d}x = \infty$. Hence, these functions are bona fide hazard rates.

In the minimax game, considered in the next theorem, the dealer chooses: an anchor h_0 from the class $\mathcal{H}$ defined in (3), parameters (α, Q, a, b) of the local Sobolev hazard class (4) and the LTRC mechanism. This information is provided to the nature that chooses the least favorable for estimation hazard and then generates a LTRC sample of size n which is used by the dealer for estimating the least favorable hazard.

Theorem 1 *Let us assume that nonnegative continuous random variables T^*, X^* and Z^* are mutually independent, $h_0 \in \mathcal{H}$, the interval of interest $[a, a+b]$ is such that*

$$\min_{x\in[a,a+b]} G^{Z^*}(x) F^{T^*}(x) G_0^{X^*}(x) > 0, \tag{6}$$

and an underlying hazard $h(x) := h^{X^*}(x)$ is from the class (4). Then the following dealer's minimax lower bound holds,

$$\inf_{\check{h}^*} \sup_{h \in \mathcal{S}(\alpha, Q, h_0, a, b, n)} \mathbb{E}_h \left\{ \int_a^{a+b} (\check{h}^*(x) - h(x))^2 \mathrm{d}x \right\}$$
$$\geq P(\alpha, Q, b) \left[\left(b^{-1} \int_a^{a+b} h_0(v) g_0^{-1}(v) \mathrm{d}v \right)^{2\alpha/(2\alpha+1)} \right] n^{-2\alpha/(2\alpha+1)} (1 + o_n(1)), \tag{7}$$

where

$$P(\alpha, Q, b) := Q^{1/(2\alpha+1)} b (2\alpha + 1)^{1/(2\alpha+1)} \left[\frac{\alpha}{\pi(\alpha + 1)} \right]^{2\alpha/(2\alpha+1)}, \tag{8}$$
$$g_0(v) := p_0^{-1} G^{Z^*}(v) F^{T^*}(v) G_0^{X^*}(v), \tag{9}$$
$$p_0 := \mathbb{P}(T^* \leq \min(X^*, Z^*) | h^{X^*} = h_0), \tag{10}$$

the infimum in the left side of (7) *is taken over all possible dealer estimators* $\check{h}^*$ *based on sample* $\{(T_1, Y_1, \Delta_1), \ldots, (T_n, Y_n, \Delta_n)\}$ *as well as on* F^{T^*}*,* G^{Z^*} *and parameters* (α, Q, h_0, a, b) *defining the underlying local Sobolev class. Furthermore, the lower bound* (7) *is attainable by a dealer estimator, and hence, it is sharp.*

Let us make several comments about the result. (i) Factor $P(\alpha, Q, 1)$ is present in the classical minimax lower bound for the density, supported on [0, 1], and direct data (see Efromovich and Pinsker 1982). (ii) In (7) the factor in the square brackets shows how an underlying hazard rate and the LTRC affect the MISE, and following Efromovich (1999) the functional

$$d := b^{-1} \int_a^{a+b} h_0^{X^*}(v) g_0^{-1}(v) \mathrm{d}v \tag{11}$$

may be referred to as the coefficient of difficulty. As we shall see shortly, variance of an efficient estimator of θ_j is equal to $dn^{-1}[1 + o_n(1) + o_j(1)]$, and this sheds light on the coefficient of difficulty. (iii) Rate $n^{-2\alpha/(2\alpha+1)}$, known for the case of direct observations, is preserved. Furthermore, using a sequence (4) of shrinking local Sobolev classes in the minimax does not affect the rate of the MISE convergence. If following Efromovich (2016) we consider a larger class by replacing in the second inequality of (5) the upper bound $1/\ln(n)$ by a positive constant ρ, then in (7) a larger constant, depending on ρ, will appear. (iv) Assumption (6) describes restrictions on supports of the three underlying variables that in general are not necessarily the same. Let us comment on the restrictions. It is well understood, without invoking any theory, that if $F^{T^*}(x_1) = 0$ and $G^{X^*}(x_1) > 0$ or $G^{Z^*}(x_2) = 0$ and $G^{X^*}(x_2) > 0$ then no consistent estimation of $h^{X^*}(x)$ (or $F^{X^*}(x)$) is possible in the former case due to truncation of all observations smaller than x_1 and in the latter case due to censoring all observations larger than x_2. Further, a known result Efromovich (2001) on estimation of the density f^{X^*} for right censored data shows that the corresponding sharp constant is proportional to $\int_a^{a+b} f^{X^*}(x)[G^{Z^*}(x)]^{-1}\mathrm{d}x$. This tells us that if $G^{Z^*}(a+b) > 0$ then the density estimation is possible. Assumption (6), together with the constant (11), allows us to shed a new light on the effect of censoring on estimation of the hazard rate. No longer the assumption $G^{Z^*}(a+b) > 0$ alone guarantees a consistent estimation of the hazard rate over $[a, a+b]$, it should be complemented by $G^{X^*}(a+b) > 0$. This is, of course, the specific and complexity of the considered problem of estimation of the hazard rate, and more discussion can be found in the Supplementary Materials.

Our final comment about the result is as follows. If interval $[a, a+b]$ is fixed and there is no interest in assessing the hazard beyond this interval, then using of class (4) is appropriate (as well as standard in the literature). On the other hand, if this interval is one of possible intervals of interest, then some important issues with the above-introduced minimax arise. First, it follows from (4) that a function from the class may not be smooth because no assumption about smoothness of the anchor hazard $h_0(x)$ is made [check (3)]. Second, even if the anchor $h_0(x)$ is continuous, a perturbation (4) can make $h(x)$ discontinuous at the boundary points a and $a+b$. Third, a perturbation can change the distribution of X^* beyond the interval of estimation. Of course, a traditional answer is that we are interested in estimation only over $[a, a+b]$ and for the dealer smoothness of an underlying anchor is irrelevant because the dealer knows the anchor. On the other hand, eventually we will use the dealer's lower bound for the analysis of estimates. Hence, is it possible that the above-mentioned issues affect the lower bound?

This is a licit question, and let us explore it. First, introduce a class of α-fold differentiable on $(0, c_h)$ (recall that this is the largest studied interval of estimation) anchor hazards,

$$\mathcal{H}^*(\alpha, c_h, a, b) := \Big\{ h_0 : \int_0^{c_h} h_0(x)\mathrm{d}x < \infty, \int_0^{c_h} \left[h_0^{(\alpha)}(x)\right]^2 \mathrm{d}x < \infty, \\ \min_{x\in[a,a+b]} h_0(x) > 0, \quad h_0(x) \geq 0, x \geq 0 \Big\}. \tag{12}$$

For an anchor $h_0 \in \mathcal{H}^*(\alpha, c_h, a, b)$, introduce a class of α-fold differentiable hazard functions which is obtained by additive perturbations of the anchor on the interval of interest $[a, a+b]$,

$$\mathcal{S}^*(\alpha, Q, h_0, c_h, a, b, n) := \{h : h(x) = h_0(x) + q(x), \tag{13}$$

$$q(x) = 0 \text{ for } x \notin (a + [3\ln(\ln(n))]^{-2}, a+b \\ -[3\ln(\ln(n))]^{-2}) \text{ and } \int_a^{a+b} q(x)\mathrm{d}x = 0, \tag{14}$$

$$\max_{x\in[a,a+b]} |q(x)| < 1/\ln(n), \tag{15}$$

$$\int_a^{a+b} [q^{(\alpha)}(x)]^2 \mathrm{d}x \leq bQ < \infty \Big\}. \tag{16}$$

Let us comment on the class. Line (13) tells us that only additive perturbations of the anchor h_0 are allowed. Line (14) yields that the hazard is equal to the anchor beyond $(a, a+b)$, and no change in the cumulative hazard (and therefore in the distribution of X^*) occurs beyond the interval of interest. Line (15) implies the local nature of the class, and line (16) means that all hazards from the class are α-fold differentiable over $(0, c_h)$. As a result, this class takes care about the above-raised issues.

Now let us comment on the sequence of events in the new minimax game. The dealer chooses parameters of the class (12), then chooses an anchor h_0 from this class as well as parameters of the class (13) and then informs nature about the class (13). Additionally, the dealer chooses the LTRC mechanism and informs nature about it. Then nature chooses a hazard, which is the least favorable for estimation, from class (13) and generates a LTRC sample of size n. The dealer uses the sample, together with information provided by nature, to estimate the hazard.

Theorem 2 *Consider a LTRC model where nonnegative continuous random variables T^*, X^* and Z^* are mutually independent, an anchor h_0 is from class (12), and the interval of interest $[a, a+b]$ is such that*

$$\min_{x\in[a,a+b]} G^{Z^*}(x)F^{T^*}(x)G_0^{X^*}(x) > 0. \tag{17}$$

Then

$$\inf_{\check{h}^*} \sup_{h^{X^*}\in\mathcal{S}^*(\alpha,Q,h_0,c_h,a,b,n)} \mathbb{E}_h\left\{\int_a^{a+b}(\check{h}^*(x)-h(x))^2\mathrm{d}x\right\}$$
$$\geq P(\alpha, Q, b)\left[\left(b^{-1}\int_a^{a+b} h_0(v)g_0^{-1}(v)\mathrm{d}v\right)^{2\alpha/(2\alpha+1)}\right]n^{-2\alpha/(2\alpha+1)}(1+o_n(1)), \tag{18}$$

where the infimum in the left side of (18) is taken over all possible dealer's estimates $\check{h}^*$ based on a sample $(T_1, Y_1, \Delta_1), \ldots, (T_n, Y_n, \Delta_n)$, F^{T^*}, G^{Z^*}, and parameters $(\alpha, Q, h_0, c_h, a, b, n)$. Furthermore, the lower bound (18) is attainable by a dealer estimator and hence it is sharp.

Note that the lower bound (18) is the same as in Theorem 1, and hence, neither smoothness of the anchor nor a possible changing by an additive perturbation of the distribution of X^* beyond the interval of interest affects the sharp lower bound. We also conclude that using the classical Sobolev class (4) is a feasible approach for the considered LTRC setting.

3.2 Sharp minimax data-driven estimation

The aim is to propose an estimator, based only on a sample $(T_1, Y_1, \Delta_1), , \ldots, (T_n, Y_n, \Delta_n)$, whose MISE attains the dealer's minimax lower bound (7). This is a challenging problem keeping in mind that the dealer knows the function class (4) and the LTRC mechanism.

To define an estimator, we need several new notations. In what follows, $\lfloor z \rfloor$ denotes the largest integer which is smaller or equal to z. Introduce a sequence of blocks $\{B_k, k = 1, 2, \ldots\}$ that partition nonnegative integers (frequencies of the cosine basis $\{\varphi_j(x), j = 0, 1, \ldots\}$) into nonoverlapping blocks of cardinality (length) L_k. 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 $k > \lfloor \ln(n) \rfloor$, this defines blocks $B_1 := \{0\}$ and $B_k := \{\sum_{s=1}^{k-1} L_s, \ldots, \sum_{s=1}^{k} L_s - 1\}$ for $k = 2, 3, \ldots$ 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 $[n^{1/3} \ln(\ln(n))]^{-2\alpha} = o_n(1) n^{-2\alpha/(2\alpha+1)}$.

The estimator is defined as

$$\hat{h}(x) := \sum_{k=1}^{K_n} \left[1 - \frac{\hat{d} n^{-1}}{L_k^{-1} \sum_{j \in B_k} \hat{\theta}_j^2} \right] I \left(L_k^{-1} \sum_{j \in B_k} \hat{\theta}_j^2 > (\hat{d} + 1/\ln(n)) n^{-1} \right) \times \sum_{j \in B_k} \hat{\theta}_j \varphi_j(x), \quad x \in [a, a+b]. \tag{19}$$

Here

$$\hat{\theta}_j := \sum_{l=1}^{n} \Delta_l \varphi_j(Y_l) \eta_l^{-1} I(Y_l \in [a, a+b]), \tag{20}$$

$$\eta_l := \sum_{s=1}^{n} I(T_s \leq Y_l \leq Y_s), \tag{21}$$

note that $\eta_l \geq 1$ and hence its reciprocal may be used in (20), and

$$\hat{d} := n b^{-1} \sum_{l=1}^{n} \Delta_l \eta_l^{-2} I(Y_l \in [a, a+b]) \tag{22}$$

is the estimate of the coefficient of difficulty (11).

Let us make a comment about observations (T_s, Y_s, Δ_s) used by the estimator. Observations with $Y_s \notin [a, a+b]$ are used only to calculate statistics η_l. Further, we need to calculate η_l only for indexes l such that $Y_l \in [a, a+b]$. As a result, in some cases only values of $Y_s \in [a, a+b]$ are needed to construct the proposed estimator. For instance, suppose that $F^T(a) = 1$. Then we can use $\eta_l' := \sum_{s=1}^{n} [I(Y_l \leq Y_s \leq a+b) + I(Y_s > a+b)]$ in place of η_l. Indeed, $\eta_l' = \eta_l$ if $Y_l \in [a, a+b]$ and it is at least 1 otherwise.

Now we need to add a new reasonable assumption about an anchor h_0 used in the local minimax. The dealer knows the anchor h_0, while an estimator is based only on data and does not know the anchor which is not necessarily even continuous according to (3). To give an estimator a chance to compete with the dealer, let us additionally assume that the anchor is sufficiently smooth on the interval of interest, namely for some positive constant β [compare with the first inequality in (5)]

$$\sum_{j=0}^{\infty}(1+j^{2\alpha+\beta})\left[\int_a^{a+b} h_0(x)\varphi_j(x)\mathrm{d}x\right]^2 < \infty. \tag{23}$$

Theorem 3 *Let the assumption of Theorem 1 and (23) hold. Then estimator (19) is sharp minimax and adapts to parameters of the local Sobolev class and nuisance functions defining the LTRC mechanism, that is,*

$$\sup_{h^{X^*}\in\mathcal{S}(\alpha,Q,h_0,a,b,n)} \mathbb{E}_{h^{X^*}}\left\{\int_a^{a+b}(\hat{h}(x)-h^{X^*}(x))^2\mathrm{d}x\right\}$$
$$\leq P(\alpha,Q,b)\left[\left(b^{-1}\int_a^{a+b} h_0(x)g_0^{-1}(x)\mathrm{d}x\right)^{2\alpha/(2\alpha+1)}\right]n^{-2\alpha/(2\alpha+1)}(1+o_n(1)). \tag{24}$$

Now let us explore robustness of the proposed estimator. It was explained in Sect. 2 that in some practical situations random variables T^* and Z^* may be dependent and/or not have a continuous joint distribution. For instance, this occurs when $Z^* = T^* + R^*$ and R^* is a nonnegative random variable which may have a mixed distribution. No sharp lower bound is known for this case (it is an open problem). At the same time, we can show that the estimator (19) is rate minimax and its MISE attains the rate $n^{-2\alpha/(2\alpha+1)}$ known for the case of direct observations.

Theorem 4 *Let us assume that a continuous random variable X^* is independent of the pair (T^*, Z^*), this pair may have a mixed (continuous and discrete) joint distribution,*

$$\inf_{x\in[a,a+b]} \mathbb{P}(T^* \leq x \leq \min(X^*, Z^*)) > 0, \tag{25}$$

and (23) holds. Then estimator (19) is rate minimax, that is,

$$\sup_{h^{X^*}\in\mathcal{S}(\alpha,Q,h_0,a,b,n)} \mathbb{E}_{h^{X^*}}\left\{\int_a^{a+b}(\hat{h}(x)-h^{X^*}(x))^2\mathrm{d}x\right\} \leq C_* n^{-2\alpha/(2\alpha+1)},\ \ C_* < \infty. \tag{26}$$

We conclude that the proposed estimator is robust.

Modification of the estimator, proposed for small samples, as well as its numerical study, confidence bands and the methodology of choosing a feasibly large interval of estimation can be found in the online Supplementary Materials. Next section explains its use for the analysis of real data.

4 Analysis of real data

The NIH breast cancer project WHEL is described in Rock et al. (2005) and Pierce et al. (2007). The project was designed to address the question of whether high intake of vegetables and fruits, rich in carotene, could reduce breast cancer recurrence. Conclusion of the project was negative, and this created a controversy. Our aim is to revisit the collected data and use the proposed estimator for its analysis. Then, at the end of the section, we also consider another example devoted to longevity at a retirement community. (Recall that colored figures, as well as more discussion and a numerical study, may be found in the Supplementary Materials.)

The NIH project included 3088 women previously treated for breast cancer and who were cancer free at the baseline. Participants were randomly assigned to either a diet intervention, done via intensive consulting, or to a comparison group. This was done from year 1995 to year 2000, and then participants were followed upon through 2006. Plasma carotenoids concentrations were measured at baseline using blood samples (only 3044 participants provided blood samples and therefore are considered). The variable of interest, X^*, is the recurrence-free survival time which is defined as the time from the date of initial breast cancer diagnosis (for all our purposes that date can be considered as the date of surgery) to the date of diagnosis of the cancer relapse. The data is right censored because some participants dropped the study (by any reason apart of the breast cancer relapse) and also due to the end of the project in 2006. The data are also left truncated because at the baseline potential participants (who had surgery in areas of the study) must be alive, cancer free, still live in areas of the study and agree to participate in the study. Note that censoring due to death, or leaving the area, or not agreeing to participate in the study may occur before truncation by the baseline, and hence $\mathbb{P}(T^* < Z^*) < 1$. The primary hypothesis is that the intervention dietary pattern implies a lesser risk of breast cancer recurrence. This hypothesis was tested via comparison of Kaplan–Meier estimates of survivor functions for censored lifetimes (the effect of left truncation was ignored due to randomization used for choosing the two groups and we will comment on this shortly). Analysis of the Kaplan–Meier estimates revealed no difference between the intervention and control groups.

The available data are LTRC, and the proposed estimator allows us to visualize and compare hazard rates for the intervention and control groups. Following the methodology of choosing an interval of estimation, outlined in Section 3 of the online Supplementary Materials, we begin with the analysis of function $1/g_0(x)$ and coefficient of difficulty $d(1, x)$ for intervention and comparison groups.

Their estimates are shown in the top row of Fig. 1 (estimates for the coefficient of difficulty are shown for interval [8, 13] to better exhibit the interesting shapes, and note that colored graphics may be found in the Supplementary Materials). They allow us to suggest the interval of estimation from 1 to 10 years; we will also exhibit results for the larger interval from 1 to 12 years to show the consequences of this decision. Hazard rate estimates and the corresponding confidence bands are exhibited in the middle and bottom rows of Fig. 1. Furthermore, estimates of the cumulative hazard $H(1, x) := \int_1^x h^{X^*}(v)dv$ are shown in Fig. 2. Note that $H(1, x) = \mathbb{E}_{h^{X^*}}\{\Delta g_0^{-1}(Y)I(Y \in [1, x])\}$ and then its natural sample mean estimate is $\hat{H}(1, x) := \sum_{l=1}^{n} \Delta_l \eta_l^{-1} I(Y_l \in [1, x])$; compare with (20). As we see, there is no evidence that the intervention and comparison groups are different. This also points upon good randomization used in the WHEL study.

We conclude that taking into effect the LT and considering hazard rates does not change the NIH conclusion. At the same time, our approach of estimating hazard rates for LTRC data allows us to shed a new light on recurrence-free survival after breast

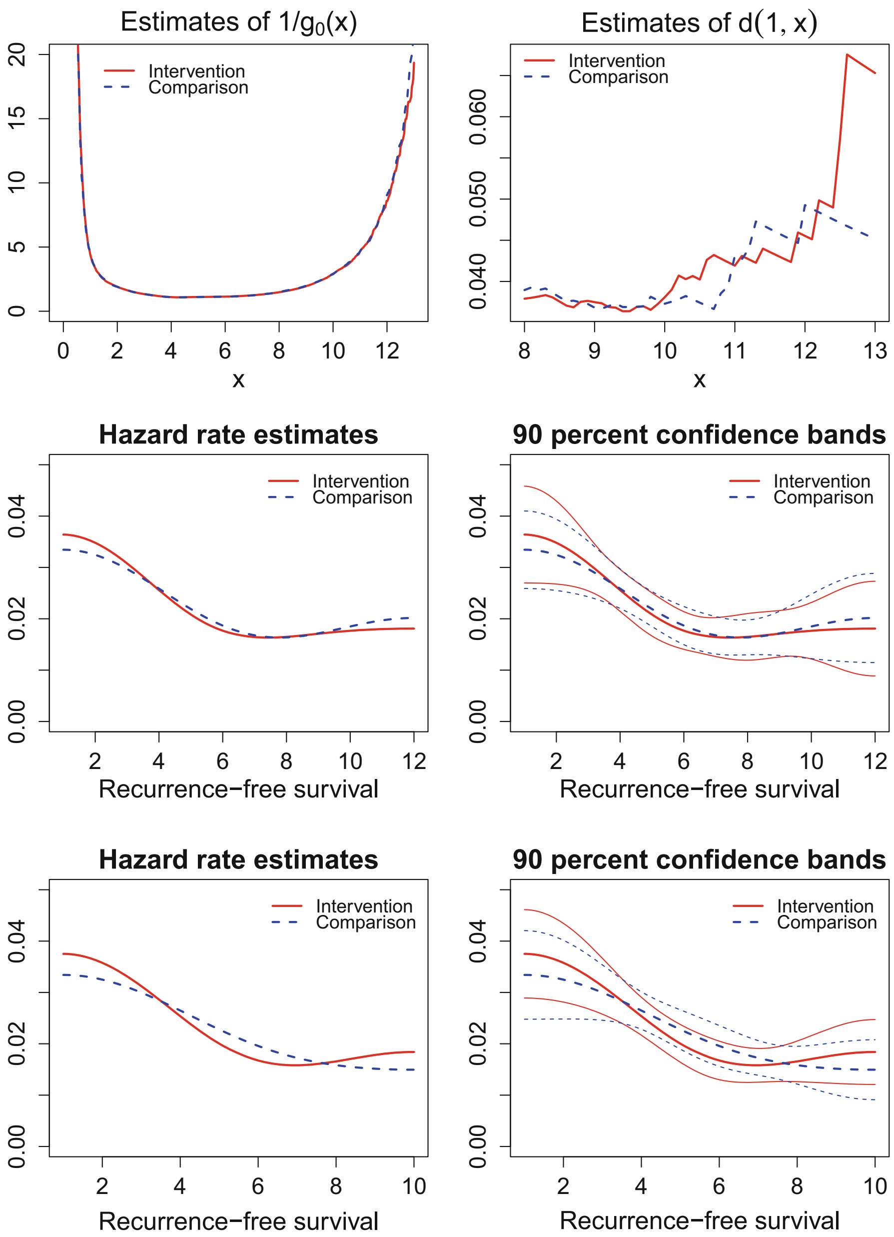

Fig. 1 Estimation for intervention and comparison groups in the WHEL study. The middle and bottom diagrams correspond to estimates for intervals [1, 12] and [1, 10], respectively.

surgery. First, hazard rate estimates allow us to visualize and appreciate how the hazard of cancer recurrence changes in time. Namely, Fig. 1 indicates that the largest hazard rate is during first several years after breast surgery, and then the rate significantly

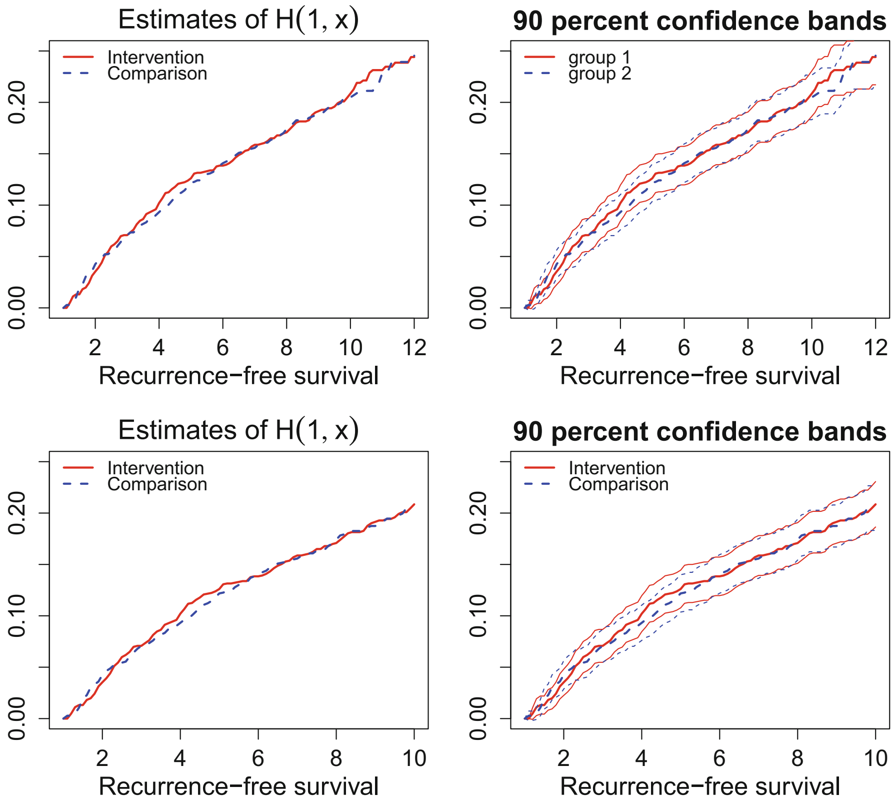

Fig. 2 Estimates of the cumulative hazard $H^{X^*}(1, x) := \int_1^x h^{X^*}(v)dv$ and corresponding 90% confidence bands for intervention and comparison groups in the WHEL study calculated for cases $[a, a+b] = [1, 12]$ (the top diagrams) and $[a, a+b] = [1, 10]$ (the bottom diagrams)

(almost in half) slows down toward the sixth year. After the sixth year, there may be a modest increase in the hazard rate.

Second, it is important to stress that the NIH conclusion does not reject benefits of a carotene-rich diet; it simply tells us that the studied diet intervention (an after-surgery recommendation to enrich diet by carotene) does not imply the desired effect. On the other hand, is there a relationship between levels of carotene, shown in blood tests collected at the baseline, and the recurrence-free survival? In other words, does the level of carotene affects the hazard rate? The proposed estimator for the LTRC can help us to answer this question. For each type of plasma carotenoids, we divide 3044 patients, who provided blood samples at baseline, into two groups using median concentration as the threshold. For example, if plasma alpha-carotene is the variable of interest, then participants whose plasma alpha-carotene was above the median value are classified as group 1 and others belong to group 2. Then for each group the proposed adaptive hazard rate estimator is used to estimate the hazard rate for the time from surgery to relapse.

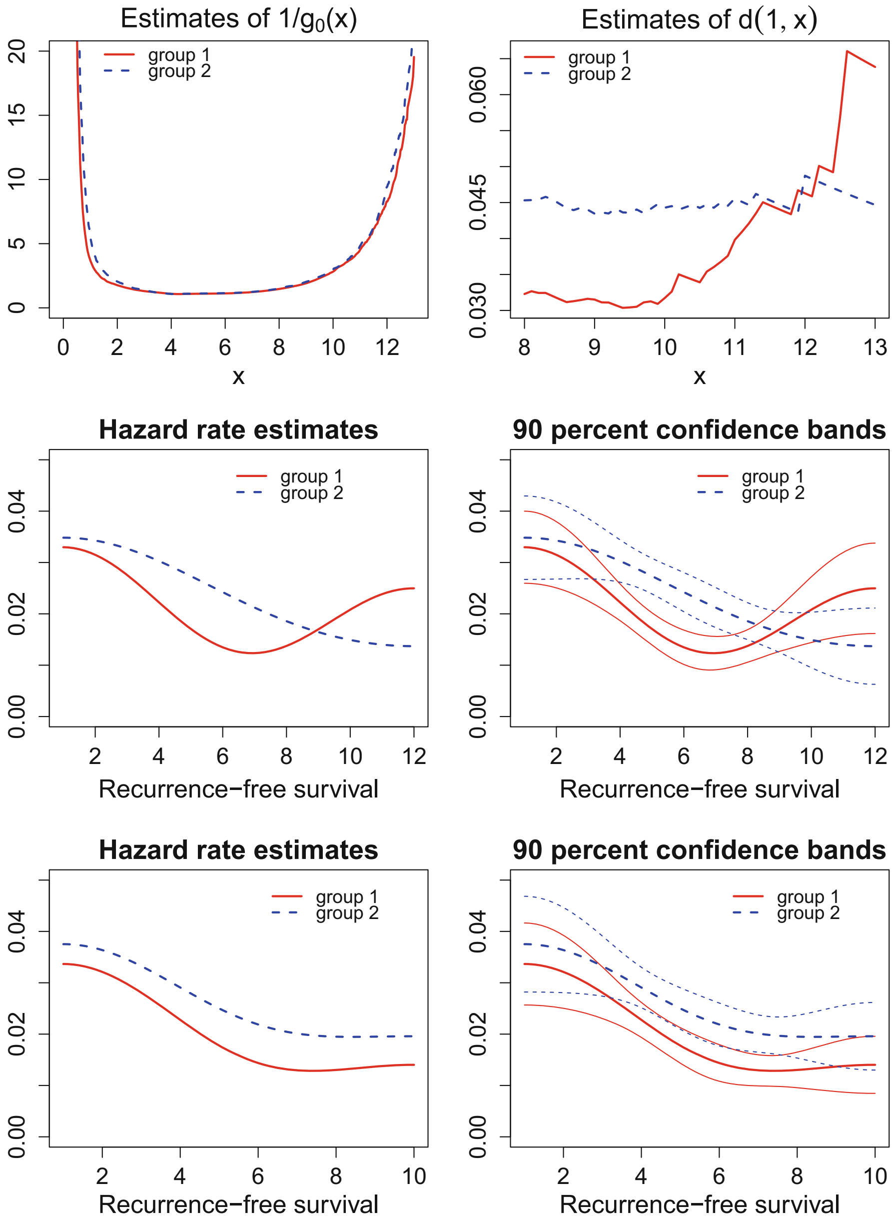

Fig. 3 Effect of plasma alpha-carotene on hazard rates. Group 1 consists of women with levels of plasma alpha-carotene larger than the median level for all participants; all other women belong to group 2. The structure of the figure is identical to Fig. 1

We begin with presenting results for alpha-carotene in Figs. 3 and 4. These figures are similar to the previously discussed Figs. 1 and 2 for the intervention and control groups. What we see is that participants with larger levels of

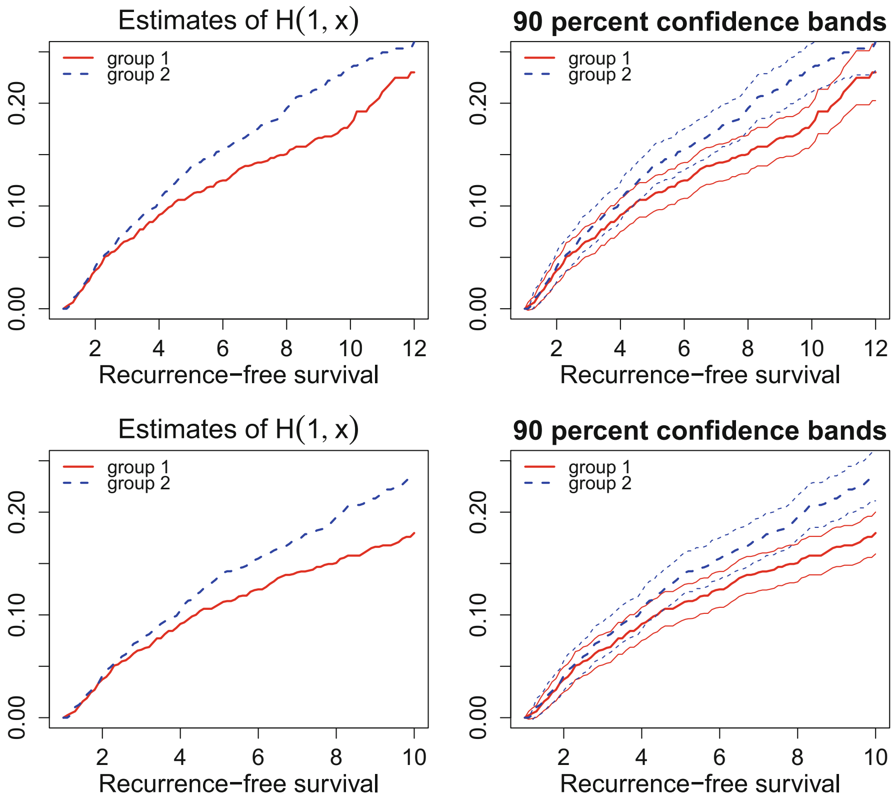

Fig. 4 Effect of plasma alpha-carotene on cumulative hazard. Group 1 consists of women with levels of plasma alpha-carotene larger than the median level for all participants; all other women belong to group 2. The structure of the figure is identical to Fig. 2

plasma alpha-carotene concentration enjoy a lesser risk of the cancer recurrence. A similar conclusion can be made from results for plasma beta-carotene shown in Figures 8 and 9 of the supplementary materials, where also the interested reader can find results for the effects of cryptoxanthin and lycopene. Our dietological conclusion is that a high level of plasma carotenoids, which reflects the history of dietary habits and lifestyle, is a positive factor in reducing the risk of the cancer relapse.

Now let us consider a familiar in the survival analysis literature example of the death times for Channing House retirement community which is traditionally used to illustrate Kaplan–Meier estimator, see a discussion in Klein and Moeschberger (2003). Channing House is a retirement center located in Palo Alto, California, and its distinctive feature is that all residents of the community are covered by a health care program with a zero deductible, that is, no additional financial burden to the residents. The data were collected between the opening of the house in January 1964 and July 1975. In that period of time, 97 men and 365 women passed through the center, and among them 130 women and 46 men died at Channing House. The resident's age at entry to the community as well as the age at leaving the community or death was recorded. Note that when a person joined the community, he or she was alive at that time, and the person's age at death was at least as great as the age at entry to the community and thus was left truncated. If the resident was alive when the study ended or the resident left the community before the study ended, right censoring had occurred. The difference between the survival hazard rates for male and female residents was the primary aim of the study, and the known conclusion is that female residents survive much better than male residents. Using the proposed adaptive estimator for the lifetime of men and women produced estimates exhibited in the left diagram of Figure 14 in the Supplementary Materials, while the ratio of the two estimates is exhibited in the right diagram. Both hazard rate estimates, shown in the left diagram, change relatively slow in the early ages below 960 months, i.e., 80 years old, and then they rise rapidly. Female residents definitely have an edge in terms of longevity, but the estimates indicate that with age the difference between male and female ability to survive at the retirement community vanishes, and furthermore, older male residents may have a slight edge. The relative hazard rate of men with respect to women indicates that the risk of death may be almost twice as large for men than for women for the age group up to 78 years and with the peak around 76 years. The main conclusion is that overall female residents better adjust to living in the retirement community. On the other hand, male residents, who are able to adjust to their new life in the community, may live as long as female residents. Corresponding confidence bands are shown in Figure 15. As we see, there is a relatively small interval (about 20 months) around 930 months where the two confidence bands do not intersect. Further, there is a larger interval in time (about 200 months) where each estimate is not covered by confidence band for another estimate. These findings shed a new light on the classical example.

5 Proofs

Proof of Upper Bounds in Theorems 1 and 2. We need to suggest dealer estimators that attain lower bounds of the theorems. We begin with the dealer estimator for the setting of Theorem 1. Recall that

$$h^{X^*}(x) = h_0(x) + \sum_{j=0}^{\infty} \theta_j \varphi_j(x) I(a \le x \le a+b), \quad \sum_{j=1}^{\infty} (\pi j/b)^{2\alpha} \theta_j^2 \le bQ. \quad (27)$$

Further, d is the coefficient of difficulty defined in (11), and note that it is known to the dealer.

Now let us make the following remark which will allow us to simplify formulas and use the below obtained results in other proofs. The dealer knows the anchor h_0, and hence, we may formally assume in the following formulas that the anchor is zero on the interval $[a, a+b]$. (To make the proof rigorous, one should in all formulas subtract from the proposed estimator $\hat{\theta}_j$ the corresponding jth Fourier coefficient of the anchor.)

Set (we just repeat definitions for the reader convenience)

$$\hat{\theta}_j := \sum_{l=1}^{n} \Delta_l \varphi_j(Y_l)\eta_l^{-1} I(Y_l \in [a, a+b]), \tag{28}$$

where

$$\eta_l := \sum_{s=1}^{n} I(T_s \le Y_l \le Y_s). \tag{29}$$

Note that $\eta_l \ge 1$ and hence its reciprocal may be used in (28). The proposed dealer estimator of the hazard (27) is

$$\check{h}(x) := h_0(x) + \left[\sum_{j=0}^{J_0} \hat{\theta}_j \varphi_j(x) + \sum_{j=J_0+1}^{J} (1 - (j/J)^{\alpha})\hat{\theta}_j \varphi_j(x)\right] I(x \in [a, a+b]), \tag{30}$$

where $J_0 := \lfloor n^{1/(2\alpha+1)}/(\ln(n))^{1/5} \rfloor$, $J := \lfloor n^{1/(2\alpha+1)}[bQ(b/\pi)^{2\alpha}(\alpha+1)(2\alpha+1)/(\alpha d)]^{1/(2\alpha+1} + 1\rfloor$.

Let us show that the dealer estimator is efficient and its MISE attains the lower bound. We begin with establishing several properties of statistic $\hat{\theta}_j$ defined in (28). Set

$$g(v) := \mathbb{P}_{h^{X^*}}(T \le v \le Y) = p^{-1} F^{T^*}(v) G^{X^*}(v) G^{Z^*}(v). \tag{31}$$

Write,

$$\begin{aligned}
\hat{\theta}_j &= n^{-1} \sum_{l=1}^{n} \frac{\Delta_l \varphi_j(Y_l) I(Y_l \in [a, a+b])}{g(Y_l)} \\
&\quad + n^{-1} \sum_{l=1}^{n} \frac{\Delta_l \varphi_j(Y_l) I(Y_l \in [a, a+b])[g(Y_l) - \eta_l/n]}{g^2(Y_l)} \\
&\quad + n^{-1} \sum_{l=1}^{n} \frac{\Delta_l \varphi_j(Y_l) I(Y_l \in [a, a+b])[g(Y_l) - \eta_l/n]^2}{g^2(Y_l)\eta_l/n}.
\end{aligned} \tag{32}$$

Now we are evaluating the expectation of $\hat{\theta}_j$ by considering 3 terms in (32) in turn. For expectation of the first term, we can write,

$$\begin{aligned}
&\mathbb{E}_{h^{X^*}} \left\{ n^{-1} \sum_{l=1}^{n} \frac{\Delta_l \varphi_j(Y_l) I(Y_l \in [a, a+b])}{g(Y_l)} \right\} \\
&= \int_a^{a+b} p^{-1} f^{X^*}(y) G^{Z^*}(y) \varphi_j(y) F^{T^*}(y) g^{-1}(y) \mathrm{d}y
\end{aligned}$$

$$= \int_a^{a+b} h^{X^*}(y)\varphi_j(y)[p^{-1}G^{X^*}(y)G^{Z^*}(y)F^{T^*}(y)g^{-1}(y)]\mathrm{d}y$$
$$= \int_a^{a+b} h^{X^*}(y)\varphi_j(y)\mathrm{d}y = \theta_j. \quad (33)$$

For evaluation of the expectation of the second term in (32), we begin with a conditional expectation,

$$\mathbb{E}_{h^{X^*}}\{(g(Y_l) - \eta_l/n)|Y_l = y\} = \mathbb{E}_{h^{X^*}}\left\{g(y) - n^{-1}\sum_{s\neq l, s=1}^{n} I(T_s \le y \le Y_s) - n^{-1}\right\}$$
$$= g(y) - (n-1)n^{-1}g(y) - n^{-1} = n^{-1}(g(y) - 1). \quad (34)$$

We are using the reciprocal of $g(y)$, hence let us show that

$$\inf_{h^{X^*}\in\mathcal{S}(\alpha,Q,h_0,a,b)} \min_{y\in[a,a+b]} g(y) > 0. \quad (35)$$

Using Cauchy–Schwarz inequality, we can write,

$$G^{X^*}(a+b) = G^{X^*}(a)e^{-\int_a^{a+b} h^{X^*}(x)\mathrm{d}x} \ge G^{X^*}(a)e^{-\left[b\int_a^{a+b}[h^{X^*}(x)]^2\mathrm{d}x\right]^{1/2}}.$$

Remember that for hazard rates from the function class $\mathcal{S}(\alpha, Q, h_0, a, b)$ we have $G^{X^*}(a) > 0$ and $\sup_{h^{X^*}\in\mathcal{S}(\alpha,Q,h_0^{X^*},\beta,a,b)} \int_a^{a+b}[h^{X^*}(x)]^2\mathrm{d}x < C < \infty$. This, together with (6), verifies (35).

Using (34) and (35) allows us to evaluate expectation of the second term in (32). Write,

$$\mathbb{E}_{h^{X^*}}\left\{n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\varphi_j(Y_l)I(Y_l\in[a,a+b])[g(Y_l)-\eta_l/n]}{g^2(Y_l)}\right\}$$
$$= \mathbb{E}_{h^{X^*}}\left\{\mathbb{E}_{h^{X^*}}\left\{n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\varphi_j(Y_l)I(Y_l\in[a,a+b])[g(Y_l)-\eta_l/n]}{g^2(Y_l)}|Y_l\right\}\right\}$$
$$= \mathbb{E}_{h^{X^*}}\left\{\frac{\Delta_l\varphi_j(Y_l)I(Y_l\in[a,a+b])n^{-1}[g(Y_l)-1]}{g^2(Y_l)}\right\}$$
$$= n^{-1}\int_a^{a+b}[h^{X^*}(y)G^{X^*}(y)G^{Z^*}(y)F^{T^*}(y)p^{-1}]\varphi_j(y)g^{-1}(y)(1-g^{-1}(y))\mathrm{d}y$$
$$= n^{-1}\int_a^{a+b}[h^{X^*}(y)(1-g^{-1}(y))]\varphi_j(y)\mathrm{d}y =: n^{-1}\kappa_j. \quad (36)$$

Using the Parseval identity and (35), we get the following relation for Fourier coefficients κ_j,

$$\sup_{h^{X^*}\in\mathcal{S}(\alpha,Q,h_0,a,b)}\sum_{j=0}^{\infty}\kappa_j^2=\sup_{h^{X^*}\in\mathcal{S}(\alpha,Q,h_0,a,b)}\int_a^{a+b}[h^{X^*}(y)(1-g^{-1}(y))]^2\mathrm{d}y<C<\infty. \tag{37}$$

Now let us evaluate the expectation of the third term in (32). Several preliminary calculations are needed. We begin with the following conditional expectation,

$$\begin{aligned}&\mathbb{E}_{h^{X*}}\left\{(g(Y_l)-\eta_l/n)^2|Y_l=y\right\}\\&\quad=\mathbb{E}_{h^{X*}}\left\{[n^{-1}\sum_{s\in\{1,\dots,n\}\setminus\{l\}}(I(T_s\le y\le Y_s)-g(y))\right.\\&\qquad\left.+n^{-1}(1-g(y))]^2\right\}\le n^{-1}g(y)(1-g(y))+n^{-2}(1-g(y))^2. \end{aligned}\tag{38}$$

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

$$\mathbb{P}\left(\sum_{i=1}^{m}V_i\ge\epsilon\right)\le e^{-2\epsilon^2/\sum_{i=1}^{m}(b_i-a_i)^2}. \tag{39}$$

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 (39) we get

$$\begin{aligned}&\mathbb{P}_{h^{X*}}(|g(Y_l)-\eta_l/n|>\epsilon|Y_l=y)\\&\quad=\mathbb{P}_{h^{X*}}\left(\left|n^{-1}\sum_{s\in\{1,\dots,n\}\setminus\{l\}}[g(y)-I(T_s\le y\le Y_s)]-n^{-1}(1-g(y))\right|>\epsilon\right)\\&\quad\le\mathbb{P}_{h^{X*}}\left(\left|n^{-1}\sum_{s\ne l,s=1}^{n}[g(y)-I(T_s\le y\le Y_s)]\right|>(\epsilon-n^{-1}(1-g(y)))\right)\\&\quad\le 2e^{-n\epsilon^2/2}.\end{aligned}\tag{40}$$

In its turn, (40) implies the following probability inequality,

$$\sum_{l=1}^{n}\mathbb{P}_{h^{X*}}(\{|g(Y_l)-\eta_l/n|>\epsilon\})\le 2ne^{-n\epsilon^2/2}. \tag{41}$$

Note that

$$\min_{y\in[a,a+b]}g(y)\ge g(a+b)F^{T^*}(a)/F^{T^*}(a+b)=:g_*>0. \tag{42}$$

Using this result, we can write down the following chain of relations for events,

$$\left\{\frac{I(Y_l \in [a,a+b])}{\eta_l/n} > \frac{2}{g_*}\right\} = \left\{\frac{\eta_l}{n} < \frac{g_*}{2}, Y_l \in [a,a+b]\right\}$$
$$\subset \{g(Y_l) - \eta_l/n > g_*/2, Y_l \in [a,a+b]\} \subset \{|g(Y_l) - \eta_l/n| > g_*/2\},$$

where g_* is defined in (42). The last relation and (41) yield

$$\sup_{h^{X^*} \in \mathcal{S}(\alpha,Q,h_0,a,b)} \sum_{l=1}^{n} \mathbb{P}_{h^{X^*}}\left(\frac{I(Y_l \in [a,a+b])}{\eta_l/n} > 2/g_*\right) \le 2ne^{-ng_*^2/8}. \tag{43}$$

Now we can evaluate the third term in (32). Note that $\eta_l \ge 1$, $|G(X_l) - \eta_l/n| \le 1$ and then with the help of (38) and (43) we can write for any h^{X^*} from $\mathcal{S}(\alpha, Q, h_0, a, b)$,

$$\mathbb{E}_{h^{X^*}}\left\{\left|n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\varphi_j(Y_l)I(Y_l \in [a,a+b])[g(Y_l) - \eta_l/n]^2}{g^2(Y_l)\eta_l/n}\right|\right\}$$
$$\le C^*\mathbb{E}_{h^{X^*}}\left\{I\left(\frac{I(Y_l \in [a,a+b])}{\eta_l/n} \le \frac{2}{g_*}\right)(g(Y_l) - \eta_l/n)^2\right\}$$
$$+ C^* n \mathbb{E}_{h^{X^*}}\left\{I\left(\frac{I(Y_l \in [a,a+b])}{\eta_l/n} > \frac{2}{g_*}\right)\right\} \le C^* n^{-1}, \tag{44}$$

where here and in what follows C^*s are generic and possibly different positive constants that do not depend on considered h^{X^*}.

Using (33), (36) and (44) in the right side of (32), we get

$$|\mathbb{E}_{h^{X^*}}\{\hat{\theta}_j\} - \theta_j| \le C^* n^{-1}. \tag{45}$$

Let us make the following remark about (45). This inequality is sufficient for our purposes, but it can be made tighter by considering four terms in (32) where the third term, similarly to the first two, uses $G(X_l)$ in place of η_l/n. Then we can evaluate the third term similarly to (37), and expectation of the fourth term will be not larger than $C^* n^{-3/2}$. This yields that in (45) a constant C^* can be replaced by some t_j satisfying $\sum_{j=0}^{\infty} t_j^2 < \infty$. The remark is useful in analysis of higher moments.

Now we are evaluating the mean squared error of $\hat{\theta}_j$. Using (32) we can write,

$$\mathbb{E}_{h^{X^*}}\{(\hat{\theta}_j - \theta_j)^2\} = \mathbb{E}_{h^{X^*}}\left\{\left[n^{-1}\sum_{l=1}^{n}\left(\frac{\Delta_l\varphi_j(Y_l)I(Y_l \in [a,a+b])}{g(Y_l)} - \theta_j\right)\right.\right.$$
$$+ n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\varphi_j(Y_l)I(Y_l \in [a,a+b])[g(Y_l) - \eta_l/n]}{g^2(Y_l)}$$
$$\left.\left. + n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\varphi_j(Y_l)I(Y_l \in [a,a+b])[g(Y_l) - \eta_l/n]^2}{g^2(Y_l)\eta_l/n}\right]^2\right\}. \tag{46}$$

It is convenient to consider the three sums in (46) in turn. For $j \geq 1$, using (33) and trigonometric formula $\varphi_j^2(y) = b^{-1} + (2b)^{-1/2}\varphi_{2j}(y)$, we get

$$\begin{aligned} &\mathbb{E}_{h^{X^*}}\left\{\left[n^{-1}\sum_{l=1}^{n}\left(\frac{\Delta_l\varphi_j(Y_l)I(Y_l\in[a,a+b])}{g(Y_l)}-\theta_j\right)\right]^2\right\} \\ &\quad= n^{-1}\left[\mathbb{E}_{h^{X^*}}\left\{\left(\frac{\Delta_l\varphi_j(Y_l)I(Y_l\in[a,a+b])}{g(Y_l)}\right)^2\right\}-\theta_j^2\right] \\ &\quad= n^{-1}\left[b^{-1}\int_a^{a+b} h^{X^*}(y)g^{-1}(y)\mathrm{d}y+\nu_j-\theta_j^2\right], \end{aligned} \tag{47}$$

where $\nu_j := (2b)^{-1/2}\int_a^{a+b} h^{X^*}(y)g^{-1}(y)\varphi_{2j}(y)\mathrm{d}y$ and according to Parseval's identity

$$\begin{aligned} &\sup_{h^{X^*}\in\mathcal{S}(\alpha,Q,h_0,a,b)}\sum_{j=1}^{\infty}\nu_j^2 \leq (2b)^{-1}\sup_{h^{X^*}\in\mathcal{S}(\alpha,Q,h_0,a,b)} \\ &\quad\int_a^{a+b}[h^{X^*}(y)]^2g^{-2}(y)\mathrm{d}y < C^* < \infty. \end{aligned} \tag{48}$$

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

$$\mathbb{E}_{h^{X^*}}\{[g(Y_l)-\eta_l/n]^{2k}\} \leq C_k n^{-k}, \;\; C_k < \infty, \tag{49}$$

where C_k depends only on k. Let us verify (49). Similarly to (38) we can write

$$\begin{aligned} &\mathbb{E}_{h^{X^*}}\{[g(Y_l)-\eta_l/n]^{2k}\} \\ &= \mathbb{E}_{h^{X^*}}\left\{\mathbb{E}_{h^{X^*}}\left\{\left[n^{-1}\sum_{s\in\{1,\dots,n\}\setminus\{l\}}(I(T_s\leq Y_l\leq Y_s)-g(Y_l))+n^{-1}(1-g(Y_l))\right]^{2k}|Y_l\right\}\right\}. \end{aligned}$$

Then (49) follows from two inequalities (see Petrov 1975): (i) For independent and mean zero random variables $V_1,\dots,V_m$

$$\mathbb{E}\left\{\left|m^{-1}\sum_{i=1}^{m}V_i\right|^p\right\} \leq C_p^* m^{-p/2-1}\sum_{i=1}^{m}\mathbb{E}\{|V_i|^p\}, \;\; p\geq 2, \tag{50}$$

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

$$|u+v|^p \leq 2^{p-1}(|u|^p+|v|^p), \;\; p\geq 1. \tag{51}$$

Now we can continue evaluation of $E_{h^{X^*}}\{(\hat{\theta}_j - \theta_j)^2\}$. For the second sum in (46),

$$\mathbb{E}_{h^{X^*}}\left\{\left[n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\varphi_j(Y_l)I(Y_l\in[a,a+b])[g(Y_l)-\eta_l/n]}{g^2(Y_l)}\right]^2\right\}$$
$$=\mathbb{E}_{h^{X^*}}\left\{n^{-2}\sum_{l,k=1}^{n}\frac{\Delta_l\Delta_k\varphi_j(Y_l)\varphi_j(Y_k)I((Y_l,Y_k)\in[a,a+b]^2)[g(Y_l)-\eta_l/n][g(Y_k)-\eta_k/n]}{g^2(Y_l)g^2(Y_k)}\right\}. \tag{52}$$

To continue the evaluation, we are considering the conditional expectation of a particular factor in (52). For any pair $(x, y) \in [a, a+b]^2$ and $k \neq l$, we can write

$$\mathbb{E}_{h^{X^*}}\{[g(Y_l)-\eta_l/n][g(Y_k)-\eta_k/n]|Y_l=x, Y_k=y\}$$
$$=\mathbb{E}_{h^{X^*}}\left\{\left[n^{-1}\sum_{s\in\{1,\ldots,n\}\setminus\{k,l\}}(I(T_s\le x\le Y_s)-g(x))+n^{-1}(1+I(y\ge x)-2g(x))\right]\right.$$
$$\left.\times\left[n^{-1}\sum_{s\in\{1,\ldots,n\}\setminus\{k,l\}}(I(T_s\le y\le Y_s)-g(y))+n^{-1}(1+I(x\ge y)-2g(y))\right]\right\}$$
$$=n^{-2}(1+I(y\ge x)-2g(x))(1+I(x\ge y)-2g(y))$$
$$+(n-2)n^{-2}[\mathbb{E}_{h^{X^*}}\{I(T\le x\le Y)I(T\le y\le Y)\}-g(x)g(y)]. \tag{53}$$

The last expectation in (53) can be simplified via using the following equality,

$$\mathbb{E}_{h^{X^*}}\{I(T\le x\le Y)I(T\le y\le Y)\}=\mathbb{P}_{h^{X^*}}(Y\ge\max(x,y),T\le\min(x,y))$$
$$=p^{-1}G^Y(\max(x,y))F^{T^*}(\min(x,y)).$$

Using this equality in (53) and then the obtained result in (52), we get with the help of (49)

$$\left|\mathbb{E}_{h^{X^*}}\left\{\left[n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\varphi_j(Y_l)I(Y_l\in[a,a+b])[g(Y_l)-\eta_l/n]}{g^2(Y_l)}\right]^2\right\}\right|$$
$$\le C^*n^{-2}+\frac{n-2}{n^4}\sum_{l=1}^{n}\sum_{k\in\{1,\ldots,n\}\setminus\{l\}}\mathbb{E}_{h^{X^*}}\left\{\frac{\Delta_l\Delta_k\varphi_j(Y_l)\varphi_j(Y_k)I((Y_l,Y_k)\in[a,a+b]^2)}{g^2(Y_l)g^2(Y_k)}\right.$$
$$\left.\times[p^{-1}G^Y(\max(Y_l,Y_k))F^{T^*}(\min(Y_l,Y_k))-g(Y_l)g(Y_k)]\right\}$$
$$=C^*n^{-2}+\frac{(n-1)(n-2)}{n^3}\mathbb{E}_{h^{X^*}}\left\{\Delta_1\Delta_2\varphi_j(Y_1)\varphi_j(Y_2)I((Y_1,Y_2)\in[a,a+b]^2)\right.$$

$$\times \frac{p^{-1}G^{X^*}(\max(Y_1, Y_2))G^{Z^*}(\max(Y_1, Y_2))F^{T^*}(\min(Y_1, Y_2)) - g(Y_1)g(Y_2)}{g^2(Y_1)g^2(Y_2)}\Bigg\}$$

$$\leq C^* n^{-1}[b'_j + n^{-1}], \qquad \sup_{h^{X^*}\in\mathcal{S}(\alpha,Q,a,b)} \sum_{j=0}^{\infty}[b'_j]^2 < C^* < \infty. \tag{54}$$

In the last line

$$b'_j := \mathbb{E}_{h^{X^*}}\Big\{\Delta_1\Delta_2\varphi_j(Y_1)\varphi_j(Y_2)I((Y_1, Y_2) \in [a, a+b]^2)$$
$$\times[p^{-1}G^{X^*}(\max(Y_1, Y_2))G^{Z^*}(\max(Y_1, Y_2))F^{T^*}(\min(Y_1, Y_2))$$
$$-g(Y_1)g(Y_2)]g^{-2}(Y_1)g^{-2}(Y_2)\Big\},$$

note that b'_j are particular Fourier coefficients of the tensor product cosine basis $\varphi_i(x)\varphi_j(y)$ on $[a, a+b]^2$, and therefore, according to the Bessel inequality and (35) we have $\sum_{j=0}^{\infty}(b'_j)^2 \leq C^*[\int_a^{a+b}[h^{X^*}(y)]^2 dy]^{1/2}$, and this verifies the last inequality in (54). Finally, inequality (49) allows us to conclude that the expectation of the squared third term in (46) is at most $C^* n^{-2}$. Combining the obtained results in (46), together with the Cauchy inequality $(c_1 + c_2)^2 \leq (1+\gamma)c_1^2 + (1+\gamma^{-1})c_2^2$, which is valid for any $\gamma > 0$, we conclude that

$$\mathbb{E}_{h^{X^*}}\{(\hat{\theta}_j - \theta_j)^2\} = \left[b^{-1}\int_a^{a+b} h^{X^*}(y)g^{-1}(y)dy\right] n^{-1}(1 + o_j^*(1) + o_n^*(1))$$
$$=: d^* n^{-1}(1 + o_j^*(1) + o_n^*(1)), \tag{55}$$

where $o_j^*(1)$ and $o_n^*(1)$ are bounded and vanish (as j and n increase) uniformly over $h^{X^*} \in \mathcal{S}(\alpha, Q, h_0, a, b)$. Note that $d^* = d(1 + o_n^*(1))$ where d is defined in (11).

Now we are establishing another general result. We are exploring the mean squared error of the shrinkage estimator $\lambda_j\hat{\theta}_j$ of θ_j where $\lambda_j \in [0, 1]$.

Write,

$$\mathbb{E}_{h^{X^*}}\{(\lambda_j\hat{\theta}_j - \theta_j)^2\}$$
$$= \lambda_j^2\mathbb{E}_{h^{X^*}}\{(\hat{\theta}_j - \theta_j)^2\} + (1-\lambda_j)^2\theta_j^2 - 2(1-\lambda_j)\lambda_j\mathbb{E}_{h^{X^*}}\{\hat{\theta}_j - \theta_j\}\theta_j. \tag{56}$$

Using (45) and (55) we continue (56) and conclude that

$$\mathbb{E}_{h^{X^*}}\{(\lambda_j\hat{\theta}_j - \theta_j)^2\} = \left[\lambda_j^2 n^{-1}d^* + (1-\lambda_j)^2\theta_j^2\right] + n^{-1}\rho_j(h^{X^*}, n, \lambda_j), \tag{57}$$

where $|\rho_j(h^{X^*}, n, \lambda_j)| \leq C^*[\lambda_j^2(o_j^*(1) + o_n^*(1)) + |\theta_j|]$.

For a positive integer J^*, introduce the following linear estimator of $h^{X^*}(x)$ for $x \in [a, a+b]$,

$$\bar{h}(x, \{\lambda_j\}) := \sum_{j=0}^{J^*}\lambda_j\hat{\theta}_j\varphi_j(x). \tag{58}$$

Using (57) and the Parseval identity, we can evaluate the MISE of the linear estimator (58),

$$\begin{aligned}
&\mathbb{E}_{h^{X^*}}\left\{\int_a^{a+b}(\bar{h}(x,\{\lambda_j\})-h^{X^*}(x))^2\mathrm{d}x\right\}\\
&\quad=\sum_{j=0}^{J^*}\left[n^{-1}d^*\lambda_j^2+(1-\lambda_j)^2\theta_j^2\right]+\sum_{j>J^*}\theta_j^2+n^{-1}\sum_{j=0}^{J^*}\rho_j(h^{X^*},n,\lambda_j)\\
&\quad\le\sum_{j=0}^{J^*}\left[n^{-1}d^*\lambda_j^2+(1-\lambda_j)^2\theta_j^2\right]+\sum_{j>J^*}\theta_j^2\\
&\qquad+\left[o_n^*(1)n^{-1}\sum_{j=0}^{J^*}\lambda_j^2+C^*n^{-1}I\left(\sum_{j=0}^{J^*}\lambda_j^2<\ln\ln(n)\right)\right],
\end{aligned}\tag{59}$$

where $o_n^*(1)\to 0$ as $n\to\infty$ uniformly over hazard rates h^{X^*} from the local Sobolev class. The indicator in (59) 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. The latter is not the case for the dealer estimator, but the general inequality (59) holds for all cases and this will allow us to use it later for any sequence λ_j.

Now it is a quick calculation to establish the upper bound for the dealer estimator $\check{h}$ defined in (30). Note that the dealer estimator is linear, and then using (27) and (59) we can write,

$$\begin{aligned}
&\mathbb{E}_{h^{X^*}}\left\{\int_a^{a+b}(\check{h}(x)-h(x))^2\mathrm{d}x\right\}\\
&\quad=\sum_{j=0}^{J_0}[n^{-1}d^*]+\sum_{j=J_0+1}^{J}[n^{-1}d^*(1-(j/J)^\alpha)^2+(j/J)^{2\alpha}\theta_j^2]\\
&\qquad+\sum_{j>J}\theta_j^2+o_n^*(1)n^{-2\alpha/(2\alpha+1)}.
\end{aligned}\tag{60}$$

Now, plug in expressions for J_0 and J, recall that $d^*=d(1+o_n^*(1))$, and then a straightforward algebra implies that for the considered $\{\theta_j\}$ the right side of (60) coincides with the lower bound (7). The upper bound of Theorem 1 is verified.

Now we are establishing that the lower bound (18) of Theorem 2 is attainable by a dealer estimator. We begin with the result of Efromovich (2001). Set

$$\alpha'=\begin{cases}\alpha & \text{if } \alpha \text{ is even,}\\ \alpha-1 & \text{if } \alpha \text{ is odd.}\end{cases}\tag{61}$$

Lemma 1 *Let a function $q(x)$ be α-times differentiable on $[a,a+b]$ and $\int_a^{a+b}[q^{(\alpha)}(x)]^2\mathrm{d}x<\infty$. Then there exists a unique polynomial–cosine expansion*

$$q(x) = \sum_{i=1}^{\alpha'} c_i x^i + \sum_{j=0}^{\infty} v_j \varphi_j(x), \quad x \in [a, a+b], \tag{62}$$

where the coefficients c_i are defined from the relations for the polynomial term $p(x) := \sum_{i=1}^{\alpha'} c_i x^i$,

$$p^{(r)}(a) = q^{(r)}(a), \quad p^{(r)}(a+b) = q^{(r)}(a+b), \tag{63}$$

that hold for all positive and odd $r < \alpha$. In (62) it is understood that for $\alpha = 1$ (and correspondingly $\alpha' = 0$) the term $\sum_{i=1}^{0} c_i x^i = 0$. Moreover,

$$\sum_{j=1}^{\infty} (\pi j/b)^{2\alpha} v_j^2 = \int_a^{a+b} [q^{(\alpha)}(x)]^2 \mathrm{d}x - (1 - \alpha + \alpha')[\alpha'! c_{\alpha'}]^2. \tag{64}$$

Relation (62) explains the underlying idea of the proposed estimator. Coefficients v_j are no longer Fourier coefficients, but they still belong to ellipsoid (64). As a result, if we can estimate coefficients c_i in (62), then we have a path for using the previous solution for finding an efficient dealer estimator. To follow this methodology, we need to introduce a new notation. It is convenient to rewrite (62) via orthonormal Legendre polynomials. Denote the Legendre polynomial basis on $[a, a+b]$ by $\{L_i(x),\ i = 0, 1, \ldots\}$ where $L_0(x) = 1$ and $L_i(x) = (x^i - \sum_{s=0}^{i-1} \langle x^i, L_s \rangle L_s(x)) \| x^i - \sum_{s=0}^{i-1} \langle x^i, L_s \rangle L_s(x) \|^{-1}$. Here $\langle f, g \rangle = \int_a^{a+b} f(x) g(x) \mathrm{d}x$ and $\| f \| = \sqrt{\langle f, f \rangle}$ denote the inner product and norm in $L_2([a, a+b])$, respectively. Also set $J' = \lfloor n^{1/(2\alpha+1)} \ln(n) \rfloor$ and rewrite (62) as

$$q(x) = \sum_{j=1}^{\alpha'} \beta_i L_i(x) + \sum_{j=0}^{\infty} v_j \varphi_j(x) = \sum_{j=0}^{J'} \kappa_j \varphi_j(x) + \sum_{i=1}^{\alpha'} \beta_{J'i} L_{J'i}(x) + \sum_{j > J'} \kappa_{J'j} \varphi_j(x), \tag{65}$$

where $\kappa_j = \langle q, \varphi_j \rangle$, $\beta_{J'i} = \langle q, L_{J'i} \rangle$, and

$$L_{J'i}(x) = [L_i(x) - \sum_{j=0}^{J'} \langle L_i, \varphi_j \rangle \varphi_j(x) - \sum_{s=1}^{i-1} \langle L_i, L_{J's} \rangle L_{J's}(x)] / N_{J'i}$$

is the $(1 + J' + i)$th element of an orthonormal system obtained by applying Gram–Schmidt orthogonalization procedure to the functions $\{\varphi_0, \ldots, \varphi_J, L_1, \ldots, L_{\alpha'}\}$; $N_{J'i} = \| L_i(x) - \sum_{j=0}^{J} \langle L_i, \varphi_j \rangle \varphi_j(x) - \sum_{s=1}^{i-1} \langle L_i, L_{J's} \rangle L_{J's}(x) \|$; $\kappa_{J'j} = \kappa_j - \langle \sum_{i=1}^{\alpha'} \beta_{J'i} L_{J'i}(x), \varphi_j \rangle$, $j > J'$.

Now, similarly to (28), we can estimate Fourier coefficients $\beta_{J'i}$ by

$$\hat{\beta}_{J'i} := \sum_{l=1}^{n} \Delta_l L_{J'i}(Y_l) \eta_l^{-1} I(Y_l \in [a, a+b]), \tag{66}$$

coefficients θ_j by

$$\tilde{\theta}_j := \sum_{l=1}^{n} \varphi_j(Y_l)\eta_l^{-1} I(Y_l \in [a, a+b]), \tag{67}$$

and coefficients β_i by

$$\tilde{\beta}_i = \left[\tilde{\beta}_{J'i} - \sum_{r=i+1}^{\alpha'} \tilde{\beta}_r \langle L_r, L_{J'i} \rangle \right] / N_{J'i}, \quad i = \alpha', \alpha' - 1, \ldots, 1. \tag{68}$$

Then the proposed dealer estimator of the hazard (compare with the dealer estimator (30) and note that the same sequences J_0 and J are used) is

$$\check{h} := \sum_{i=1}^{\alpha'} \tilde{\beta}_i L_i(x) + \sum_{j=0}^{J_0} \tilde{\theta}_j \varphi_j(x) + \sum_{j=J_0+1}^{J} (1 - (j/J)^{\alpha}) \tilde{\theta}_j \varphi_j(x). \tag{69}$$

To establish that this dealer estimator is minimax, we make several steps. First, let us note that estimators (66) and (67) are sample mean estimators of Fourier coefficients, and hence (45), as well as (55) for $\tilde{\theta}_j$, are applied here. Second, a direct calculation shows that $\langle L_l, L_{J'i} \rangle = O(1)[J']^{-3}$ for $\imath \geq i$. This allows us to write,

$$E_{h_{X^*}}\{(\tilde{\beta}_i - \beta_i)^2\} \leq C^*[J']^{-2\alpha+3} = C^* n^{-(2\alpha+3)/(2\alpha+1)} [\ln(n)]^{-2\alpha+3} \text{ for } i = 1, \ldots, \alpha'. \tag{70}$$

Third, we note that

$$\sum_{i=1}^{\alpha'} \tilde{\beta}_i L_i(x) + \sum_{j=0}^{J'} \tilde{\theta}_j \varphi_j(x) = \sum_{j=0}^{J'} \tilde{\kappa}_j \varphi_j(x) + \sum_{i=1}^{m'} \tilde{\beta}_i (L_i(x) - \sum_{j=0}^{J'} \langle L_i, \varphi_j \rangle \varphi_j(x)).$$

Now we can use the above-presented general proof of the upper bound for the dealer estimator (30). Following that proof and using established properties of Fourier estimates, we verify that the MISE of the dealer estimator (69) attains the lower bound of Theorem 2. □

Acknowledgements The research is supported by NSF Grant DMS-1513461. Suggestions of the reviewers and the AE are appreciated.

References

Andersen, P. K., Borgan, O., Gill, R. D., Keiding, N. (1993). *Statistical models based on counting processes*. New York: Springer.

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

Bagkavos, D., Patil, P. (2012). Variable bandwidths for nonparametric hazard rate estimation. *Communications in Statistics - Theory and Methods*, *38*, 1055–1078.

Bindera, N., Schumache, M. (2014). Missing information caused by death leads to bias in relative risk estimates. *Journal of Clinical Epidemiology*, *67*, 1111–1120.
Bremhorsta, V., Lamberta, F. (2016). Flexible estimation in cure survival models using Bayesian p-splines. *Computational Statistics & Data Analysis*, *93*, 270–284.
Brody, T. (2012). *Clinical trials*: *Study design, endpoints and biomarkers, drug safety, and FDA and ICH guidelines*. New York: Elsevier.
Brunel, E., Comte, F. (2008). Adaptive estimation of hazard rate with censored data. *Communications in Statistics - Theory and Methods*, *37*, 1284–1305.
Cao, R., Janssen, P., Veraverbeke, N. (2005). Relative hazard rate estimation for right censored and left truncated data. *Test*, *14*, 257–280.
Cox, D. R., Oakes, D. (1984). *Analysis of survival data*. London: Chapman&Hall.
Daepp, I., Hamilton, M., West, G., Bettencourt, L. (2015). The mortality of companies. *Journal of the Royal Society Inference*, *12*, 1–8.
Döhler, S., Rüschendorf, L. (2002). Adaptive estimation of hazard functions. *Probability and Mathematical Statistics*, *22*, 355–379.
Efromovich, S. (1999). *Nonparametric curve estimation*: *Methods, theory and applications*. New York: Springer.
Efromovich, S. (2001). Density estimation under random censorship and order restrictions: From asymptotic to small samples. *Journal of the American Statistical Association*, *96*, 667–685.
Efromovich, S. (2016). Minimax theory of nonparametric hazard rate estimation: Efficiency and adaptation. *Annals of the Institute of Mathematical Statistics*, *68*, 25–75.
Efromovich, S., Pinsker, M. (1982). Estimation of a square-integrable probability density of a random variable. *Problems of Information Transmission*, *18*, 19–38.
Fleming, T., Harrington, D. (2011). *Counting processes and survival analysis*. New York: Wiley.
Gill, R. (2006). *Lectures on survival analysis*. New York: Springer.
Hagar, Y., Dukic, V. (2015). Comparison of hazard rate estimation in R. arXiv: 1509.03253v1
Huang, C., Qin, J. (2013). Semiparametric estimation for the additive hazard model with left-truncated and right-censored data. *Biometrika*, *100*, 877–888.
Jankowski, H., Wellner, J. (2009). Nonparametric estimation of a convex bathtub-shaped hazard function. *Bernoulli*, *15*, 1010–1035.
Klein, J. P., Moeschberger, M. L. (2003). *Survival analysis*: *Techniques for censored and truncated data*. New York: Springer.
Klugman, S., Panjer, H., Willmot, G. (2012). *Loss models*: *From data to decisions*. New York: Wiley.
Lee, E., Wang, J. (2013). *Statistical methods for survival data analysis*. New York: Wiley.
Lu, X., Min, L. (2014). Hazard rate function in dynamic environment. *Reliability Engineering and System Safety*, *130*, 50–60.
Müller, H., Wang, J. (2007). Density and hazard rate estimation. In F. Ruggeri, R. Kenett, F. Faltin (Eds.), *Encyclopedia of statistics in quality and reliability* (pp. 517–522). Chichester: Wiley.
Patil, P., Bagkavos, D. (2012). Histogram for hazard rate estimation. *Sankhya B*, *74*, 286–301.
Petrov, V. (1975). *Sums of independent random variables*. New York: Springer.
Pierce, J., Natarajan, L., Caan, B., Parker, B., Greenberg, R., Flatt, S., et al. (2007). Influence of a diet very high in vegetables, fruit, and fiber and low in fat on prognosis following treatment for breast cancer: The women's healthy eating and living (WHEL) randomized trial. *Journal of the American Medical Association*, *298*, 289–298.
Qian, J., Betensky, R. (2014). Assumptions regarding right censoring in the presence of left truncation. *Statistics and Probability Letters*, *87*, 12–17.
Rock, C., Flatt, S., Nataryan, L., Thomson, C., Bardwell, W., Newman, V., et al. (2005). Plasma carotenoids and recurrence-free survival in women with a history of breast cancer. *Journal of Clinical Oncology, American Society of Clinical Oncology*, *23*, 6631–6638.
Shi, J., Chen, X., Zhou, Y. (2015). The strong representation for the nonparametric estimation of length-biased and right-censored data. *Statistics and Probability Letters*, *104*, 49–57.
Silverman, B. (1986). *Density estimation for statistics and data analysis*. London: Chapman&Hall.
Su, Y., Wang, J. (2012). Modeling left-truncated and right censored survival data with longitudinal covariates. *Annals of Statistics*, *40*, 1465–1488.
Talamakrouni, M., Van Keilegom, I., El Ghouch, A. (2016). Parametrically guided nonparametric density and hazard estimation with censored data. *Computational Statistics and Data Analysis*, *93*, 308–323.

Uzunogullari, U., Wang, J. (1992). A comparison of hazard rate estimators for left truncated and right censored data. *Biometrika*, *79*, 297–310.

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

Wang, Y., Liang, B., Tong, X., Marder, K., Bressman, S., Orr-Urtreger, A., et al. (2015). Efficient estimation of nonparametric genetic risk function with censored data. *Biometrika*, *102*, 515–532.

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

Zhang, F., Zhou, Y. (2013). Analyzing left-truncated and right-censored data under cox model with long-term survivors. *Acta Mathematicae Applicatae Sinica*, *29*, 241–252.