Supplementary Materials for the paper

Hazard Rate Estimation for Left Truncated and Right Censored Data

Small LTRC Samples and Lower Bounds in Hazard Rate Estimation

by Sam Efromovich and Jufen Chu

In what follows (k) means kth line in the paper, while a kth line in the supplementary materials is denoted as (A.k). Here Figures 1-4 of the paper are Figures 4-8 and all figures are colored.

1. ESTIMATOR FOR SMALL SAMPLES

The proposed estimator is motivated by the asymptotic theory and an intensive numerical study presented in the next section. The cosine basis, used in the asymptotic analysis, performs exceptionally well for a majority of hazards discussed in the literature but can be improved for hazards that change rapidly near boundaries. For instance, there is an important class of bathtub-shaped hazards that can be seen in the bottom row of Figure 1. Note that the hazards are almost vertical near boundaries. According to Jankowski and Welner (2009), all traditional nonparametric estimators do not perform well for bathtub hazards, and hence in that paper a smart procedure of piece-wise approximations is proposed. Our numerical study of bathtub hazards shows that cosine estimators perform well in terms of the ISE, but their visual appeal near boundaries is not satisfactory. To resolve the boundary issue, it is recommended to complement the cosine basis by linear and quadratic terms as is explained in the proof of Theorem 2. This simple method, as we will see shortly, yields satisfactory visualization of bathtub shapes. At the same time, the recommended estimator uses linear and quadratic terms only if there is a statistically significant evidence for the need of a boundary correction.

To utilize the proposed method, we are using the Gram-Schmidt orthonormalization procedure. Note that it is always better to deal with a basis defined on a fixed interval, say $[0,1]$ and then correspondingly rescale data. Let J be a nonnegative integer. Choose $J+1$ elements of the cosine basis $\varphi_0^*(x) := 1$, $\varphi_j^*(x) := 2^{1/2}\cos(\pi j x)$, $j = 1, 2, \ldots$ and complement them by linear and quadratic terms using the Gram-Schmidt orthonormalization,

$$\psi_{J,1}(x) := \frac{x - 1/2 - 2\pi^{-2}\sum_{j=1}^{J} j^{-2}((-1)^j - 1)\cos(\pi j x)}{[1/12 - 2\pi^{-4}\sum_{j=1}^{J} j^{-4}((-1)^j - 1)^2]^{1/2}}, \tag{A.1}$$

$$\psi_{J,2}(x) := \frac{x^2 - 1/3 - (2/\pi)^2\sum_{j=1}^{J} j^{-2}(-1)^j\cos(\pi j x) - c_*\psi_{J,J+1}(x)}{[4/45 - c_*^2 - 8\pi^{-4}\sum_{j=1}^{J} j^{-4}]^{1/2}}. \tag{A.2}$$

In (A.1) and (A.2) it is assumed that $\sum_{j=1}^{0} x_j := 0$ and in (A.2)

$$c_* := \frac{1/12 + 4\pi^{-4}\sum_{j=1}^{J} j^{-4}((-1)^j - 1)}{[1/12 - 2\pi^{-4}\sum_{j=1}^{J} j^{-4}((-1)^j - 1)^2]^{1/2}}. \tag{A.3}$$

The proposed estimator for an interval of interest $[a, a+b]$ is

$$\tilde{h}^{X^*}(x) := b^{-1}[\sum_{j=0}^{\tilde{J}} \max(1 - \hat{d}/(n\tilde{\theta}_j^2), 0)\tilde{\theta}_j\varphi_j^*((x-a)/b) \tag{A.4}$$

$$+ \sum_{i=1}^{2} I(\tilde{\kappa}_{J,i}^2 > 2\ln(n)\tilde{d}n^{-1})\tilde{\kappa}_{J,i}\psi_{J,i}((x-a)/b), \;\; x \in [a, a+b]. \tag{A.5}$$

Here

$$\tilde{\theta}_j := \sum_{l=1}^{n} \Delta_l\varphi_j^*([Y_l - a]/b)\eta_l^{-1} I(Y_l \in [a, a+b]) \tag{A.6}$$

and

$$\tilde{\kappa}_{J,i} := \sum_{l=1}^{n} \Delta_l\psi_{J,i}([Y_l - a]/b)\eta_l^{-1} I(Y_l \in [a, a+b]) \tag{A.7}$$

are estimators of Fourier coefficients,

$$\hat{d} := \hat{d}_{a,b} := nb^{-1} \sum_{l=1}^{n} \Delta_l \eta_l^{-2} I(Y_l \in [a, a+b]) \tag{A.8}$$

is the estimator of the coefficient of difficulty,

$$\tilde{J} := \text{argmin}_{0 \le J \le 4+(1/2)\ln(n)} [2(J+3)\tilde{d}n^{-1} - \sum_{j=0}^{J} \tilde{\theta}_j^2 - \tilde{\kappa}_{J,1}^2 - \tilde{\kappa}_{J,2}^2] \tag{A.9}$$

is the cutoff which minimizes the empirical ISE, and

$$\eta_l := \sum_{s=1}^{n} I(T_s \le Y_l \le Y_s). \tag{A.10}$$

Because $\eta_l \ge 1$, using its reciprocal is correct.

Let us comment on the estimator. Not surprisingly, some of its components are the same as in the asymptotic estimators of Section 3. The cosine part (A.4) is motivated by the asymptotic theory, and it is a classical cosine series estimator. Two-terms polynomial part (A.5) is used solely for correcting boundary effects, and it is used only if there is a statistically significant evidence for doing this; the latter is done by the universal thresholding. Because estimators of Fourier coefficients are sample mean estimators, it is straightforward to establish pointwise confidence bands, see Wasserman (2005). Let us also comment on a question which is often asked about the used cosine basis complemented by two polynomials. Why do not use a polynomial basis instead? The answer is that while this basis is excellent for polynomial functions, in general polynomial approximations are not visually appealing and there is no suitable approximation theory for differentiable functions; see a discussion in Efromovich (1999,ch.2). Recall that the latter is the reason why a smooth piece-wise polynomial approximation is often recommended, which is called the approximation by splines. Interestingly, spline bases, used for the asymptotic analysis, resemble the cosine basis, and this sheds another light on the recommended basis.

2. NUMERICAL STUDY OF THE ESTIMATOR

To evaluate performance of the proposed estimator (A.4) for small sample sizes, it is convenient to have a benchmark to compare with. The chosen benchmark is an oracle-estimate that knows an underlying LTRC mechanism as well as an estimated hazard rate. This oracle is a familiar kernel estimator (see, for instance, Uzunogullari and Wang 1992 as well as the review in Hagar and Dukic 2015 and the R-package *muhaz*) which is transformed into oracle by: (i) Using the golden-rule bandwidth based on all unknown to a statistician underlying functions (see more in Marron and Wand 1992 and Efromovich 1999). This solves the issue of adaptation to unknown smoothness of an underlying hazard rate; (ii) Using all observations (not just belonging to $[a, a+b]$). This solves the familiar boundary problem. Then the oracle-estimator is defined as

$$\check{h}(x,\nu(x)) := n^{-1}\sum_{l=1}^{n} K\Big(\frac{x-Y_l}{\nu(x)}\Big)\frac{\Delta_l}{\nu(x)g(Y_l)} \tag{A.11}$$

where $K(x)$ is the Gaussian kernel, $g(x) = \mathbb{P}(T \le x \le Y)$, and $\nu(x)$ is the golden-rule oracle-bandwidth

$$\nu(x) := n^{-1/5}\frac{[h^{X^*}(x)\int K^2(t)dt]^{1/5}}{[h''(x)\int t^2K(t)dt]^{2/5}[g(x)]^{1/5}}. \tag{A.12}$$

Note that (A.11) is not a data-driven estimator and it is used to create a benchmark. Further, using the positive Gaussian kernel for small samples is recommended in Marron and Wand (1992) because, together with the optimal adaptation, this kernel oracle outperforms oracles using asymptotically optimal kernels taking negative values.

Now let us describe statistical experiments used in the study. Underlying distributions of X^* are either Weibull $W(\gamma,\beta)$, where γ and β are the shape and scale parameters, respectively, or it is a Bathtub (BT) distribution generated by $X^* := \min(V_1, V_2)$ with V_1 and V_2 being $W(0.3, 1)$ and $W(15, 1)$, respectively. Weibull distribution is of interest because its hazard rate is decreasing, constant and increasing for the shape parameter $\gamma < 1$,

$\gamma = 1$ and $\gamma > 1$, respectively. Bathtub distributions, implying a convex bathtub-shaped hazard function, occur in a number of applications; an interesting discussion can be found in Jankowski and Welner (2009). The above-described distributions allow us to test performance of the estimator on increasing, decreasing, and convex bathtub-shaped hazard functions, see solid lines in Figure 1. Figure 1 also exhibits series and oracle estimates for a single simulation, and note how well the oracle performs near boundaries due to using observations beyond the interval of estimation. In experiments with independent random variables the underlying distributions of T^* and Z^* are either exponential or uniform, and experiments with dependent T^* and Z^* will be explained shortly. Several different intervals $[a, b]$ and $n \in \{100, 200, 300, 400, 500, 1000\}$ are considered (next section is devoted to discussion of the interval of estimation). For each experiment, defined by underlying distributions as well as by an interval $[a, a + b]$ and a sample size, 5000 simulations are conducted. For each simulation the empirical integrated squared error of the oracle (ISEO) and the empirical integrated squared error of the proposed data-driven series estimator (ISEE) are calculated. Then the median ratio (over 5000 simulations) of ISEO/ISEE, denoted as M, is shown in Table 1 together with the average number of observations fallen within a studied interval $[a, a + b]$, denoted as m. The corresponding entry in Table 1 is written as M/m. The number m is informative because empirical Fourier coefficients (A.6) and (A.7) are based only on $Y_l \in [a, a + b]$. In other words, we can say that m is the "effective" sample size used to estimate Fourier coefficients (of course, all n observations are used to estimate nuisance functions).

Five bottom rows in Table 1 are devoted to experiments where T^* and Z^* are dependent. In these experiments $Z^* := T^* + R^*$ where T^* and R^* are independent and X^* is independent of (T^*, R^*). To define R^*, denote by $U(c_1, c_2)$ a uniform on $[c_1, c_2]$ random variable which is independent of all other random variables involved in an experiment. In "Case1" we set $R^* = U(0, 2)$ with probability 0.7 and $R^* = 2$ with probability 0.3. This is the case when some subjects may leave a study before its end and others are censored by the end of the

study. In "Case2" we add the possibility for a subject to leave a study during a single follow-up. Namely, we set $R^* = U(0,2)$ with probability 0.4, $R^* = 1$ with probability 0.3, and $R^* = 2$ with probability 0.3. In "Case3" a subject is additionally "allowed" to leave a study immediately at the baseline, and then we set $R^* = U(0,5)$ with probability 0.2, $R^* = 0$ with probability 0.2, $R^* = 2.5$ with probability 0.2, and $R^* = 5$ with probability 0.4.

Table 1. Results of Monte Carlo simulations. Distributions are denoted as $W(\gamma,\beta)$, BT, $U(c_1,c_2)$ and $E(\lambda)$ for Weibull with shape parameter γ and scale parameter β, Bathtub corresponding to the minimum of random variables with distributions $W(0.3,1)$ and $W(15,1)$, uniform on the interval $[c_1,c_2]$, and exponential with $1/\lambda$ being the mean, respectively. An entry in the Table is written as M/m where M is the median of 5000 ratios ISEO/ISEE and m is the average number of observations fallen within interval $[a,a+b]$.

					n					
X^*	T^*	Z^*	$[a,a+b]$	d	100	200	300	400	500	1000
$W(3,4)$	$U(0,3)$	$U(3,10)$	$[0.5,4]$	0.54	0.80/59	0.92/118	1.07/177	1.20/237	1.15/295	1.41/590
$W(3,4)$	$U(0,3)$	$U(3,10)$	$[1,4]$	0.61	0.71/58	0.87/116	0.88/174	1.10/232	1.01/290	1.32/580
$W(3,4)$	$U(0,3)$	$U(3,10)$	$[1,5]$	1.72	0.71/80	0.89/161	1.02/241	1.03/320	1.26/400	1.41/800
$W(1.2,5)$	$E(1)$	$E(0.05)$	$[1,5]$	0.41	0.92/48	1.03/95	1.27/143	1.33/200	1.45/251	1.43/480
$W(0.5,2)$	$E(2)$	$E(0.05)$	$[0.5,8]$	0.63	0.72/46	0.80/91	0.86/138	0.97/182	1.07/230	1.44/462
$W(0.5,2)$	$E(5)$	$E(0.05)$	$[0.1,3]$	0.65	0.69/47	0.83/95	0.90/142	0.94/190	1.01/237	1.24/474
$W(0.3,1)$	$E(2)$	$E(0.05)$	$[0.2,6]$	0.44	0.73/32	0.92/64	1.00/96	1.08/128	1.19/160	1.58/320
$W(3,2)$	$E(1)$	$E(0.15)$	$[1,2]$	1.90	0.91/38	0.99/74	1.13/111	1.11/151	1.24/188	1.56/371
$W(3,2)$	$E(1)$	$E(0.15)$	$[0.5,2]$	1.37	0.84/45	0.91/91	1.09/137	1.04/182	1.15/229	1.42/456
$W(3,2)$	$E(1.5)$	$E(0.1)$	$[0.5,2.5]$	3.33	0.76/72	0.94/144	1.03/215	1.12/287	1.10/360	1.47/720
BT	$E(80)$	$E(0.5)$	$[0.05,0.9]$	2.04	0.78/45	0.90/90	1.08/135	1.16/180	1.38/225	1.40/450
$W(2.5,1)$	$E(2)$	Case1	$[0.1,1]$	1.98	0.70/51	0.93/102	1.02/153	1.08/204	1.19/255	1.23/511
$W(4,1)$	$E(3)$	Case2	$[0.4,1]$	3.12	0.84/56	0.91/112	1.03/168	1.05/223	1.06/278	1.10/561
$W(5,2)$	$E(1)$	Case2	$[0.3,2.1]$	2.46	0.70/64	0.78/130	0.93/195	1.07/260	1.10/324	1.29/654
$W(1.1,1)$	$E(5)$	Case3	$[0.3,2]$	4.65	1.31/53	1.23/106	1.26/158	1.30/211	1.38/364	1.19/528
$W(1.1,7)$	$E(0.5)$	Case3	$[0.5,5]$	0.39	1.01/48	1.25/96	1.27/144	1.25/192	1.32/240	1.47/480

We begin discussion of results, presented in Table 1, with the "effective" sample sizes m. They are relatively small with respect to n. Just as an example, for $n = 200$ and X^* being $W(0.3,1)$, we have $m = 64$ and this is a very small (for a nonparametric setting) number of "efficient" observations. As a result, it is not surprising that overall the kernel oracle-estimator, that knows much more than an estimator, outperforms our data-driven estimator for smallest sample sizes 100 and 200 (this can be observed from values of M

being smaller than 1). For larger sample sizes the series estimator performs on par and even begins to outperform the oracle for n larger 400. Marron and Wand (1992) argue that the ratio M larger than 0.75 is a good outcome when a data-driven estimator is compared with an oracle. If we look at the ratios for $n = 100$, the smallest $M = 0.69$ and this is not a bad outcome because the corresponding effective size is $m = 47$. Furthermore, for $n = 200$ the smallest ratio is 0.78 and it passes the above-cited threshold. Parameter d, shown in column 5, is the coefficient of difficulty (11) and it sheds light on relative complexity of a particular experiment.

Results shown in five bottom raws of Table 1 support conclusion of Theorem 4 that the proposed estimator is not sensitive to deviations from the main model which assumes that hidden random variables are independent and continuous. Overall, we may conclude that the data-driven estimator performs reasonably well for small sample sizes and the variety of underlying models.

3. INTERVAL OF ESTIMATION

The literature on the effect/choosing of the interval of estimation $[a, a + b]$ is next to none even for the case of direct observations when the hazard rate is traditionally estimated over an interval $[0, b]$ with $b = 1$ being a popular choice. It is a difficult problem indeed, and hence its discussion is of a special interest.

The asymptotic theory, presented in Section 3 of the paper, sheds light on the effect of the interval on the MISE convergence. Namely, if the interval is fixed then asymptotically

$$MISE = P(\alpha, Q, 1)b^{1/(2\alpha+1)}\Big[b^{-1}\int_a^{a+b} h_0^{X^*}(v)g_0^{-1}(v)dv\Big]^{2\alpha/(2\alpha+1)} n^{-2\alpha/(2\alpha+1)}(1 + o_n(1)). \tag{A.13}$$

Expression in the square brackets is the coefficient of difficulty $d(a, a + b)$ (recall formula (11)), and function $g_0(x)$ is defined in (9) and it reflects the effect of the LTRC on the MISE.

Note that function $g_0(v)$ vanishes as $v \to 0$ because of the LT and it also vanishes as $v \to \infty$ due to the RC and the fact that $G_0^{X^*}(v) = o_v(1)$. Of course, in general it may even

vanish beyond some interval $[c_1, c_2]$, $0 < c_1 < c_2 < \infty$. This is what may preclude us from estimation over an interval dictated by substantive issues stemming from the motivating problem. Underlying functions F^{T^*}, G^{Z^*} and $G_0^{X^*}$, defining the function g_0, are unknown, and this explains complexity of the problem of choosing a feasible interval of estimation for the LTRC data.

To shed light on a possible solution, let us look at an example of particular functions $1/g_0(x)$ and $d(1, x)$ for the experiment 3 of Section 2. In Figure 2, the top diagram shows us function $1/g_0(x)$ for $x \in [1, 5]$. Beyond this interval the reciprocal of $g_0(x)$ increases very fast, for instance it takes on values 46 and 447 for $x = 6$ and $x = 7$, respectively. Similar outcomes have been observed for all other experiments, with the reciprocal of $g_0(x)$ sharply increasing beyond some specific intervals. This observation hints upon a method of choosing reasonable intervals of estimation based on the analysis of the reciprocal of $g_0(x)$. The coefficient of difficulty $d(1, x)$, $x = 1 + b$, $a = 1$ is shown in the second from the top diagram. In addition to the graphic, note that $d(1, 6) = 8$ and $d(1, 7) = 66$. Not surprisingly, the coefficient of difficulty also has sharply increasing tails.

The two bottom diagrams allow us to understand the effect of the interval $[a, a+b] = [1, x]$ on the proposed estimate (A.4). The solid line exhibits the underlying hazard. The hazard increases but its derivate is relatively small and makes no significant impact on the coefficient of difficulty (look again at the two top diagrams and make your own conclusion about effects of $1/g_0(x)$ and $h^{X^*}(x)$ on the coefficient of difficulty). Now let us look at estimates of the hazard for different intervals of estimation. The estimates are based on the same sample of size $n = 500$. The second from the bottom diagram shows us, by the long-dashed and short-dashed lines, estimates for intervals $[1, 4]$ and $[1, 5]$, respectively. Note that the estimate for the smaller interval, which corresponds to reasonably small values of $1/g_0(x)$ and moderate values of the coefficient of difficulty, is truly good. The estimate for the larger interval is still good but we may already notice signs of deterioration in the quality of estimation. This is due to relatively large values of $1/g_0(x)$ for $x > 4.5$. The situation changes more dramatically for larger intervals. The bottom diagram shows us estimates for intervals $[1, 6]$ (the long-dashed line) and $[1, 7]$ (the short-dashed line), and recall that the corresponding coefficients

of difficulty are 8 and 66. Right tails of the estimates wrongly exhibit the underlying hazard rate, and we can conclude that estimation of the hazard over intervals with large values of $1/g_0(x)$ should be avoided.

Repeated simulations for the experiments of Section 2 revealed similar outcomes. Our conclusion is that the analysis of $1/g_0(x)$, and to the lesser degree of the coefficient of difficulty (because its estimation involves the function $1/g_0(x)$), may be used in choosing the interval of estimation. But how well these two functions can be estimated?

Figure 3 sheds light on the issue. Here the sample of Figure 2 is used. The estimator of the coefficient of difficulty is defined in (A.8), and $1/g_0(x)$ is estimated by $n/\sum_{s=1}^{n} I(T_s \le x \le Y_s)$ (compare with (A.10) and note that the estimator is motivated by the definition $g_0(x) := \mathbb{P}(T \le x \le Y)$). The dotted lines in Figure 3 show us estimates of $1/g_0(x)$ and $d(1,x) = d(1,x)$, and the solid lines show the underlying functions. As we see from the diagrams, the estimates do a good job. Similar outcomes were observed for other experiments of Section 2. We may conclude that visualization of the estimate of $1/g_0(x)$, complemented by visualization of the estimate of the coefficient of difficulty, may give us a good idea of how to choose an interval of estimation.

Let us recall that the proposed approach was used for the analysis of the WHEL breast cancer data.

We finish this section by the following remark about the used minimax. Formula (7y) is important, and it is necessary to keep in mind that it is obtained for homogeneous Sobolev classes where parameters (α, Q) do not depend on the interval $[a, a+b]$. In general, parameters (α, Q) may depend on the interval. This fact does not change the main conclusion of the theory because our estimator adapts to (α, Q). At the same time, considering inhomogeneous Sobolev classes is a new and open problem.

4. Proofs of lower bounds in Theorems 1 and 2

Let us make a preliminary comment about verifying the two lower bounds. The proposed verification contains several steps. The first one is to replace function classes (4) and (13)

by a parametric subclass (the same for both function classes). The second step is to convert the study of the minimax MISE into the study of a sequence of minimax parametric mean squared errors. The third step is a traditional one, where a minimax risk is replaced by a Bayes risk with the least favorable prior distribution. The final step is to evaluate the sum of Bayes risks. These steps, of course, are known in the literature and go back to Pinsker (1980) and Efromovich and Pinsker (1982) for simpler settings of filtering and density estimation for direct observations, correspondingly. The considered problem of the hazard estimation for the LTRC is more complicated. Recently, Efromovich (2016) used a similar methodology to establish a lower bound for hazard estimates based on a sample from the random variable of interest X^*. While that proof cannot be used here directly due to the LTRC complications, some of its technical results may help us to make the proposed proof shorter; to make references on those results more transparent we are using the same notation whenever possible. In what follows $o_n(1)$ and Cs denote generic vanishing sequences and positive constants, and we may add the asterisk and write $o^*_n(1)$ 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, and recall that $n > 20$ so in what follows all sequences in n are well defined. Let us also stress that the lower bounds are asymptotic, and hence all assertions should be valid only for large sample sizes.

Now we begin step 1 of the proof by introducing a parametric function subclass of (4) and (13). According to (13), a perturbation $q(x)$ should be zero near boundaries. Also, (7) implies that the shape of anchor $h_0(x)$ affects the sharp minimax constant. To take care about these issues, we divide the interval $[a, a+b]$ into increasing (as $n \to \infty$) number of subintervals, then define parametric function classes on each subinterval, and finally smooth functions from each subinterval to satisfy (5) and (16). Denote the total number of subintervals as $s := s_n := 2+3\lfloor \ln(\ln(n)) \rfloor$. At the boundary subintervals we set $q(x) = 0$ and note that $[3\ln(\ln(n))]^{-2} < 1/s$; the latter is important for satisfying (14). Then for the $s-2$ inner subintervals, defined by index $k = 1, \ldots, s-2$, introduce notation $\mathcal{I}_{sk} := h_0^{-1}(a+kb/s)p_0^{-1}G^{Z^*}(a+kb/s)F^{T^*}(a+kb/s)G_0(a+kb/s)$ where $G_0(x) := e^{-\int_0^x h_0(v)dv}$, $p_0 := \mathbb{P}_{h_0}(\min(X^*, Z^*) > T^*)$,

$\mathcal{I}_s := [\sum_{k=1}^{s-2} \mathcal{I}_{sk}^{-1}]^{-1}$, $Q_{sk} := (1-1/s)\mathcal{I}_s\mathcal{I}_{sk}^{-1}Q$, $J := \lfloor b[n(2\alpha+1)(\alpha+1)s^{-2\alpha}(\alpha\pi^{2\alpha})^{-1}(1-s^{-1})\mathcal{I}_sQ]^{1/(2\alpha+1)}\rfloor$, $J_* = \lfloor J(s)/\ln(n)\rfloor$, $\varphi_{skj}(x) = (2s/b)^{1/2}\cos(\pi j[sb^{-1}(x-a)-k])$. Set $\vec{\nu}_{sk} := \{\nu_{skJ_*}, \ldots, \nu_{skJ}\}$.

Now introduce a sequence $\phi(n,v)$, $v \in (-\infty,\infty)$ of flattop nonnegative 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 $(-\infty,\infty)$, $0 \le \phi(n,v) \le 1$, $\phi(n,v) = 1$ for $2(\ln(n))^{-2} \le v \le 1-2(\ln(n))^{-2}$, and $|\phi^{(r)}(n,v)| \le C(\ln(n))^{2r}$ for $r = 1,\ldots,\alpha$. The kernel is constructed using a mollifier defined in Efromovich (1999, ch.7). Finally, set $\phi_{sk}(x) := \phi(n, sb^{-1}(x-a)-k)$.

For each of $s-2$ "inner" subintervals define a parametric function class,

$$\mathcal{H}_{sk} := \Big\{ f_{sk} : \ f_{sk}(x|\vec{\nu}_{sk}) := \sum_{j=J_*}^{J} \nu_{skj}\varphi_{skj}(x), \ \sum_{j=J_*}^{J} (\pi j s/b)^{2\alpha}\nu_{skj}^2 \le bQ_{sk},$$

$$|f_{sk}(x|\vec{\nu}_{sk})| \le [s^4\ln(n)]^{1/2} n^{-\alpha/(2\alpha+1)}, \quad x \in [a,a+b]\Big\}, k = 1,\ldots,s-2, \tag{A.14}$$

and then introduce a new parametric function class on $[a,a+b]$,

$$\mathcal{H}_s := \Big\{ h : \ h(x|\vec{\nu}_s) = h_0(x) + g(x|\vec{\nu}_s),$$

$$g(x|\vec{\nu}_s) := \sum_{k=1}^{s-2} [f_{sk}(x|\vec{\nu}_{sk}) - \mu_{sk}]\phi_{sk}(x),$$

$$\mu_{sk} := \frac{\int_a^{a+b} f_{sk}(z|\vec{\nu}_{sk})\phi_{sk}(z)dz}{\int_a^{a+b}\phi_{sk}(z)dz},$$

$$f_{sk} \in \mathcal{H}_{sk}, \quad h(x|\vec{\nu}_s) \ge 0\Big\}. \tag{A.15}$$

Here $\vec{\nu}_s := (\vec{\nu}_{s1}, \ldots, \vec{\nu}_{s(s-2)})$ and $\vec{\nu}_{sk} := (\nu_{skJ_*}, \ldots, \nu_{skJ})$. In what follows we may also use notation $\tilde{f}_{sk}(x) := f_{sk}(x) - \mu_{sk}$.

The term μ_{sk} is used solely to make function $g(x|\vec{\nu}_{sk})$, defined in (A.15), integrable to zero on $[a,a+b]$. The term is negligibly small for all our purposes. To show this, let us recall a known technical result (see, for instance, Efromovich 2001)).

Lemma S.1. *Let a function $q(x)$ be α-times differentiable on $[a, a+b]$, $\int_a^{a+b}[q^{(\alpha)}(x)]^2dx < \infty$, and for all positive and odd $r < \alpha$*

$$q^{(r)}(a) = q^{(r)}(a+b) = 0. \tag{A.16}$$

Then cosine coefficients $\theta_j := \int_a^{a+b} q(x)\varphi_j(x)dx$ of the function q satisfy the Parseval-type identity

$$\sum_{j=1}^{\infty}(\pi j/b)^{2\alpha}\theta_j^2 = \int_a^{a+b}[q^{(\alpha)}(x)]^2dx. \tag{A.17}$$

This result, together with (A.14), the assumed α-fold differentiability of $\phi(x)$ and the Cauchy-Schwarz inequality, allows us to write,

$$\mu_{sk}^2 = [\sum_{j=J_*}^{J}\nu_{skj}\int_a^{a+b}\varphi_{skj}(z)\phi_{sk}(z)dz/\int_a^{a+b}\phi_{sk}(z)dz]^2$$

$$\leq C^*s^2\sum_{j=J_*}^{J}\nu_{skj}^2\sum_{j=J_*}^{J}[\int_a^{a+b}\varphi_{skj}(z)\phi_{sk}(z)dz]^2 \leq C^*(J/\ln^2(n))^{-4\alpha}. \tag{A.18}$$

Similarly we establish that

$$[\partial\mu_{sk}/\partial\nu_{skj}]^2 \leq C^*s^2(J/\ln^2(n))^{-2\alpha}. \tag{A.19}$$

Now we continue Step 1 and would like to show that the parametric function class (A.15) is a subset of function classes (4) and (13). Let us check this. The second requirement (14) is valid due to the second line in (A.15). For all sufficiently large n, parametric functions h from (A.15) satisfy the second inequality in (5) as well as (15). The first relation in (14) is valid due to zero additive perturbations on the two boundary subintervals of $[a, a+b]$. We are left with verification of the first inequality in (5) and relation (16). We begin with (16).

Recall that $\alpha \ge 1$ and write,

$$\int_a^{a+b} [q^{(\alpha)}(x)]^2 dx = \int_a^{a+b} [g^{(\alpha)}(x|\vec{\nu}_s)]^2 dx = \sum_{k=1}^{s-2} \int_{a+kb/s}^{a+(k+1)b/s} [g^{(\alpha)}(x|\vec{\nu}_s)]^2 dx$$

$$= \sum_{k=1}^{s-2} \int_{a+kb/s}^{a+(k+1)b/s} [(\tilde{f}_{sk}(x|\vec{\nu}_{sk})\phi_{sk}(x))^{(\alpha)}]^2 dx, \tag{A.20}$$

and here we used our notation $\tilde{f}_{sk}(x|\vec{\nu}_{sk}) := f_{sk}(x|\vec{\nu}_{sk}) - \mu_{sk}$. Using Cauchy inequality we get

$$\int_{a+kb/s}^{a+(k+1)b/s} [(\tilde{f}_{sk}(x|\vec{\nu}_{sk})\phi_{sk}(x))^{(\alpha)}]^2 dx$$

$$\le (1+1/s) \int_{a+kb/s}^{a+(k+1)b/s} [\tilde{f}_{sk}^{(\alpha)}(x|\vec{\nu}_{sk})]^2 dx$$

$$+(1+s) \int_{a+kb/s}^{a+(k+1)b/s} [(\tilde{f}_{sk}(x|\vec{\nu}_{sk})(1-\phi_{sk}(x)))^{(\alpha)}]^2 dx =: A_{1k} + A_{2k}. \tag{A.21}$$

Using (A.14) and the Parseval identity we can evaluate A_{1k},

$$A_{1k} = (1+1/s) \sum_{i=J_*}^{J} (\pi i s/b)^{2\alpha} \nu_{ski}^2 \le (1+1/s) b Q_{sk}.$$

To evaluate A_{2k} we note that $\max_{0\le r\le\alpha} \int_a^{a+b} [(\phi_{sk}(x)-1)^{(r)}]^2 dx < C^*(s(\ln^2(n))^{2\alpha}$ and $[(\tilde{f}_{sk}(x|\vec{\nu}_{sk}))^{(\alpha-r)}]^2 = o_n^*(1) J^{-1/2}$ for $0 < r \le \alpha$. This, together with the Leibnitz rule of differentiation of the product of two functions, yield $A_{2k} = o_n^*(1) s^{-2}$. We conclude that for all sufficiently large n

$$\int_a^{a+b} [g^{(\alpha)}(x|\vec{\nu}_s)]^2 dx < bQ. \tag{A.22}$$

Using this in (A.20) verifies (16).

Lemma S.1 and (A.22) imply validity of the first inequality in (5). This finishes step 1 of the proposed proof, and from now on in the proof of the two lower bounds the original function classes can be replaced by their parametric subclass (A.15).

Step 2 is to convert the study of the minimax MISE into the study of a minimax sum of mean squared errors of parameters from the class (A.15). This step is usually straightforward

as, for instance, in Efromovich (2016), but not here due to the specific of function $g(x|\vec{\nu}_s)$ in (A.15). Here we proceed as follows. Recall that we are establishing a lower bound for a dealer-estimator $\check{h}$, which knows the anchor h_0, and hence we can set $\check{h} := h_0 + \check{g}$ and then write for hazards from the parametric class (A.15),

$$\int_a^{a+b} (\check{h}(x) - h(x))^2 dx$$

$$= \sum_{k=1}^{s-2} \int_{a+kb/s}^{a+(k+1)b/s} (\check{g}(x) - \tilde{f}_{sk}(x|\vec{\nu}_{sk})\phi_{sk}(x))^2 dx.$$

Using the Cauchy-Schwarz inequality, we can continue

$$\int_a^{a+b} (\check{h}(x) - h(x))^2 dx$$

$$\geq (1 - 1/s) \sum_{k=1}^{s-2} \int_{a+kb/s}^{a+(k+1)b/s} (\check{g}(x) - \tilde{f}_{sk}(x|\vec{\nu}_{sk}))^2 dx$$

$$-s \sum_{k=1}^{s-2} \int_{a+kb/s}^{a+(k+1)b/s} (\tilde{f}_{sk}(x|\vec{\nu}_{sk})(\phi_{sk}(x) - 1))^2 dx$$

$$=: A_3 - A_4. \tag{A.23}$$

Set $\check{\nu}_{skj} := \int_{a+kb/s}^{a+(k+1)b/s} \check{g}(x)\varphi_j(x)dx$. Using the Bessel inequality implies that

$$A_3 \geq (1 - 1/s) \sum_{k=1}^{s-2} \sum_{j=J_*}^{J} (\check{\nu}_{skj} - \nu_{skj})^2. \tag{A.24}$$

In the right side of (A.24) we have the wished sum of squared errors for the parameters. Let us show that A_4 is smaller in order than the verified lower bound, namely that $A_4 = o^*(1)n^{-2\alpha/(2\alpha+1)}$. Using (A.14), (A.18) and definition of the flattop kernel allows us to write,

$$A_4 \leq s \sum_{k=1}^{s-2} \int_{a+kb/s}^{a+(k+1)b/s} [\tilde{f}_{sk}(x|\vec{\nu}_{sk})(\phi_{sk}(x) - 1)]^2 dx$$

$$< s^2[4/(s\ln^2(n))][s^4\ln(n)n^{-2\alpha/(2\alpha+1)}] = o^*(1)[\ln(n)]^{-1/2}n^{-2\alpha/(2\alpha+1)}. \tag{A.25}$$

Using obtained bounds (A.24) and (A.25) in (A.23) we conclude that

$$\int_a^{a+b}(\check{h}(x)-h(x))^2dx$$

$$\geq (1-1/s)\sum_{k=1}^{s-2}\sum_{j=J_*}^{J}(\check{\nu}_{skj}-\nu_{skj})^2 - o^*(1)[\ln(n)]^{-1/2}n^{-2\alpha/(2\alpha+1)}. \tag{A.26}$$

We converted the nonparametric estimation under the ISE into estimation of a finite number of parameters under the SE. To finish step 2 we need to add appropriate minimax expressions in both sides of (A.26), and to do this we need a new notation. For each functional class $\mathcal{H}_{sk}$, defined in (A.14), let us introduce the corresponding set of vector-parameters $\vec{\nu}_{sk}$. Set $V_{sk} := \dot{V}_{sk} \cap \ddot{V}_{sk}$ where $\dot{V}_{sk} := \{\vec{\nu}_{sk} : \sum_{j=J_*}^{J}(\pi sj/b)^{2\alpha}\nu_{skj}^2 \leq bQ_{sk}\}$ and $\ddot{V}_{sk} := \{\vec{\nu}_{sk} : \max_{x\in[a,a+1]}|\sum_{j=J_*}^{J}\nu_{skj}\varphi_{skj}(x)| \leq [s^4\ln(n)]^{1/2}n^{-\alpha/(2\alpha+1)}\}$. For $\mathcal{H}_s$ the corresponding set of parameters is $V_s := \prod_{k=0}^{s-1}V_{sk} := V_{s0}\times\ldots\times V_{s(s-1)}$. In what follows we may also use negative subscripts to indicate that a specific part of a vector or a set is skipped. For instance, $V_{-sr} := \prod_{k\in\{0,\ldots,s-1\}\setminus r}V_{sk}$ and $\vec{\nu}_{-s2} := (\vec{\nu}_{s1},\vec{\nu}_{s3},\ldots,\vec{\nu}_{s(s-2)})$. Also, denote by $\vec{v}_{-skj}$ a vector $\vec{v}_s$ with its element v_{skj} being replaced by zero.

Using this notation and (A.26) we can write

$$\inf_{\check{h}^*}\sup_{h^{X^*}\in\tilde{\mathcal{S}}}\mathbb{E}_{h^{X^*}}\Big\{\int_a^{a+b}(\check{h}^*(x)-h^{X^*}(x))^2dx\Big\} \geq$$

$$\inf_{\check{h}^*}\sup_{h^{X^*}\in\mathcal{H}_s}\mathbb{E}_{h^{X^*}}\Big\{\int_a^{a+b}(\check{h}^*(x)-h^{X^*}(x))^2dx\Big\} \geq$$

$$\geq (1-s^{-1})\inf_{\vec{\check{\nu}}_s}\sup_{\vec{\nu}_s\in V_s}\sum_{k=1}^{s-2}\sum_{j=J_*}^{J}\mathbb{E}_{\vec{\nu}_s}\{(\check{\nu}_{skj}-\nu_{skj})^2\} + o_n^*(1)n^{-2\alpha/(2\alpha+1)}.$$

$$= (1-s^{-1})\sum_{k=1}^{s-2}\Big[\inf_{\vec{\check{\nu}}_{sk}}\sup_{\vec{\nu}_{sk}\in V_{sk}}\sum_{j=J_*}^{J}\mathbb{E}_{\vec{\nu}_s}\{(\check{\nu}_{skj}-\nu_{skj})^2\}\Big] + o_n^*(1)n^{-2\alpha/(2\alpha+1)}, \tag{A.27}$$

where $\tilde{\mathcal{S}}$ is either class (4) or (13). The last equality in (A.27) is valid because $\vec{\nu}_{sk}$ are

mutually unconstraint for different k (different subintervals). This finishes the second step of the proposed proof.

Step 3 is a traditional one when a minimax risk is replaced by a Bayes risk, and here we can restrict our attention to a particular kth subinterval. We need to introduce a prior for $\vec{\nu}_{sk}$, and this could be a simple problem if parameters in the vector were not mutually restricted. The latter is not the case according to definition (A.14) of the class $\mathcal{H}_{sk}$ (also recall the corresponding set V_{sk}). To take those restrictions on the parameters into account, the idea is to introduce independent normal random variables and then project them onto the set V_{sk}. This is how we will implement this approach. First, for each triplet (s, k, y) we define a Gaussian random variable ζ_{skj} which has zero mean and variance

$$\sigma^2_{skj} := n^{-1}(1 - 3s^{-1/2})\mathcal{I}^{-1}_{sk} \max(s^{-1/2}, \min(s^{1/2}, (J/j)^{\alpha} - 1)). \tag{A.28}$$

This choice takes into account the restrictions of class (A.14) because we can show that $\mathbb{P}(\vec{\zeta}_{sk} \in V_{sk}) = 1 + o^*_n(1)$. Indeed, recall our notation $V_{sk} = \dot{V}_{sk} \cap \ddot{V}_{sk}$, and note that $\mathbb{P}(\vec{\zeta}_{sk} \in \dot{V}_{sk}) = 1 + o^*_n(1)$ (the latter follows from Pinsker 1980), while Theorem 6.2.2 in Kahane (1985) yields $\mathbb{P}(\vec{\zeta}_{sk} \in \ddot{V}_{sk}) = 1 + o^*_n(1)$. We have proved that if we define a Gaussian stochastic process as a function from $\mathcal{H}_{sk}$ with ν_{skj} being replaced by random ζ_{skj}, then the probability for the Gaussian stochastic process to belong to the class $\mathcal{H}_{sk}$ tends to one. This sheds light on the choice of parameters (A.28), and we define the prior as the conditional distribution of $\vec{\zeta}_{sk}$ given $\vec{\zeta}_{sk} \in V_{sk}$.

Recall that a minimax risk is not smaller than a corresponding Bayes risk, and our proposed step 4 of the proof is to evaluate the Bayes risk. We begin with calculating the classical Fisher information for parameter ν_{skj}. Specific of the calculation is that we need to take into account the LTRC mechanism and that the dealer knows everything apart of an underlying vector-parameter $\vec{\nu}_s$. Recall that the Fisher information of interest is defined as

$$\mathcal{I}_{skj} := \mathbb{E}_{-\vec{\nu}_{skj}, h_0, F^{T^*}, G^{Z^*}}\{[\partial \ln(f^{Y,T,R}_{\vec{\nu}_s}(Y, T, R))/\nu_{skj}]^2\}, \ \vec{\nu}_s \in V_s. \tag{A.29}$$

The joint density of the observed triplet of random variables is

$$f^{Y,T,\Delta}_{\vec{\nu}_s}(y,t,\delta) = p^{-1}_{\vec{\nu}_s} f^{T^*}(t) I(t \le y)[f^{X^*}_{\vec{\nu}_s}(y) G^{Z^*}(y) I(\delta = 1) + G^{X^*}_{\vec{\nu}_s}(y) f^{Z^*}(y) I(\delta = 0)], \quad \text{(A.30)}$$

where

$$f^{X^*}_{\vec{\nu}_s}(y) = h_{\vec{\nu}_s}(y) e^{-\int_0^y h_{\vec{\nu}_s}(v)dv}, \quad \text{(A.31)}$$

$$p_{\vec{\nu}_s} = \int_0^\infty f^{T^*}(t) G^{X^*}_{\vec{\nu}_s}(t) G^{Z^*}(t) dt, \quad \text{(A.32)}$$

and

$$G^{X^*}_{\vec{\nu}_s}(t) = \int_t^\infty f^{X^*}_{\vec{\nu}_s}(x) dx. \quad \text{(A.33)}$$

Let us calculate the partial derivative of the density. Write for $0 \le t \le y < \infty$

$$\partial f^{Y,T,\Delta}_{\vec{\nu}_s}(y,t,\delta)/\partial \nu_{skj} \quad \text{(A.34)}$$

$$= [\partial p^{-1}_{\vec{\nu}_s}/\partial \nu_{skj}] f^{T^*}(t)[f^{X^*}_{\vec{\nu}_s}(y) G^{Z^*}(y) I(\delta = 1) + f^{Z^*}(y) G^{X^*}_{\vec{\nu}_s}(y) I(\delta = 0)]$$

$$+ p^{-1}_{\vec{\nu}_s}[\partial f^{X^*}_{\vec{\nu}_s}(y)/\partial \nu_{skj}] f^{T^*}(t) G^{Z^*}(y) I(\delta = 1)$$

$$+ p^{-1}_{\vec{\nu}_s}[\partial G^{X^*}_{\vec{\nu}_s}(y)/\partial \nu_{skj}] f^{T^*}(t) f^{Z^*}(y) I(\delta = 0)$$

$$=: A_1(y,t,\delta) + A_2(y,t,\delta) + A_3(y,t,\delta). \quad \text{(A.35)}$$

Now we are considering these three terms in turn. To evaluate A_1 we note that

$$\partial p^{-1}_{\vec{\nu}_s}/\partial \nu_{skj} = -p^{-2}_{\vec{\nu}_s} \partial \mathbb{P}_{\vec{\nu}_s}(Y^* \ge T^*)/\partial \nu_{skj}$$

$$= -p^{-2}_{\vec{\nu}_s} \int_0^\infty f^{T^*}(t) G^{Z^*}(t)[\partial G^{X^*}_{\vec{\nu}_s}(t)/\partial \nu_{skj}] dt. \quad \text{(A.36)}$$

Set $\mu'_{skj} := \partial \mu_{sk}/\partial \nu_{skj}$. Derivative of the survivor function can be written as follows,

$$\partial G^{X^*}_{\vec{\nu}_s}(t)/\partial \nu_{skj} = \partial e^{-\int_0^t h_{\vec{\nu}_s}(v)dv}/\partial \nu_{skj} = -\int_0^t [\partial h_{\vec{\nu}_s}(v)/\partial \nu_{skj}] dv G^{X^*}_{\vec{\nu}_s}(t)$$

$$= -[\int_0^t \varphi_{skj}(v)\phi_{sk}(v)dv]G^{X^*}_{\vec{\nu}_s}(t) + \mu'_{skj}[\int_0^t \phi_{sk}(v)dv]G^{X^*}_{\vec{\nu}_s}(t). \quad (A.37)$$

Consider the integral in the first term of right side of (A.37). First, let us mention several properties of rescaled flattop kernels. Function $\phi_{sk}(v)$ is zero beyond $(a + bk/s, a + b(k + 1)/s)$, it is equal to 1 on $[a + bk/s + 2b[s\ln^2(n)]^{-1}, a + b(k + 1)/s - 2b[s\ln^2(n)]^{-1}]$ and $|\phi^{(r)}_{sk}(v)| \leq C(s/b)^r(\ln(n))^{2r}$ for $r = 1, \ldots, \alpha$. Second, remember that we are considering $j > Cn^{1/(2\alpha+1)}/[s\ln(n)]$. Using these facts and integration by parts we can write,

$$|\int_0^t \varphi_{skj}(v)\phi_{sk}(v)dv| = |\int_a^{\min(t,a+b)} \varphi_{skj}(v)\phi_{sk}(v)dv|I(a \leq t)$$

$$= (\pi sj/b)^{-1}(2s/b)^{1/2}I(a \leq t)|\sin(\pi j[sb^{-1}(\min(t, a + b) - a) - k])\phi_{sk}(\min(t, a + b))$$

$$- \int_a^{\min(t,a+b)} \sin(\pi j[sb^{-1}(\min(t, a + b) - a) - k])(d\phi_{sk}(v)/dv)dv| = o^*_n(1)n^{-1/(2\alpha+2)}. \quad (A.38)$$

Using (A.38) and (A.19) in (A.37) we get

$$|\partial G^{X^*}_{\vec{\nu}_s}(t)/\partial\nu_{skj}| = o^*_n(1)n^{-1/(2\alpha+2)}. \quad (A.39)$$

This yields that

$$|\partial p^{-1}_{\vec{\nu}_s}/\partial\nu_{skj}| = o^*_n(1)n^{-1/(2\alpha+2)}p^{-2}_{\vec{\nu}_s}\int_0^\infty f^{T^*}(t)G^{Z^*}(t)dt = o^*_n(1)n^{-1/(2\alpha+2)}. \quad (A.40)$$

We conclude that the term $A_1(y, t, \delta)$ satisfies the following relation,

$$|\sum_{\delta=0}^{1}\int_{\{(t,y):\, 0\leq t\leq y<\infty\}}[A_1(y, t, \delta)]^2[f^{Y,T,\Delta}_{\vec{\nu}_s}(y, t, \delta)]^{-1}dydt = o^*_n(1)n^{-1/(2\alpha+2)}. \quad (A.41)$$

Now we are considering the second term $A_2(y, t, \delta)$ in (A.35). Write

$$\partial f^{X^*}_{\vec{\nu}_s}(y)/\partial\nu_{skj} = \partial[h_{\vec{\nu}_s}(y)e^{-\int_0^y h_{\vec{\nu}_s}(v)dv}]/\partial\nu_{skj}$$

$$= (\varphi_{skj}(y) - \mu'_{skj})\phi_{sk}(y)e^{-\int_0^y h_{\vec{\nu}_s}(v)dv} - \int_0^y (\partial h_{\vec{\nu}_s}(v)/\partial\nu_{skj})dv f^{X^*}_{\vec{\nu}_s}(y)$$

$$= \varphi_{skj}(y)\phi_{sk}(y)e^{-\int_0^y h_{\nu_s}(v)dv} - \int_0^y \varphi_{skj}(v)\phi_{sk}(v)dv f_{\vec{\nu}_s}^{X^*}(y) - \mu'_{skj}\phi_{sk}(y)e^{-\int_0^y h_{\nu_s}(v)dv}. \quad \text{(A.42)}$$

The last term is evaluated using (A.19). Using (A.38) we can evaluate the integral in the second term of (A.42),

$$|\int_0^y \varphi_{skj}(v)\phi_{sk}(v)dv| = o_n^*(1)n^{-1/(2\alpha+2)}. \quad \text{(A.43)}$$

For the first term in (A.42), using definitions of ϕ_{sk} and $A_2(y,t,\delta)$, we can write

$$\int_{\{(t,y):0\le t\le y<\infty\}} \frac{[p_{\vec{\nu}_s}^{-1} f^{T^*}(t)G^{Z^*}(y)\varphi_{skj}(y)\phi_{sk}(y)e^{-\int_0^y h_{\vec{\nu}_s}(v)dv}]^2}{p_{\vec{\nu}_s}^{-1} f^{T^*}(t)G^{Z^*}(y)f_{\vec{\nu}_s}^{X^*}(y)}dtdy$$

$$= p_{\vec{\nu}_s}^{-1}\int_{a+kb/s}^{a+(k+1)b/s} \varphi_{skj}^2(y)\phi_{sk}^2(y)G^{Z^*}(y)[h_{\vec{\nu}_s}(y)]^{-1}e^{-\int_0^y h_{\vec{\nu}_s}(v)dv}[\int_0^y f^{T^*}(t)dt]dy$$

$$= p_{\vec{\nu}_s}^{-1}G^{Z^*}(a+kb/s)[h_{\vec{\nu}_s}^{X^*}(a+kb/s)]^{-1}e^{-\int_0^{a+kb/s} h_{\vec{\nu}_s}^{X^*}(v)dv}F^{T^*}(a+kb/s)(1+o_n^*(1)).$$

This is the main term in the evaluated Fisher information $\mathcal{I}_{skj}$, and one can compare it with defined in the beginning of the proof quantity $\mathcal{I}_{sk}$. Finally, using (26) we evaluate $A_3(y,t,\delta)$ and conclude that

$$\mathbb{E}_{-\vec{\nu}_{skj},h_0,F^{T^*},G^{Z^*}}\{A_3^2(Y,T,\Delta)[f_{\vec{\nu}_{-skj}}^{Y,T,\Delta}(Y,T,\Delta)]^{-2}dydt\} = o_n^*(1). \quad \text{(A.44)}$$

Combining obtained results we get an important relation which explains our choice of used sequences,

$$\mathcal{I}_{skj} = \mathcal{I}_{sk}(1+o_n^*(1)). \quad \text{(A.45)}$$

Now we have arrived at the previously explained point where we can begin using some technical results from the proof in Efromovich (2016). The technical results begin with line (45) in that proof. We need to make one comment about notation. In that paper b and q denote specific parameters not related to our parameters, and for our purposes we can set $q = b = s^{1/2}$. Then in line (81) of Efromovich (2016) the factor $[1+o_n^*(1)+o_q^*(1)+o_b^*(1)]$

can be replaced by $[1+o_n^*(1)]$. This allows us to write,

$$\inf_{\vec{\nu}_s}\sup_{\vec{\nu}_s\in V_s}\sum_{k=1}^{s-2}\sum_{j=J_*}^{J}\mathbb{E}_{\vec{\nu}_s}\{(\check{\nu}_{skj}-\nu_{skj})^2\}\geq\Big[\sum_{k=1}^{s-2}\sum_{j=J_*}^{J}\frac{(\mathcal{I}_{sk}n)^{-1}\sigma_{skj}^2}{(\mathcal{I}_{sk}n)^{-1}+\sigma_{skj}^2}\Big](1+o_n^*(1)). \tag{A.46}$$

Note that a particular term in the sum (A.46) is a familiar variance of the Bayes estimator for the mean of a normal random variable with a conjugate normal prior, that is, we indeed were able to evaluate from below the minimax MSE by the classical Bayes one. Then a direct calculation yields that

$$\begin{aligned}\sum_{j=J_*}^{J}\frac{(\mathcal{I}_{sk}n)^{-1}\sigma_{skj}^2}{(\mathcal{I}_{sk}n)^{-1}+\sigma_{skj}^2}&\geq(\mathcal{I}_{sk}n)^{-1}\sum_{J_*}^{J}[1-(j/J)^{\alpha}](1+o_n^*(1))\\&=(\mathcal{I}_{sk}n)^{-1}J(1-(1+\alpha)^{-1})(1+o_n^*(1))\\&=\mathcal{I}_{sk}^{-1}[(b/s)^{2\alpha}\mathcal{I}_s]^{1/(2\alpha+1)}P(\alpha,bQ,1)n^{-2\alpha/(2\alpha+1)}(1+o_n^*(1)).\end{aligned} \tag{A.47}$$

Using this inequality in the right side of (A.46), we continue evaluation of the left side of (A.46),

$$\begin{aligned}&\inf_{\vec{\nu}_s}\sup_{\vec{\nu}_s\in V_s}\sum_{k=1}^{s-2}\sum_{j=J_*}^{J}\mathbb{E}_{\vec{\nu}_s}\{(\check{\nu}_{skj}-\nu_{skj})^2\}\\&\geq\sum_{k=1}^{s-2}\mathcal{I}_{sk}^{-1}[(b/s)^{2\alpha}\mathcal{I}_s]^{1/(2\alpha+1)}P(\alpha,bQ,1)n^{-2\alpha/(2\alpha+1)}(1+o_n^*(1))\\&=\mathcal{I}_s^{-1}[(b/s)^{2\alpha}\mathcal{I}_s]^{1/(2\alpha+1)}P(\alpha,bQ,1)n^{-2\alpha/(2\alpha+1)}(1+o_n^*(1))\\&=[(b/s)\mathcal{I}_s^{-1}]^{-2\alpha/(2\alpha+1)}P(\alpha,bQ,1)n^{-2\alpha/(2\alpha+1)}(1+o_n^*(1)).\end{aligned} \tag{A.48}$$

Note that

$$[(b/s)\mathcal{I}_s^{-1}]^{2\alpha/(2\alpha+1)}=[bd]^{2\alpha/(2\alpha+1)}(1+o_n^*(1)), \tag{A.49}$$

where d is defined in (11). Using (A.49) in (A.48), and then the obtained inequality in

(A.27), we conclude that

$$\inf_{\check{h}^*} \sup_{h^{X^*}\in\tilde{\mathcal{S}}} \mathbb{E}_{h^{X^*}}\Big\{\int_a^{a+b}(\check{h}^*(x)-h^{X^*}(x))^2dx\Big\} \geq [d/n]^{2\alpha/(2\alpha+1)}P(\alpha,Q,b)(1+o_n^*(1)). \quad \text{(A.50)}$$

Remember that $\tilde{S}$ denotes either the class (4) or (13), and inequality (A.50) is valid for both these classes. This verifies validity of the dealer's lower bounds of Theorems 1 and 2. Note that they are the same for the two classes, and this is due to the fact we used the same parametric class (A.15) which is the subset of both (4) and (13). What was wished to show. □

Proof of Theorem 3. The proof is written in such a way that it can be also used to verify assertion of Theorem 4. The proof also uses some technical results of the previous proofs so we continue to use the same notation. We begin with some preliminary calculations. Recall that J^* in (59) is an arbitrary integer number, and then we can write for the first term in (59),

$$\sum_{k=1}^{K_n}\sum_{j\in B_k}[n^{-1}d^*\lambda_j^2+(1-\lambda_j)^2\theta_j^2]$$

$$=\sum_{k=1}^{K_n}L_k[n^{-1}d^*\mu_k^2+(1-\mu_k)^2\Theta_k]+[n^{-1}d^*\sum_{k=1}^{K_n}\sum_{j\in B_k}(\lambda_j^2-\mu_k^2)$$

$$+\sum_{k=1}^{K_n}\sum_{j\in B_k}(2-\mu_k-\lambda_j)(\mu_k-\lambda_j)\theta_j^2]$$

$$=: A+B, \quad \text{where } \mu_k := \max_{j\in B_k}\lambda_j, \quad \text{(A.51)}$$

and recall 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 (A.51). 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

$$A\geq n^{-1}d^*\sum_{k=1}^{K_n}L_k\Theta_k[\Theta_k+n^{-1}d^*]^{-1}. \quad \text{(A.52)}$$

To evaluate from below the term B in (A.51), we note that in B the second sum is nonnegative. For the first sum, which is nonpositive, let us additionally assume that

$$\mu_{k+1} \leq \min_{j \in B_k} \lambda_j, \;\; k = 1, \ldots, K_n - 1. \tag{A.53}$$

(Note that (A.53), as well as the earlier made assumptions about $\{\lambda_j\}$, hold for the shrinkage coefficients used by the linear estimate, and this is the reason why we add these restrictions). Write,

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

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

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

$$\leq (1 + o_n^*(1)) \sum_{k=1}^{K_n} \sum_{j \in B_k} \lambda_j^2 + C^* \ln(n). \tag{A.54}$$

Using (A.54) we conclude that given (A.53) the term B in (A.51) can be bounded from below,

$$B \geq -|o_n^*(1)| n^{-1} d^* \sum_{k=1}^{K_n} \sum_{j \in B_k} \lambda_j^2 - C^* n^{-1} \ln(n). \tag{A.55}$$

Using (A.52) and (A.55) in (A.51), and then using the obtained inequality in (59) we conclude that given (A.53) the following inequality holds uniformly over all considered sets $\{\lambda_j\}$,

$$n^{-1} d^* \sum_{k=1}^{K_n} L_k \frac{\Theta_k}{\Theta_k + d^* n^{-1}}$$

$$\leq (1 + o_n^*(1)) E_{h^{X^*}} \{ \int_a^{a+b} (\bar{h}(x, \{\lambda_j\}) - h^{X^*}(x))^2 dx \} + C^* n^{-1} \ln(n). \tag{A.56}$$

Now we are introducing an oracle-estimator

$$\tilde{h}^*(x,g,h^{X^*}) := \sum_{k=1}^{K_n} \Theta_k[\Theta_k + d^*n^{-1}]^{-1} \sum_{j\in B_k} \hat{\theta}_j \varphi_j(x).$$

The oracle uses an underlying hazard rate for optimal blockwise smoothing. As we shall see shortly, this oracle is sharp minimax for the considered setting, and the oracle motivates the proposed data-driven estimator.

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 $\sum_{k=1}^{K_n}\sum_{j\in B_k}\lambda_j^2 \to \infty$ as $n\to\infty$ because otherwise the hazard rate on $[a,a+b]$ is parametric and it is estimated by the oracle-estimator with the MISE proportional to n^{-1}. Using (59), together with the remark made below that line, we can write,

$$|\mathbb{E}_{h^{X^*}}\{\int_a^{a+b}(\tilde{h}^*(x,g,h^{X^*}) - h^{X^*}(x))^2dx\} - [n^{-1}d^*\sum_{k=1}^{K_n}\frac{L_k\Theta_k}{\Theta_k + d^*n^{-1}} + \sum_{j>J^*}\theta_j^2]|$$

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

$$+Cn^{-1}I\Big(\sum_{k=1}^{K_n} L_k\Big[\frac{\Theta_k}{\Theta_k + d^*n^{-1}}\Big]^2 < \ln(\ln(n))\Big). \tag{A.57}$$

Set $\lambda_j^* := [1-(j/J)^\alpha]I(j<J)$. A direct calculation shows that

$$\sup_{h^{X^*}\in\mathcal{S}(\alpha,Q,h_0,a,b)} \mathbb{E}_{h^{X^*}}\{\int_a^{a+b}(\bar{h}(x,\{\lambda_j^*\}) - h^{X^*}(x))^2dx\}$$

$$\le P(\alpha,Q,b)[dn^{-1}]^{2\alpha/(2\alpha+1)}(1+o_n(1)). \tag{A.58}$$

This inequality, together with (A.56) and (A.57), proves sharp-minimaxity of the oracle-estimator $\tilde{h}^*(x,g,h^{X^*})$.

Now we are going to show that the proposed data-driven estimator matches MISE of the oracle-estimator. Using the Cauchy inequality we can bound the estimator's MISE via the

oracle's MISE

$$\mathbb{E}_{h^{X^*}}\{\int_a^{a+b}(\hat{h}(x)-h^{X^*}(x))^2dx\}$$

$$\le \mathbb{E}_{h^{X^*}}\{\int_a^{a+b}(\tilde{h}^*(x,g,h^{X^*})-h^{X^*}(x))^2dx\}(1+\gamma)$$

$$+\mathbb{E}_{h^{X^*}}\{\int_a^{a+b}(\tilde{h}^*(x,g,h^{X^*})-\hat{h}(x))^2dx\}(1+\gamma^{-1}),\ \gamma>0. \tag{A.59}$$

The MISE of oracle-estimator $\tilde{h}^*(x,g,h^{X^*})$ is already evaluated, and we are considering the second expectation in (A.59) using the Parseval identity,

$$\mathbb{E}_{h^{X^*}}\{\int_a^{a+b}(\tilde{h}^*(x,g,h^{X^*})-\hat{h}(x))^2dx\} = \mathbb{E}_{h^{X^*}}\Big\{\sum_{k=1}^{K_n}[\frac{\Theta_k}{\Theta_k+d^*n^{-1}}$$

$$-\frac{L_k^{-1}\sum_{j\in B_k}\hat{\theta}_j^2-\hat{d}n^{-1}}{L_k^{-1}\sum_{j\in B_k}\hat{\theta}_j^2}I(L_k^{-1}\sum_{j\in B_k}\hat{\theta}_j^2>(\hat{d}+1/\ln(n))n^{-1})]^2\sum_{j\in B_k}\hat{\theta}_j^2\Big\}. \tag{A.60}$$

Let us consider a particular $k\in\{1,\dots,K_n\}$ in the sum, set $\hat{\Theta}_k := L_k^{-1}\sum_{j\in B_k}\hat{\theta}_j^2 - d^*n^{-1}$ (note that this is not a statistic due to using d^*), and write

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

$$=\frac{n^{-2}L_k[d^*(\Theta_k-\hat{\Theta}_k)+(\hat{d}-d^*)(\Theta_k+d^*n^{-1})]^2}{(\Theta_k+d^*n^{-1})^2(\hat{\Theta}_k+d^*n^{-1})}I(\hat{\Theta}_k>(\hat{d}-d^*+1/\ln(n))n^{-1})$$

$$+\frac{\Theta_k^2L_k(\hat{\Theta}_k+d^*n^{-1})}{(\Theta_k+d^*n^{-1})^2}I(\hat{\Theta}_k\le(\hat{d}-d^*+1/\ln(n))n^{-1}) := A_{k1}+A_{k2}. \tag{A.61}$$

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

$$A_{k1}\le\frac{2n^{-2}L_k[(d^*(\hat{\Theta}_k-\Theta_k))^2+(\hat{d}-d^*)^2(\Theta_k+d^*n^{-1})^2]}{(\Theta_k+d^*n^{-1})^2(\hat{\Theta}_k+d^*n^{-1})}$$

$$\times I(\hat{\Theta}_k>(\hat{d}-d^*+1/\ln(n))n^{-1}). \tag{A.62}$$

Now we need two directly verified inequalities which are established following the proof of

(55),

$$\mathbb{E}_{h^{X^*}}\{(\hat{d}-d^*)^2\} \leq C^* n^{-1}, \quad \mathbb{E}_{h^{X^*}}\{(\hat{\Theta}_k - \Theta_k)^2\} \leq C^* L_k^{-1} n^{-1}(\Theta_k + n^{-1}), \quad k \in \{1, \ldots, K_n\}, \tag{A.63}$$

where here and in what follows generic constants C^*s are uniformly bounded for all considered hazard rates h^{X^*} 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

$$\mathbb{E}_{h^{X^*}}\{\frac{n^{-2} L_k (d^*)^2 (\hat{\Theta}_k - \Theta_k)^2}{(\Theta_k + d^* n^{-1})^2 (\hat{\Theta}_k + d^* n^{-1})} I(\hat{\Theta}_k > (\hat{d} - d^* + 1/\ln(n))n^{-1})\} \leq C^* \ln(n) n^{-1},$$

and

$$\mathbb{E}_{h^{X^*}}\{\frac{n^{-2} L_k (\hat{d} - d^*)^2}{(\hat{\Theta}_k + d^* n^{-1})} I(\hat{\Theta}_k > (\hat{d} - d^* + 1/\ln(n))n^{-1})\} \leq C^* \ln(n) n^{-2} L_k.$$

Using the last two inequalities and (A.62) imply that for $k \in \{1, \ldots, K_n\}$

$$\mathbb{E}_{h^{X^*}}\{A_{k1}\} \leq C^* \ln(n) n^{-1}. \tag{A.64}$$

Now let us evaluate $\mathbb{E}_{h^{X^*}}\{A_{k2}\}$. Write,

$$\mathbb{E}_{h^{X^*}}\{A_{k2}\} = \frac{L_k \Theta_k^2}{(\Theta_k + d^* n^{-1})^2} \mathbb{E}_{h^{X^*}}\{(\hat{\Theta}_k + d^* n^{-1}) I(\hat{\Theta}_k \leq (\hat{d} - d^* + 1/\ln(n))n^{-1})\}$$

$$=: \frac{L_k \Theta_k^2}{(\Theta_k + d^* n^{-1})^2} D_k. \tag{A.65}$$

To evaluate D_k we can write,

$$D_k = \mathbb{E}_{h^{X^*}}\{(\hat{\Theta}_k + d^* n^{-1}) I(\hat{\Theta}_k \leq (\hat{d} - d^* + 1/\ln(n))n^{-1})\}$$

$$= \mathbb{E}_{h^{X^*}}\{(\hat{\Theta}_k + d^* n^{-1})[I(|\hat{d} - d^*| < 1/\ln(n)) + I(|\hat{d} - d^*| \geq 1/\ln(n))]$$

$$\times I(\hat{\Theta}_k \leq (\hat{d} - d^* + 1/\ln(n))n^{-1})\}.$$

Using (A.63) and the Chebyshev inequality we can continue,

$$D_k \le C^* n^{-1} \mathbb{E}_{h^{X^*}}\{I(\hat{\Theta}_k \le 2n^{-1}/\ln(n))\}$$

$$+\mathbb{E}_{h^{X^*}}\{(|\hat{d} - d^*| + 1/\ln(n) + d^*)n^{-1} I(|\hat{d} - d^*| \ge 1/\ln(n))\}$$

$$\le C^* n^{-1} \mathbb{E}_{h^{X^*}}\{I(\hat{\Theta}_k \le 2n^{-1}/\ln(n))\}[I(\Theta_k < 4n^{-1}/\ln(n))$$

$$+I(\Theta_k \ge 4n^{-1}/\ln(n))] + C^* n^{-2}\ln^2(n).$$

Note that

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

Using this, with the help of (A.63) and the Chebyshev inequality, we continue evaluation of D_k,

$$D_k \le C^* n^{-1} I(\Theta_k < 4n^{-1}/\ln(n))$$

$$+C^* n^{-1} L_k^{-1/2} n^{-1/2}(\Theta_k + n^{-1})^{1/2}\Theta_k^{-1} I(\Theta_k \ge 4n^{-1}/\ln(n)) + C^* n^{-2}\ln^2(n).$$

Using the last inequality in (A.65) we get

$$E_{h^{X^*}}\{A_{k2}\} \le \frac{C^* L_k \Theta_k n^{-1}}{\Theta_k + d^* n^{-1}}\Big[\frac{\Theta_k}{\Theta_k + d^* n^{-1}} I(\Theta_k < 4n^{-1}/\ln(n))$$

$$+\frac{\Theta_k L_k^{-1/2} n^{-1/2}(\Theta_k + n^{-1})^{1/2}}{(\Theta_k + d^* n^{-1})\Theta_k} I(\Theta_k \ge 4n^{-1}/\ln(n)) + C^* n^{-1}\ln^2(n)\Big]$$

$$\le \frac{C^* L_k \Theta_k n^{-1}}{\Theta_k + d^* n^{-1}}$$

$$\times\Big[\ln^{-1}(n) I(\Theta_k < 4n^{-1}/\ln(n)) + L_k^{-1/2} I(\Theta_k > 4n^{-1}/\ln(n)) + C^* n^{-1}\ln^2(n)\Big]. \qquad \text{(A.66)}$$

In (A.66) let us look at the effect of the second term in the square brackets (it contains factor $L_k^{-1/2}$). Set $t := \lfloor 3\ln(n)(\ln\ln(n))^2 \rfloor$, note that for all sufficiently 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, $L_t > [1 + \ln^{-1}(n)/\ln(\ln(n))]^t > C^* e^{3\ln(\ln(n))} = C^*\ln^3(n)$. We conclude that

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

$$\leq n^{-1}C^*\ln^{9/2}(n) + C^*\ln^{-3/2}(n)[n^{-1}\sum_{k=1}^{K_n}\frac{L_k\Theta_k}{\Theta_k + d^* n^{-1}} I(\Theta_k > 4n^{-1}/\ln(n)]. \tag{A.67}$$

Using (A.67) in (A.66) we conclude that

$$\mathbb{E}_{h^{X^*}}\{\sum_{k=1}^{K_n} A_{k2}\} \leq C^*\sum_{k=1}^{K_n}\frac{L_k n^{-1}\Theta_k}{\Theta_k + d^* n^{-1}}\ln^{-1}(n) + C^* n^{-1}\ln^{9/2}(n). \tag{A.68}$$

Now we can return to (A.59). With the help of (A.60), (A.61), (A.64) and (A.68) we conclude that

$$\mathbb{E}_{h^{X^*}}\{\int_a^{a+b}(\hat{h}(x) - h^{X^*}(x))^2 dx\} \leq \mathbb{E}_{h^{X^*}}\{\int_a^{a+b}(\tilde{h}^*(x, g, h^{X^*}) - h^{X^*}(x))^2 dx\}(1+\gamma)$$

$$+\Big[C^*\sum_{k=1}^{K_n}\frac{L_k n^{-1}\Theta_k}{\Theta_k + d^* n^{-1}}\ln^{-1}(n) + C^* n^{-1}\ln^{9/2}(n)\Big](1+\gamma^{-1}). \tag{A.69}$$

Set $\gamma = \ln^{-1/2}(n)$, and continue (A.69) using (A.57),

$$\mathbb{E}_{h^{X^*}}\{\int_a^{a+b}(\hat{h}(x) - h^{X^*}(x))^2 dx\}$$

$$\le \mathbb{E}_{h^{X^*}}\{\int_a^{a+b}(\tilde{h}^*(x,g,h^{X^*}) - h^{X^*}(x))^2 dx\}[1 + C^* \ln^{-1/2}(n)] + C^* n^{-1}\ln^5(n). \qquad \text{(A.70)}$$

What we see in (A.70) is the so-called oracle-inequality which relates the MISEs of the estimate and the oracle-estimate. Now remember that constants C^*s do not depend on h^{X^*} and n, and then assertion of Theorem 3 follows from (A.58) and (A.70). □

Proof of Theorem 4. Consider moments of a variable $V_j := \Delta\varphi_j(Y)I(Y \in [a, a+b])g^{-1}(Y)$, where $g(v) := \mathbb{P}_{h^{X^*}}(T \le v \le Y)$. Recall the assumption $\inf_{v\in[a,a+b]} g(v) > C_* > 0$. The first moment of V is

$$\mathbb{E}_{h^{X^*}}\{V_j\} = \mathbb{E}_{h^{X^*}}\{\Delta\varphi_j(Y)I(Y \in [a, a+b])g^{-1}(Y)\}$$

$$= \int_a^{a+b} h^{X^*}(y)g(y)\varphi_j(y)g^{-1}(y)]dy = \int_a^{a+b} h^{X^*}(y)\varphi_j(x)dx = \theta_j. \qquad \text{(A.71)}$$

We conclude that V_j is unbiased estimate of the jth Fourier coefficient whenever $g(v)$ is known (otherwise it is an oracle-estimate). For the second moment we can write,

$$\mathbb{E}_{h^{X^*}}\{V_j^2\} = \int_a^{a+b} h^{X^*}(y)g(y)\varphi_j^2(y)g^{-2}(y)dy$$

$$\le (2/b)\int_a^{a+b} h^{X^*}(y)g^{-1}(y)dy \le C^* < \infty. \qquad \text{(A.72)}$$

These two results allow us to follow along lines of the proof of Theorem 3 (recall that it was explained that the proof was written in such a way that it could be used here) and establish rate-minimaxity of the estimate (19). Theorem 4 is proved. □

Acknowledgements The research is supported by NSF Grant DMS-1513461 and CAS-2016 Grant. Suggestions of reviewers and the Associate Editor are greatly appreciated.

References

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

Hagar,Y. and Dukic, V. (2015). Comparison of hazard rate estimation in R. arXiv: 1509.03253v1

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

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

Marron, J.S. and Wand, M.P. (1992). Exact mean integrated squared error. *Annals of Statistics, 20*, 712-736.

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

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

Wasserman, L. (2005). *All of Nonparametric Statistics.* New York: Springer.

APPENDIX B. Useful Formulae for Understanding the LTRC Model

Let us present formulae that shed light on relationship between the hidden triplet (T^*, X^*, Z^*) and, given $T^* \le Y^* := \min(X^*, Z^*)$, the observed statistic (T, Y, Δ) where $T := T^*$, $Y := Y^*$, and $\Delta := \Delta^* := I(X^* \le Z^*)$. Recall that $p := \mathbb{P}(T^* \le \min(X^*, Z^*))$ is defined in (2), and then we can write for $(t, y) \in \{(u, v) : \ u \in [0, \infty), v \in [u, \infty)\}$ and $r \in \{0, 1\}$ that the joint cumulative distribution function is

$$F^{T,Y,\Delta}(t, y, \delta) = F^{T^*,Y^*,\Delta^*|T^* \le Y^*}(t, y, \delta) = p^{-1}\mathbb{P}(T^* \le t, T^* \le Y^* \le y, \Delta^* = \delta)$$

$$= p^{-1} \int_0^t f^{T^*}(\tau)[\int_\tau^{\max(\tau,y)} f^{X^*}(x)G^{Z^*}(x)dxI(\delta = 1) + \int_\tau^{\max(\tau,y)} f^{Z^*}(z)G^{X^*}(z)dzI(\delta = 0)]d\tau. \tag{B.1}$$

Via direct differentiation, this formula for the cumulative distribution function yields the following formula for the joint probability density,

$$f^{T,Y,\Delta}(t, y, \delta) = p^{-1} f^{T^*}(t)I(t \le y)[f^{X^*}(y)G^{Z^*}(y)I(\delta = 1) + f^{Z^*}(y)G^{X^*}(y)I(\delta = 0)]. \tag{B.2}$$

Next, we have

$$F^{Y,\Delta=1}(y) = F^{X^*,X^* \le Z^*|T^* \le \min(X^*,Z^*)}(y) = \mathbb{P}(X^* \le y, X^* \le Z^*|T^* \le \min(X^*, Z^*))$$

$$= p^{-1}\mathbb{P}(X^* \le y, X^* \le Z^*, T^* \le X^*) = p^{-1} \int_0^y f^{X^*}(x)G^{Z^*}(x)F^{T^*}(x)dx, \ \ y \ge 0. \tag{B.3}$$

This yields the corresponding probability density

$$f^{Y,\Delta=1}(y) = p^{-1} f^{X^*}(y)G^{Z^*}(y)F^{T^*}(y)I(y \ge 0)$$

$$= h^{X^*}(y)[p^{-1}G^{Z^*}(y)F^{T^*}(y)G^{X^*}(y)I(y \ge 0)]. \tag{B.4}$$

In its turn, for x such that $G^{Z^*}(x)F^{T^*}(x) > 0$, the last relation yields for the density of the

random variable of interest X^* the expression

$$f^{X^*}(x) = \frac{f^{Y,\Delta=1}(x)}{p^{-1}G^{Z^*}(x)F^{T^*}(x)} \quad \text{whenever } G^{Z^*}(x)F^{T^*}(x) > 0. \tag{B.5}$$

Now we focus on the denominator in (B.5). Write for $v \in (0, \infty)$,

$$\mathbb{P}(T \leq v \leq Y) = \mathbb{P}(T^* \leq v \leq Y^* | T^* \leq Y^*) = [p^{-1}G^{Z^*}(v)F^{T^*}(v)]G^{X^*}(v) =: g(v). \tag{B.6}$$

Note that the right side of (B.6) contains, as a factor in the square brackets, the denominator of the ratio in (B.5). This yields the following expression for the probability density of X^*,

$$f^{X^*}(x) = \frac{f^{Y,\Delta=1}(x)G^{X^*}(x)}{\mathbb{P}(T \leq x \leq Y)} \quad \text{whenever } G^{Z^*}(x)F^{T^*}(x) > 0. \tag{B.7}$$

The last formula points upon a feasible estimator of the density, but we can also realize that another unique characteristic of the distribution, the hazard rate function, has a nice expression via estimable functions,

$$h^{X^*}(x) = \frac{f^{Y,\Delta=1}(x)}{\mathbb{P}(T \leq x \leq Y)} \quad \text{whenever } G^{Z^*}(x)F^{T^*}(x) > 0. \tag{B.8}$$

Finally, because the hazard rate is estimable, we may use the following familiar formula that relates the density, survival function and the hazard rate function, $f^{X^*}(x) := h^{X^*}(x)e^{-\int_0^x h^{X^*}(v)dv}$ and $G^{X^*}(x) = e^{-\int_0^x h^{X^*}(v)dv}$.

Conclusions from the formulae: (i) For estimation, if $\theta_j = \int_a^{a+b} h^{X^*}(x)\varphi_j(x)dx$ then its reasonable estimator is defined in (20)-(21). The underlying idea of η_l, defined in (20), is that $n^{-1}\sum_{s=1}^{n} I(T_s \leq v \leq Y_s)$ is the sample mean estimator of the probability $\mathbb{P}(T \leq v \leq Y)$.
(ii) For establishing a lower bound, the following relation is useful,

$$f^{T,Y,\Delta}(t,y,\delta) = f^{T^*,Y^*,\Delta^*|T^*\leq Y^*}(t,y,\delta)$$

$$= p^{-1}I(t \leq y)f^{X^*}(y)f^{T^*}(t)G^{Z^*}(y)I(\delta = 1) + p^{-1}I(t \leq y)f^{Z^*}(y)f^{T^*}(t)G^{X^*}(y)I(\delta = 0).$$

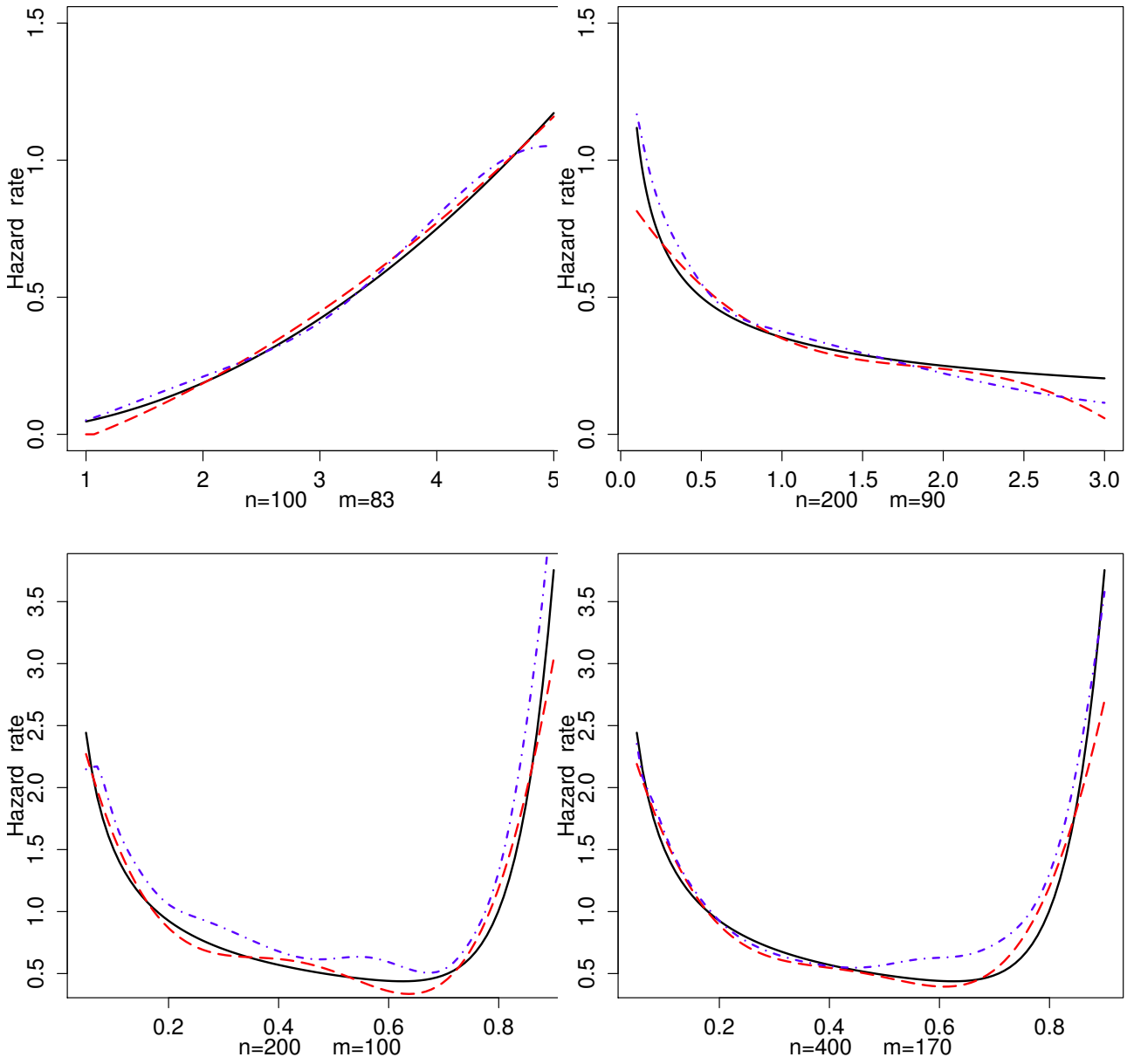

Figure 1: Estimation of the hazard rate function (solid line) by oracle kernel estimate (dot-dash line) and proposed estimate (dash line). Top left, top right and two bottom diagrams correspond to the third, sixth and eleventh experiments described in Section 2. Subtitles show the total sample size n and the number m of Ys observed within an interval of estimation.

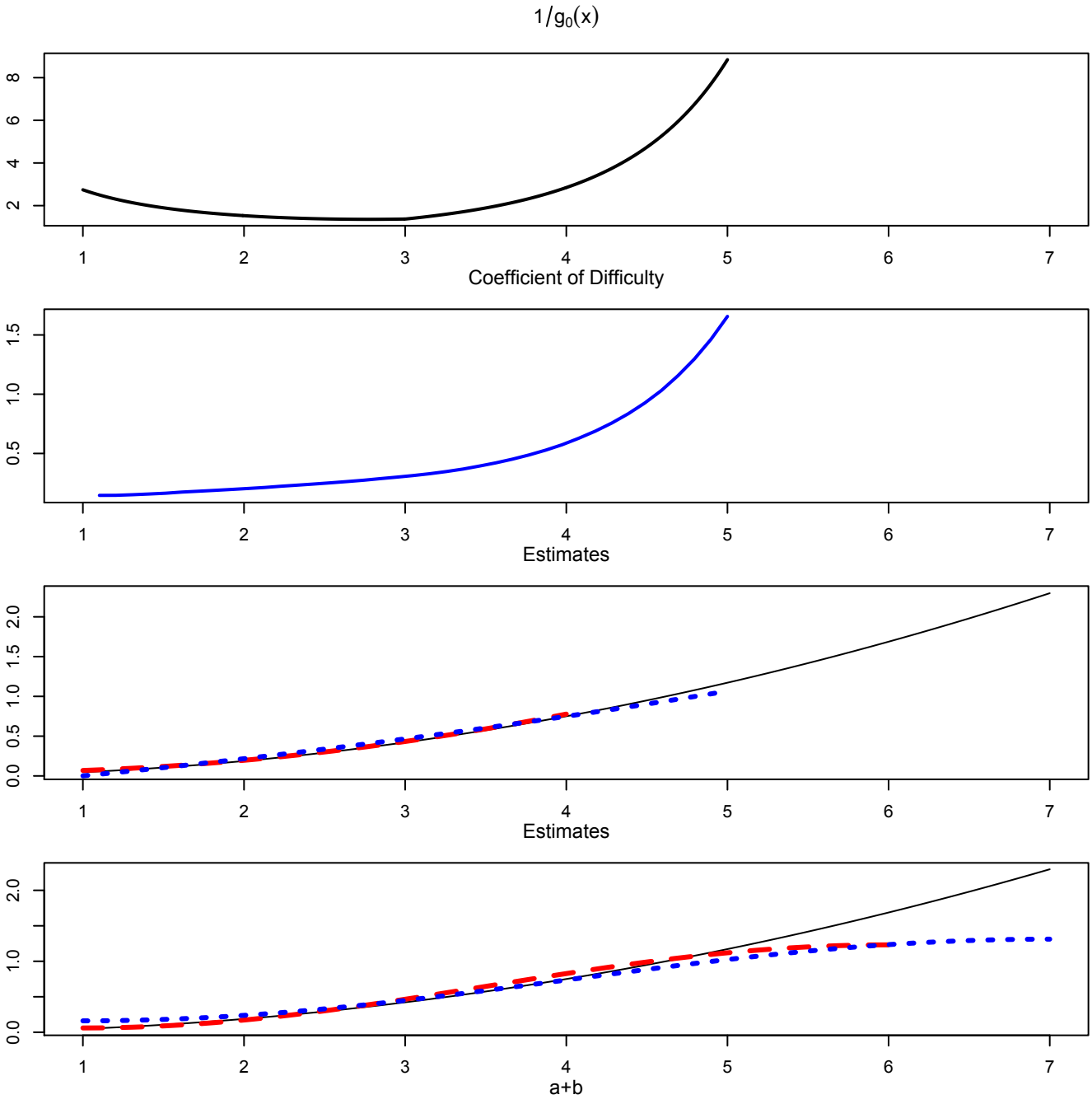

Figure 2: Effect of the interval of estimation $[a, a+b]$. Experiment 3 of Section 2 is considered with $a = 1$ and $x = a+b \geq 1$.Two top diagrams show functions only for $x \in [1, 5]$ because they increase rapidly beyond the interval. In particular, for values 6 and 7, the function $1/g_0(x)$ takes on values 46 and 447 while the coefficient of difficulty is 8 and 66, respectively. Two bottom diagrams show the underlying hazard rate by the solid line, and estimates are shown by the long-dashed and short-dashed lines over corresponding intervals with $x = 4, 5, 6, 7$.

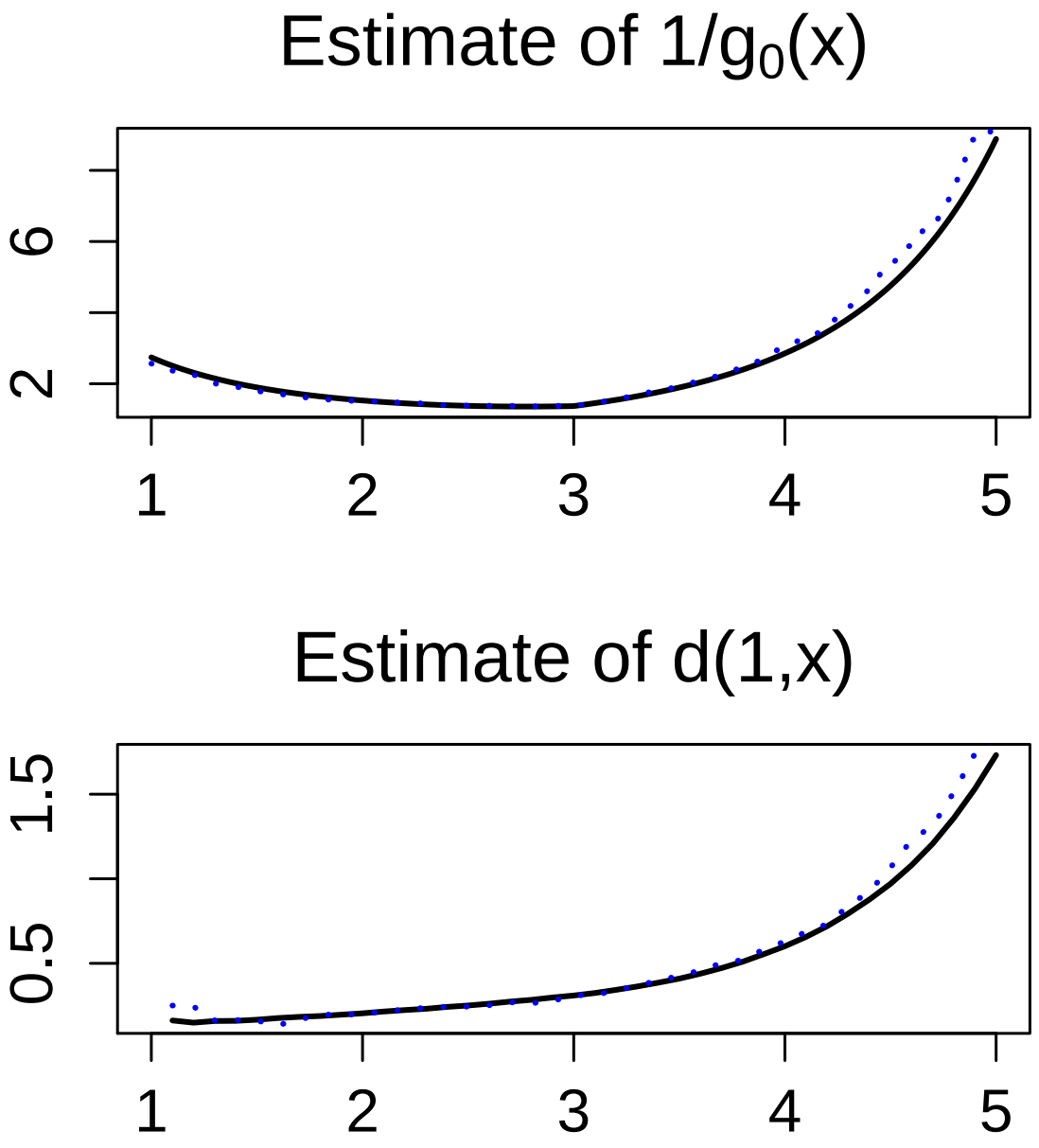

Figure 3: Estimates of $1/g_0(x)$ and coefficient of difficulty $d(1, x)$. The sample is the same as in Figure 2.

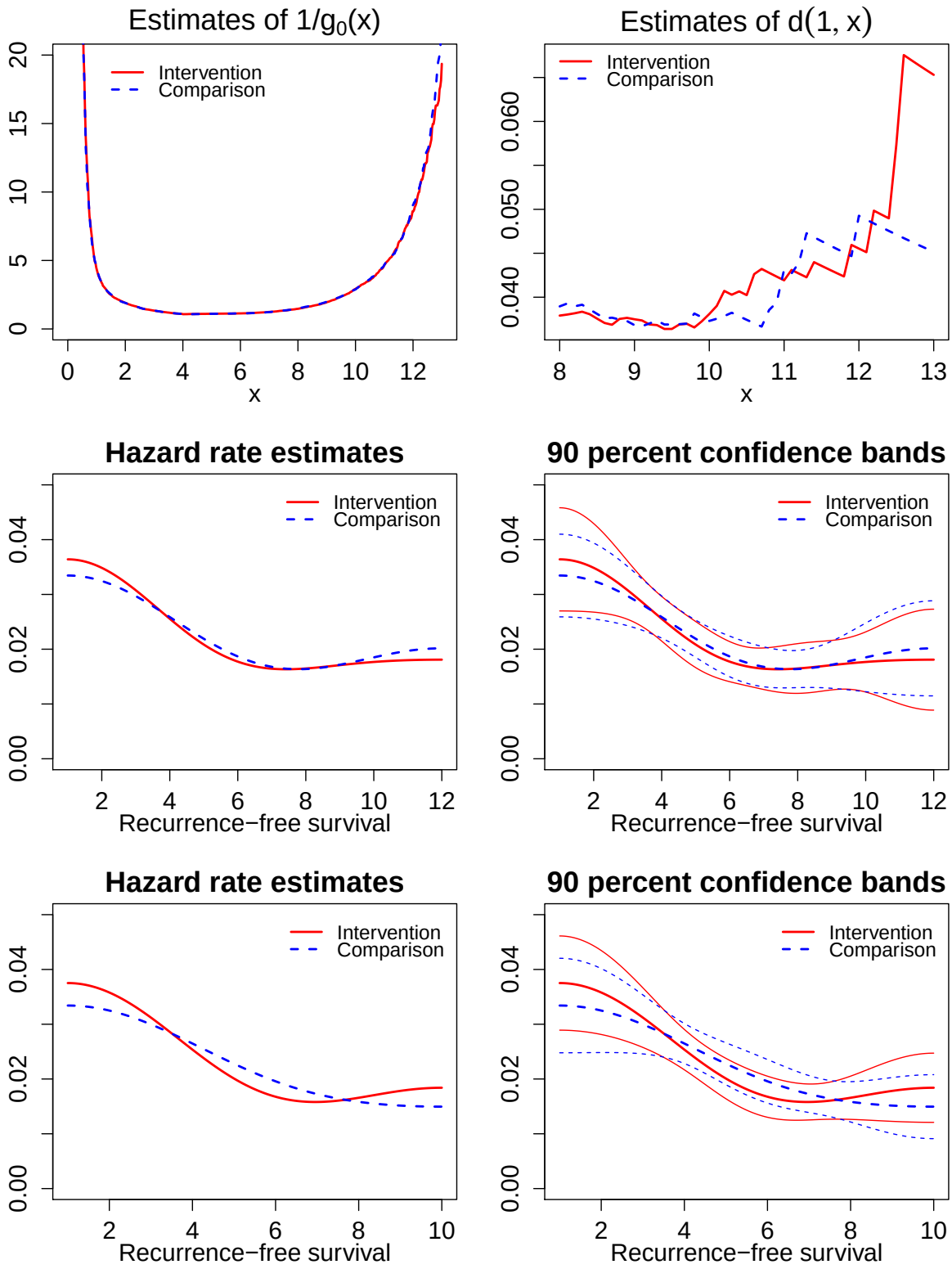

Figure 4: Estimation for intervention and comparison groups in the WHEL study. The middle and bottom diagrams correspond to estimates for intervals $[1, 12]$ and $[1, 10]$, respectively.

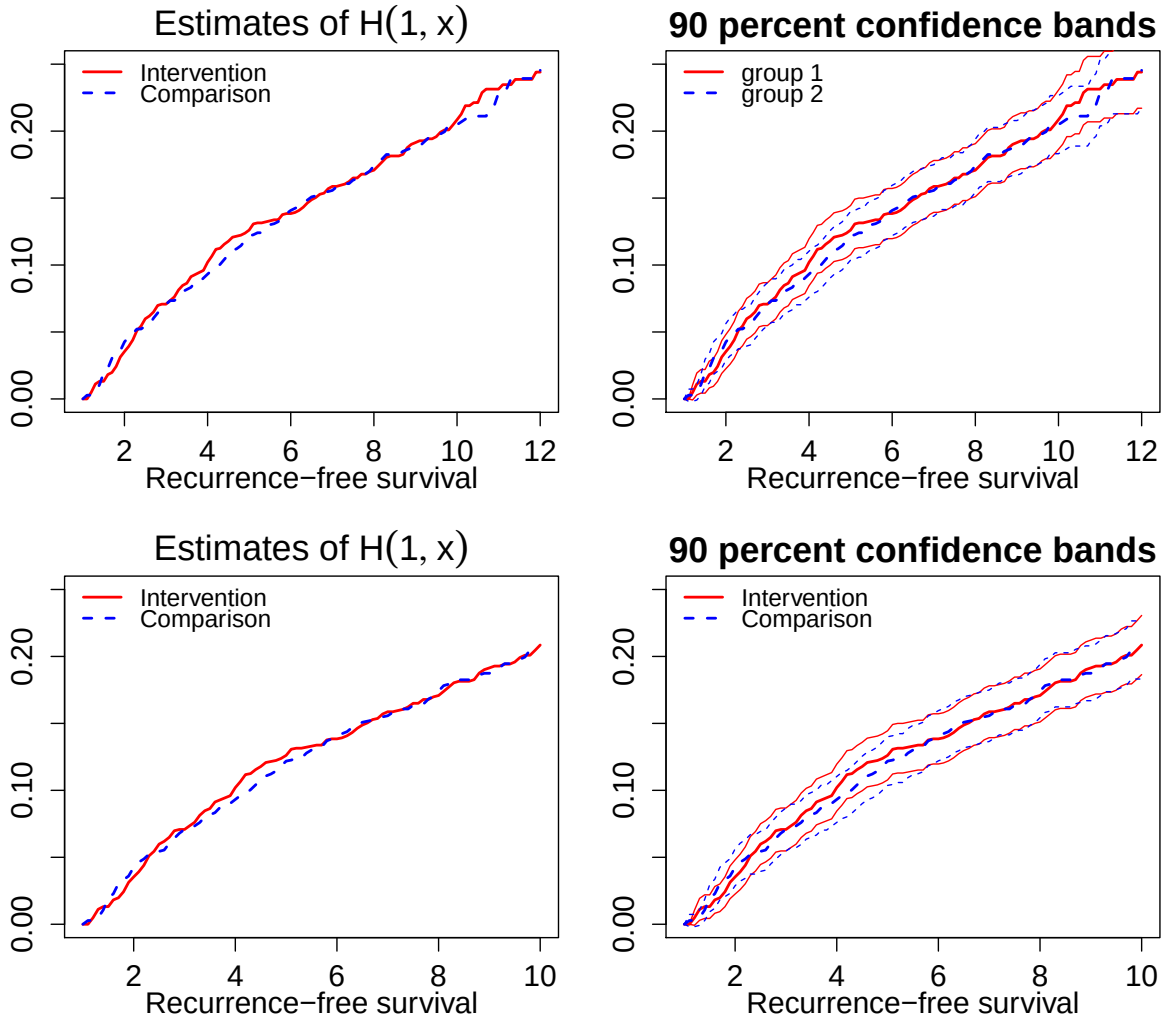

Figure 5: Estimates of the cumulative hazard $H^{X^*}(1,x) := \int_1^x h^{X^*}(v)dv$ and corresponding 90 percent confidence bands for intervention and comparison groups in the WHEL study calculated for cases $[a, a+b] = [1, 12]$ (the top diagrams) and $[a, a+b] = [1, 10]$ (the bottom diagrams).

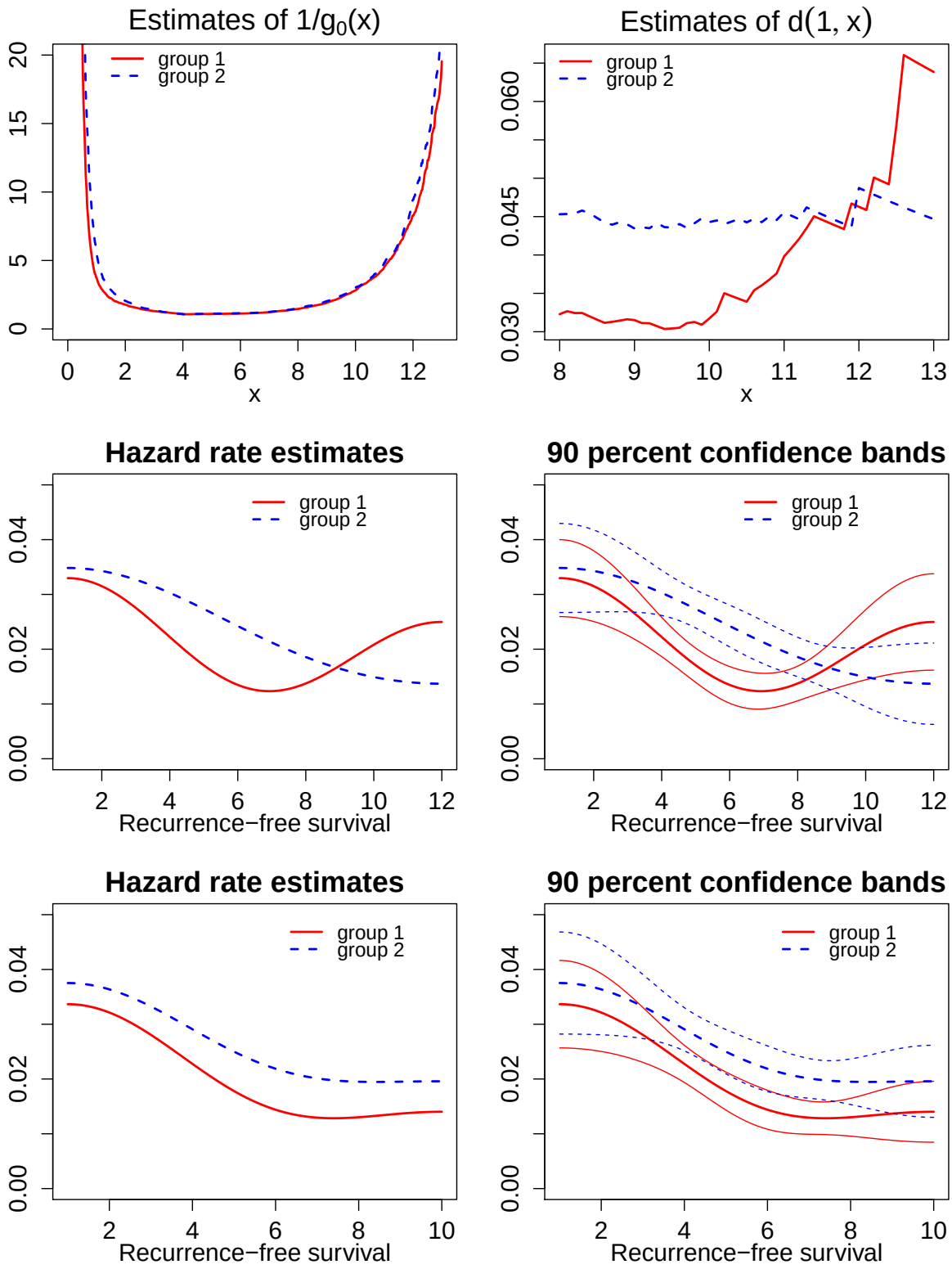

Figure 6: The effect of plasma alpha-carotene on hazard rates. Group 1 consists of women with levels of plasma alpha-carotene larger than the median level for all participants, all other women belong to group 2. The structure of the figure is identical to Figure 4.

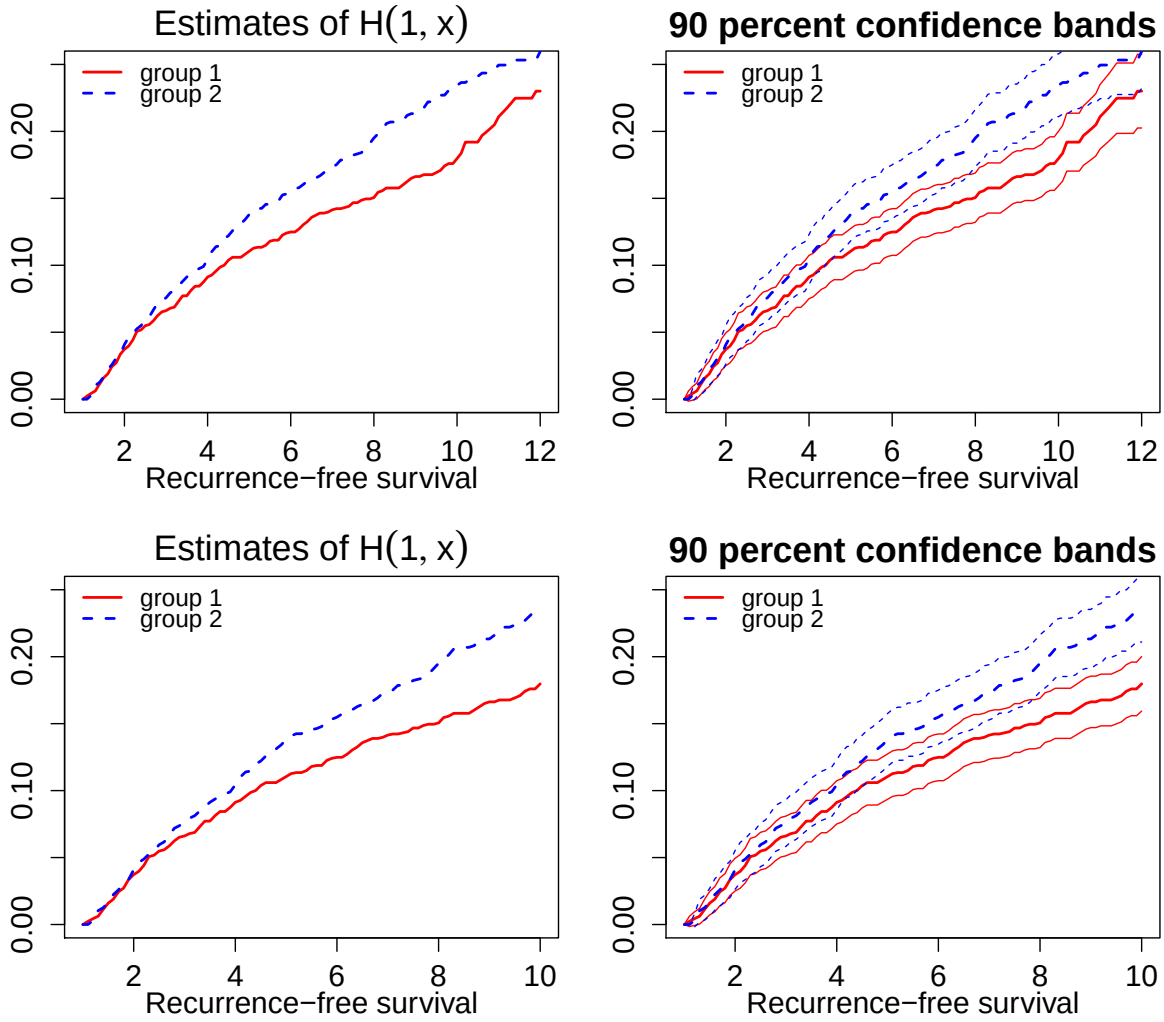

Figure 7: The effect of plasma alpha-carotene on cumulative hazard. Group 1 consists of women with levels of plasma alpha-carotene larger than the median level for all participants, all other women belong to group 2. The structure of the figure is identical to Figure 5.

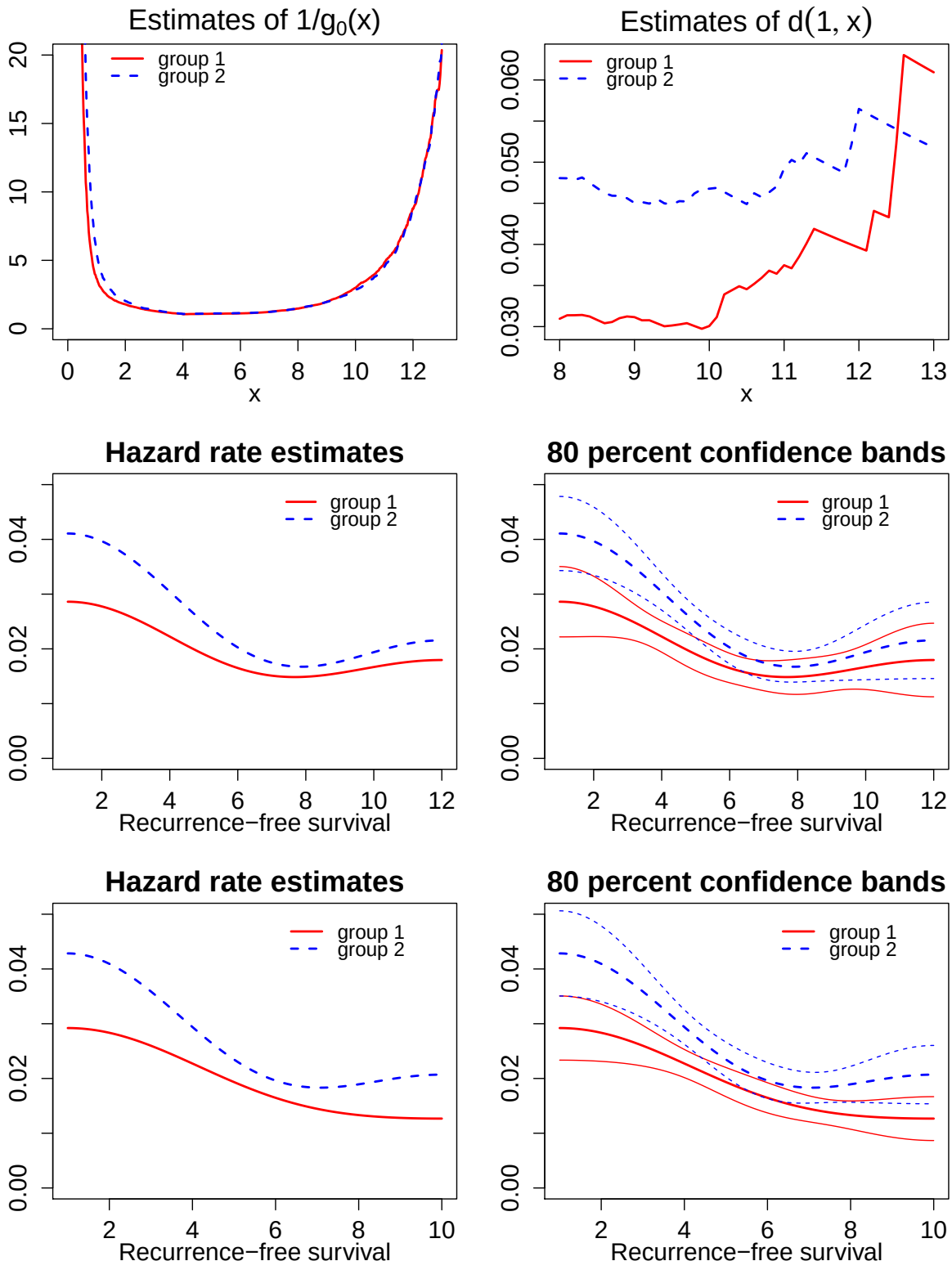

Figure 8: The effect of plasma beta-carotene on the hazard rate. The structure of the figure is identical to Figure 4.

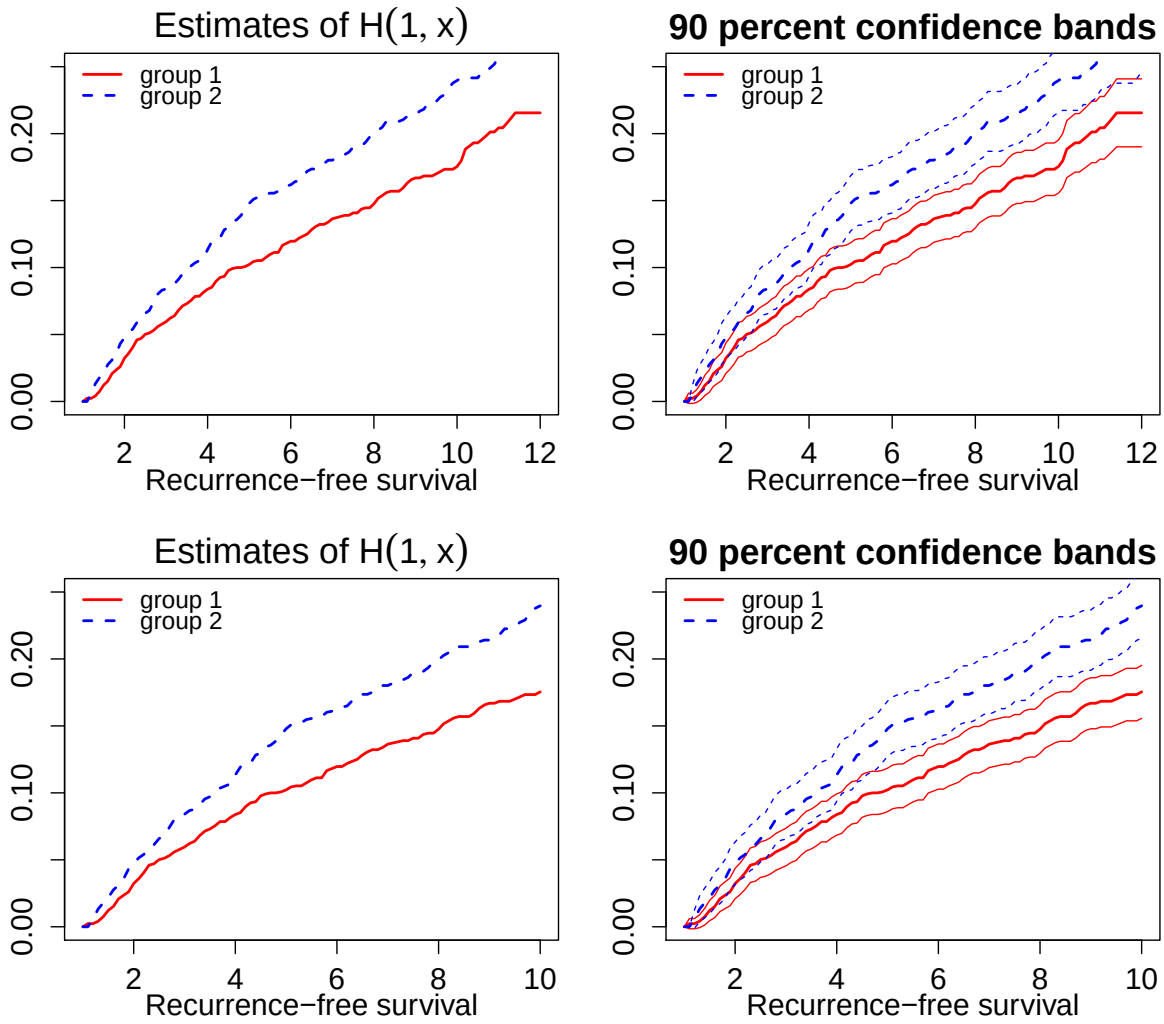

Figure 9: The effect of plasma beta-carotene on the cumulative hazard. The structure of the figure is identical to Figure 5.

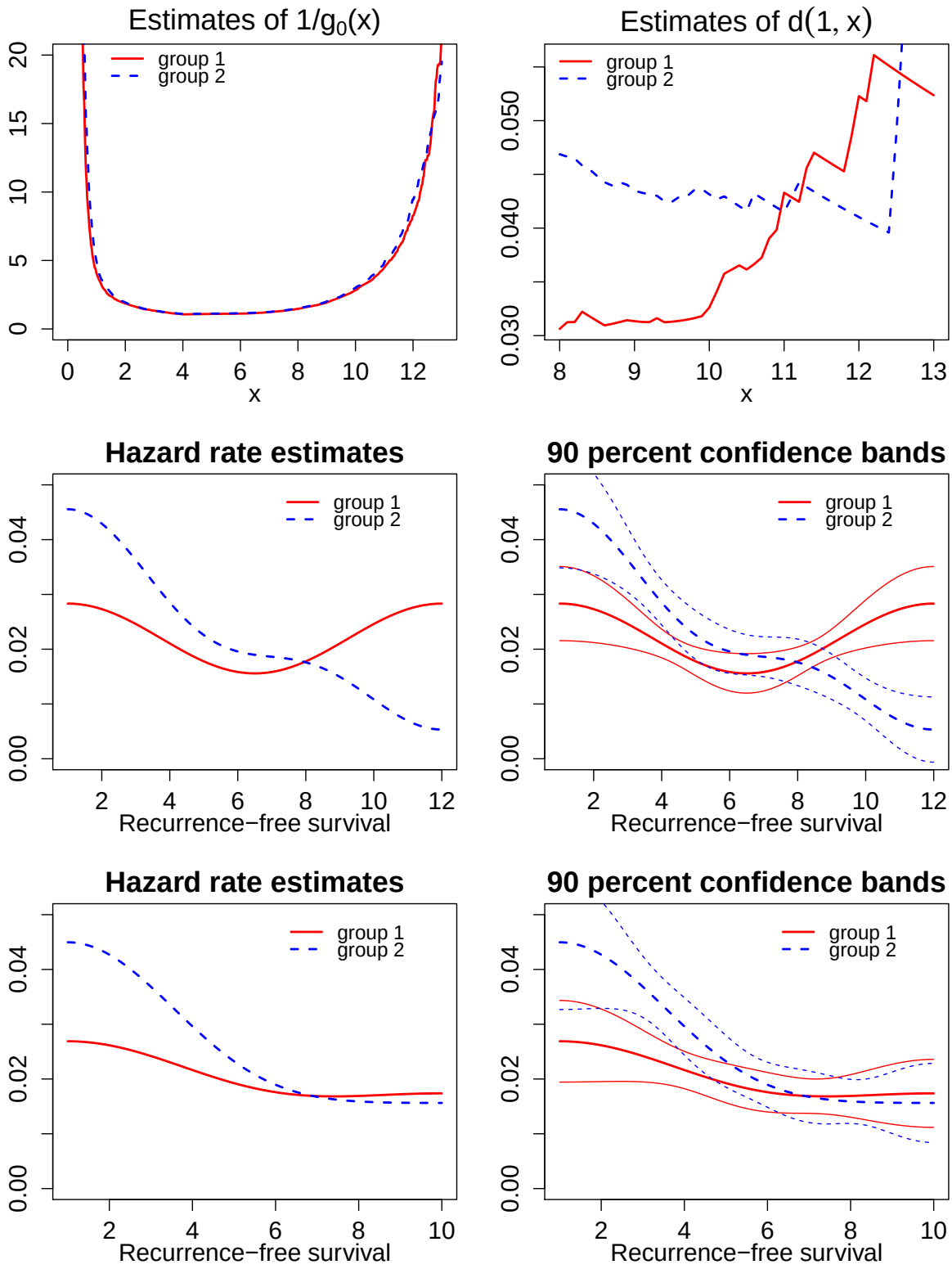

Figure 10: The effect of plasma lycopene on the hazard rate. The structure of the figure is identical to Figure 4.

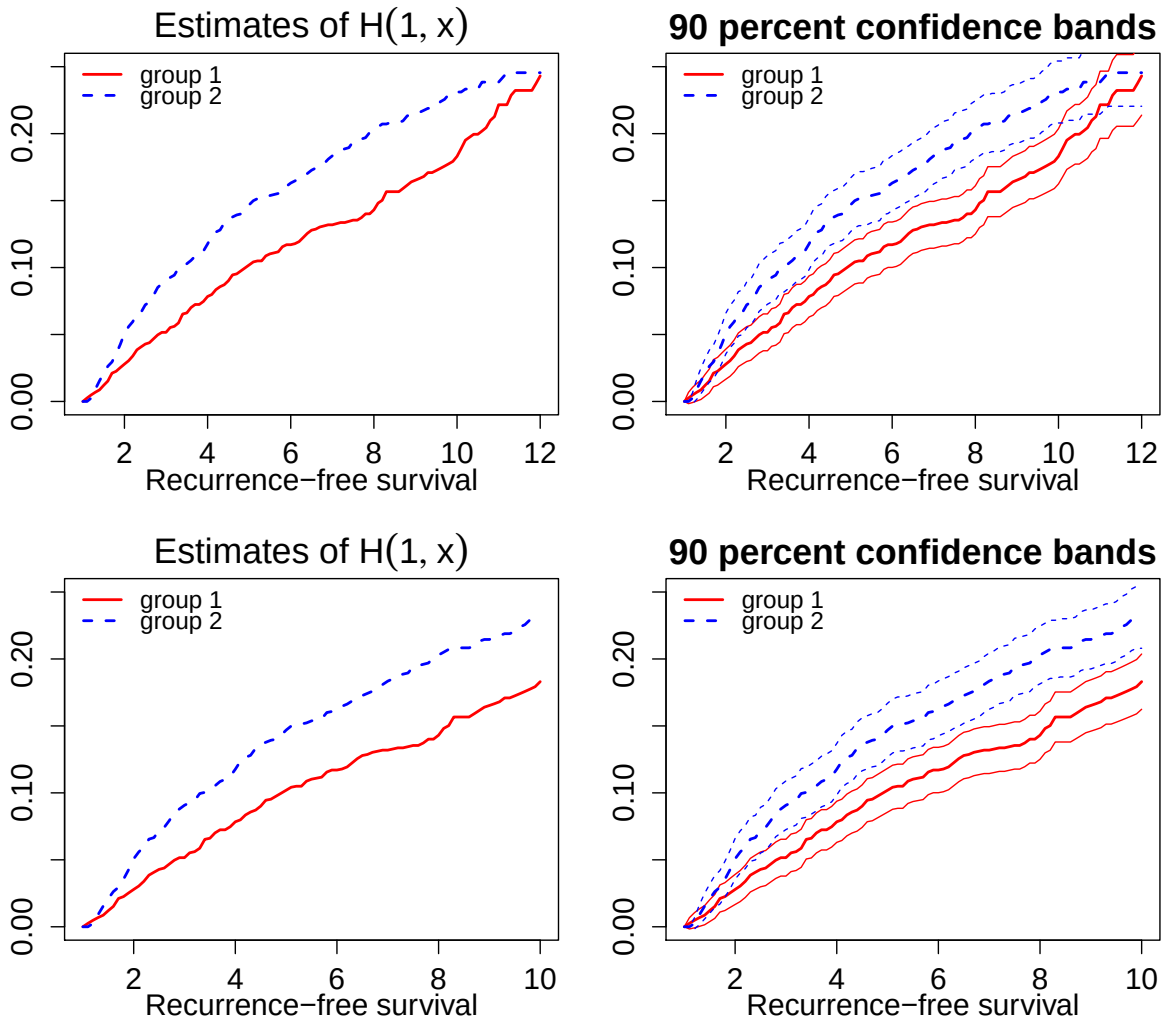

Figure 11: The effect of plasma lycopene on the cumulative hazard. The structure of the figure is identical to Figure 5.

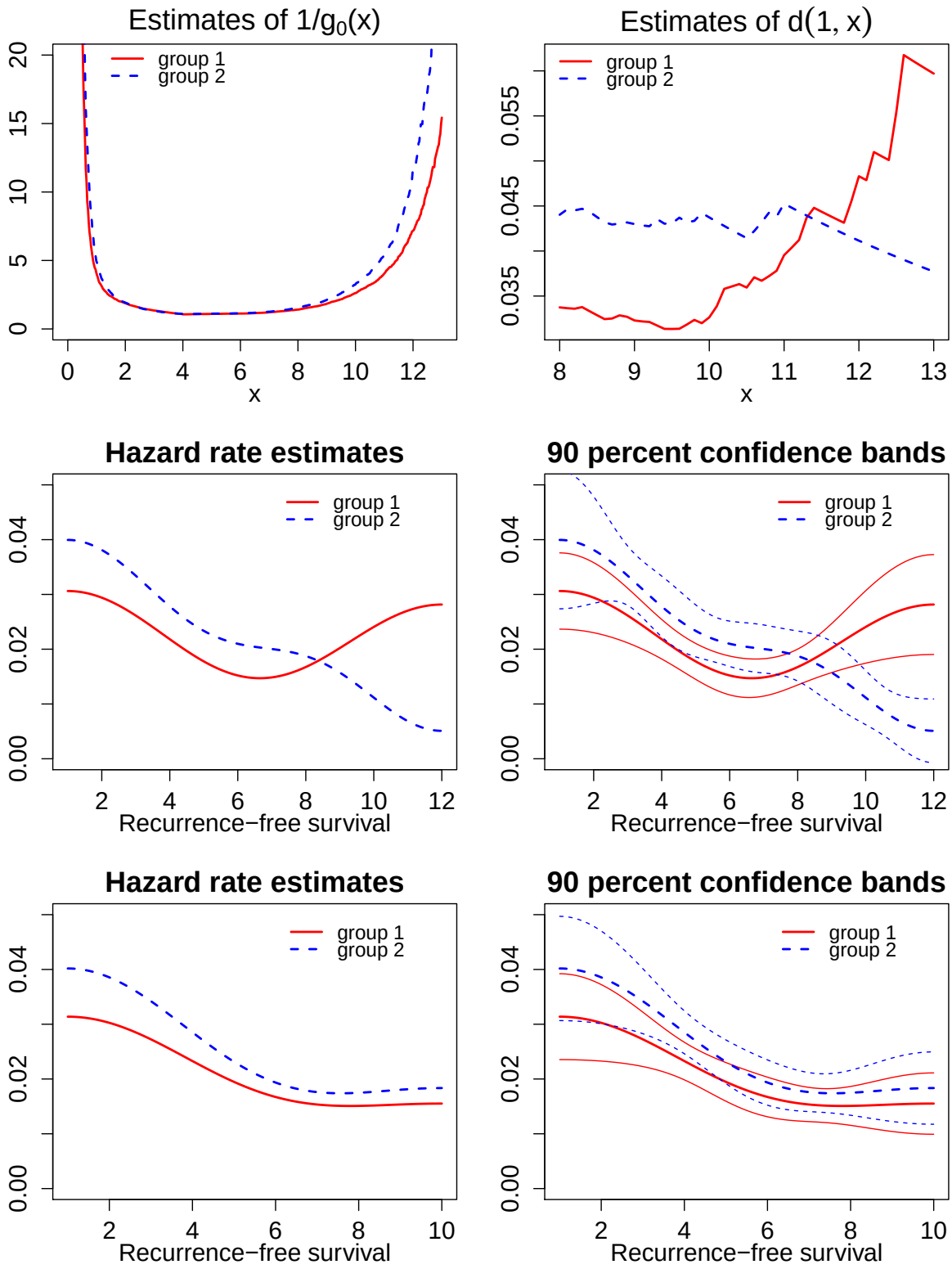

Figure 12: The effect of plasma cryptoxanthin on the hazard rate. The structure of the figure is identical to Figure 4.

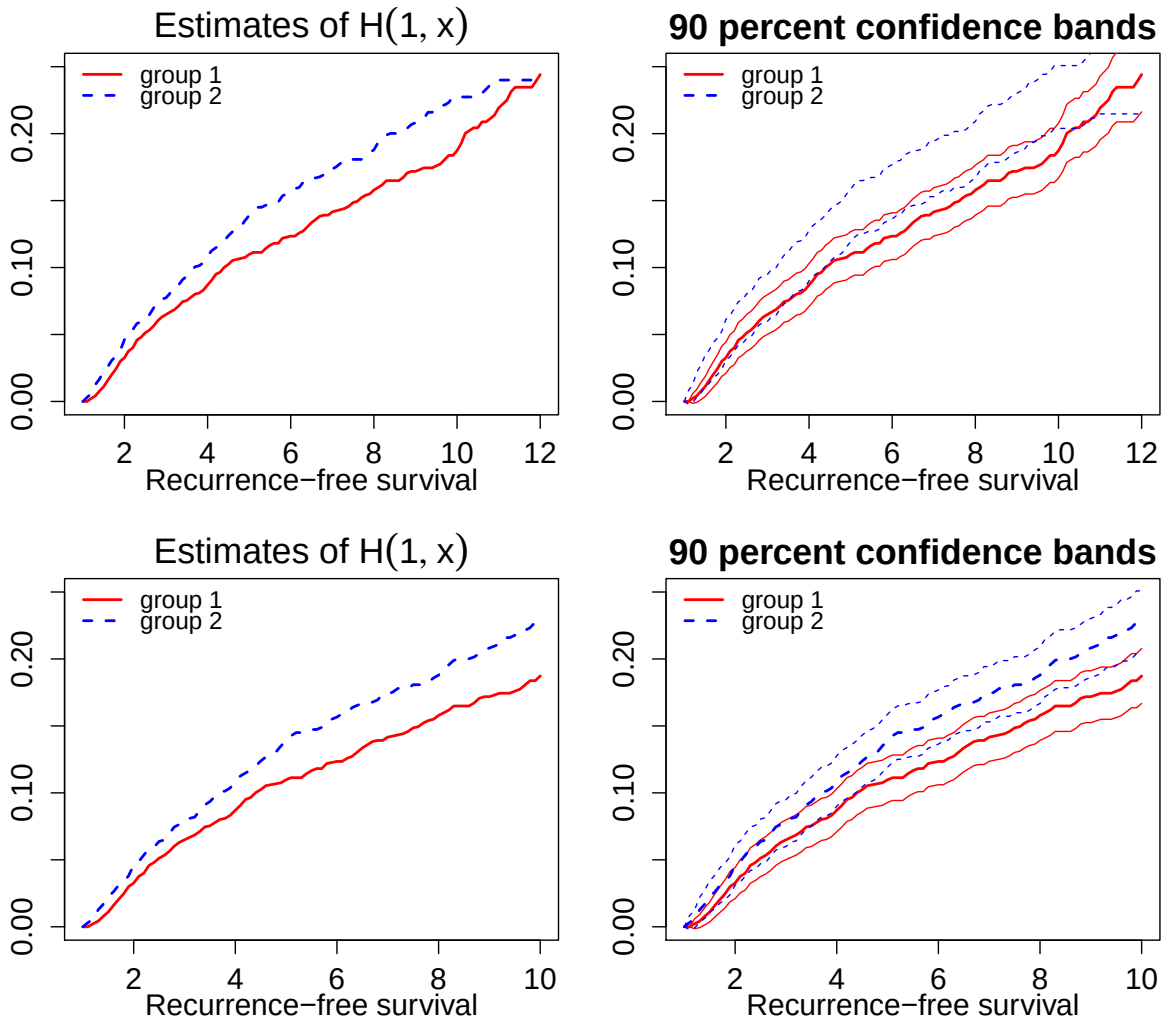

Figure 13: The effect of plasma cryptoxanthin on the cumulative hazard. The structure of the figure is identical to Figure 5.

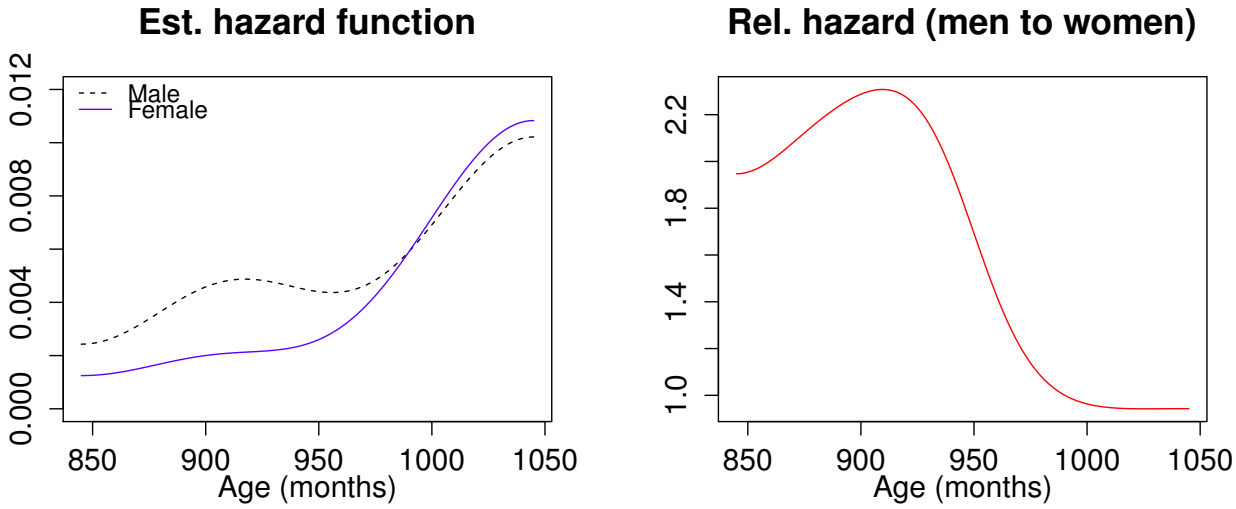

Figure 14: Analysis of the Channing House data. The left diagram shows estimated hazard rates for male (the dotted line) and female (the solid line) residents. The right diagram shows the relative hazard rate which is the ratio of the male's estimated hazard rate to the female's estimated hazard rate.

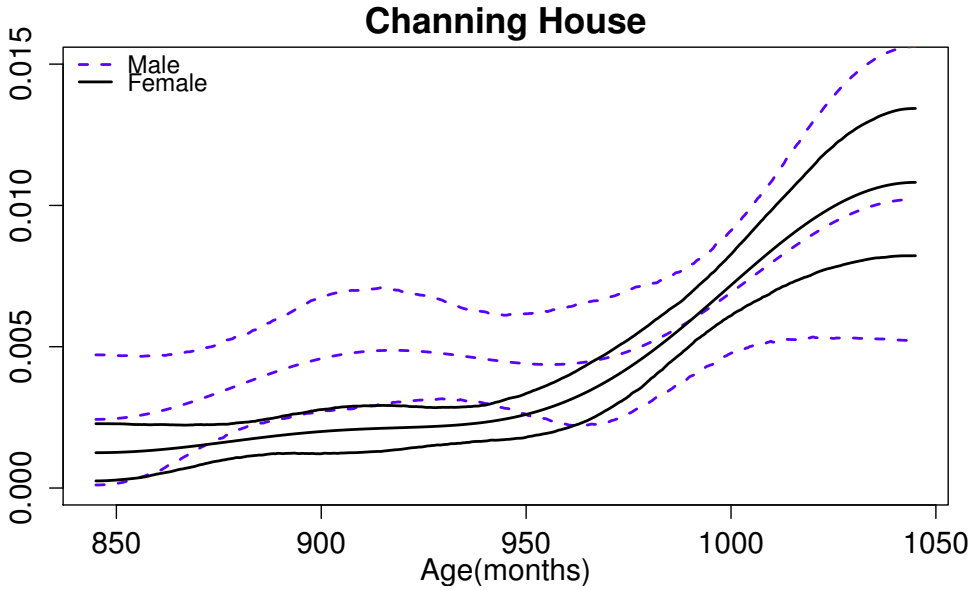

Figure 15: 80 percent confidence band for Channing House data.