Biometrika (2024), **xx**, x, *pp*. 1–32
Printed in Great Britain

Superefficient estimation of future conditional hazards based on time-homogeneous high-quality marker information

BY D. BAGKAVOS
Department of Mathematics, University of Ioannina, Ioannina, 45100, Greece
d.bagkavos@uoi.gr

A. ISAKSON
Bayes Business School, City St George's, University of London, EC1Y 8TZ, UK
alex.isakson.2@city.ac.uk

E. MAMMEN
Institute of Applied Mathematics, Heidelberg University, 69120 , Germany
mammen@math.uni-heidelberg.de

J.P. NIELSEN
Bayes Business School, City St George's, University of London, EC1Y 8TZ, UK
jens.nielsen.1@city.ac.uk

AND C. PROUST–LIMA
Univ. Bordeaux, INSERM, Bordeaux Population Health Research Center, U1219, F-33000 Bordeaux, France
cecile.proust-lima@u-bordeaux.fr

SUMMARY

We introduce a new concept for forecasting future events based on marker information. The model is developed in the nonparametric counting process setting under the assumptions that the marker is of so-called high quality and with a time-homogeneous conditional distribution. Despite the model having nonparametric parts it is established herein that it attains a parametric rate of uniform consistency and uniform asymptotic normality. In usual nonparametric scenarios, reaching such a fast convergence rate is not possible, so one can say that the proposed approach is superefficient. These theoretical results are employed in the construction of simultaneous confidence bands directly for the hazard rate. Extensive simulation studies validate and compare the proposed methodology with the joint modeling approach and illustrate its robustness for mild violations of the assumptions. Its use in practice is illustrated in the computation of individual dynamic predictions in the context of primary biliary cirrhosis of the liver.

Some key words: Counting processes; Dynamic prediction; Kernel hazard estimation; Nonparametric smoothing; Survival Analysis

1. INTRODUCTION

This paper investigates a novel approach to understanding future survival when the hazard depends on a developing marker process. Given some natural assumptions on the marker process and the marker dependent hazard, we establish herein that the proposed technique achieves parametric rates of convergence. That is so even though the specifications of the marker process and the marker dependent hazard are fully nonparametric. Nonparametric estimators with this property have been called superefficient, see e.g. Nielsen (1999). This should not be mixed up with the notion of superefficiency used in asymptotic parametric statistics when discussing optimality of estimators. The estimator proposed here allows one to analyse and visualise survival forecasting using methodology related to square-root-n consistent mathematical statistics. In particular, we derive uniform confidence bands based on approximations by Gaussian processes, proceeding similarly as in statistical inference based on cumulative distribution functions, Kaplan–Meier estimators or Nelson–Aalen estimators. While most of the mathematics of this paper is new, interpretation of its results requires only conventional intuition. A key asymptotic requirement of the proposed technique is that the nonparametric components involved are undersmoothed with vanishing asymptotic bias. In this sense our approach is related to semiparametric statistics. Implementation of the estimator in practice is facilitated via a fully automatic smoothing methodology, also developed herein, based on an adapted version of cross-validation. Our final approach is therefore fully data driven. Additionally, the present research develops theory for uniform confidence bands for future development of conditional hazards for individuals with a certain present marker level, thus allowing its first practical implementation.

Development of the new methodology contained in this article requires two crucial assumptions. The first one is that conditional hazards depend on time only through the value of the marker process at this time point. We regard such markers as being of high quality, a notion that goes back to Nielsen (1999), see also Nielsen (2000). Secondly, for the marker process we make a Markov-type assumption that allows us to predict the further development of the marker. In our model we take a purely nonparametric view. The main intuition behind our new approach to modeling the full survival system was already given in Nielsen (2000) with some improved technical indications in Mammen & Nielsen (2007). However, the practicalities and full technical consequences of this new approach were never investigated and as a result this new approach has never entered mainstream statistics nor it has ever been implemented on real data up to now.

The proposed methodology applies in many fields such as Data Mining or Asset-Liability management, see Nielsen (2000). Below we will discuss a health research example which requires the forecast of the hazard rate function as a disease progresses. Specifically, in analysing the primary biliary cirrhosis (PBC) data set of the Mayo Clinic (Therneau & Grambsch, 2000), the objective is the prediction of clinical progression of patients diagnosed with PBC, based on repeated measures of different biological markers and time to clinical progression. There exist multiple examples in the Biostatistical literature with similar modeling objective, including e.g. prostate specific antigen and prostate cancer recurrence (Proust-Lima & Taylor, 2009) or CD4 counts and HIV infection (Fusaro et al., 1993; Cui et al., 2023).

For the analysis of such data, individual dynamic prediction techniques have been proposed, e.g. landmarking (Anderson et al., 1983; Van Houwelingen, 2007) or joint modelling (Henderson et al., 2000; Rizopoulos, 2012). The landmarking approach uses individuals alive at t and information on the biomarker up to t to fit a proportional hazards model and estimate the probability of surviving up to $t + s$ (Ferrer et al., 2019); see Van Houwelingen (2007) for an implementation where smoothing with respect to t has been included. The joint modeling methodology estimates the distribution of the biomarker and the time-to-event and derive the conditional future event

probability using past biomarker data (Rizopoulos, 2012; Proust-Lima & Taylor, 2009). In this model, the marker process is usually described through a linear mixed model with a stochastic component W_i for individual i, and the time-to-event is usually modelled via a mixed proportional hazards model that includes the same stochastic component W_i. The two submodels for marker and survival are linked by sharing the same stochastic component W_i, and the marker process is thus conditionally independent of the survival times given W_i. Both approaches have been shown as being powerful tools in statistical inference, especially for individual dynamic prediction (Ferrer et al., 2019). However both methodologies do not fully capture the stochastic structure of the data. This is so because in landmarking one has a separate model for each value of the landmark time t with biomarker data only considered up to t so that the two processes are not temporally and mutually linked for all times. This induces that predictions at landmark t are not consistent (in the sense of Jewell & Nielsen 1993) with predictions at other landmark times (Suresh et al., 2017; Ferrer et al., 2019). In contrast, the joint model approach relies on the joint distribution of the two processes and is thus likely to provide consistent predictions. However, in practice, joint models suffer from simplifying parametric assumptions that highly reduce the flexibility and stochasticity of the model and association between the two processes (Ferrer et al., 2019). For instance, the shared stochastic component W_i is almost always limited to a vector of time-independent random effects.

This paper contains a full asymptotic study for the case of one-dimensional marker processes. We will show that under our two assumptions, the full survival system can be estimated with parametric rates. This has important technical consequences. For example, our resulting forecast of the future hazard can be estimated uniformly consistently for any individual starting the forecasting at any value of the marker process. Extended to this methodology, a wild bootstrap approach, see Mammen (1992), is developed and provides both pointwise and uniform confidence bands of the entire range of the future hazard. This is very powerful and yields a real alternative to the commonly used parametric approach. Due to the mathematical challenges of our approach, we have developed the full mathematical insight described above for the one-dimensional marker case only. Possible extensions to higher dimensional markers are briefly discussed at the end of the paper.

2. MODEL FORMULATION

Consider n individuals observed in the time interval $[0, T]$ with survival times $T_1, \ldots, T_n$. The methodology proposed herein is developed under the same model formulation as in Nielsen (2000). For $i = 1, \ldots, n$ let N_i be the counting process, which indicates the observed event for the ith individual and Z_i the exposure process taking values in $\{0, 1\}$, with 1 indicating that the ith individual is at risk to encounter the event of interest. Furthermore, additional information for every individual is available in the form of a one-dimensional predictable cadlag marker process $X_i(t)$, $t \in [0, T]$. We assume that $N^{(n)} = (N_1, \ldots, N_n)$ is an n-dimensional counting process with respect to the increasing, right continuous filtration $\mathcal{F}_t = \sigma(N^{(n)}(s), X^{(n)}(s), Z^{(n)}(s); s \leq t)$, $t \in [0, T]$, where $Z^{(n)} = (Z_1, \ldots, Z_n)$ and $X^{(n)} = (X_1, \ldots, X_n)$. The observed data are denoted by $\mathbb{X}_n = \{X^{(n)}(t), Z^{(n)}(t), N^{(n)}(t); t \in [0, T]\}$. Note that when for each individual $i, X_i(t) = x_i \in \mathbb{R}$ for all $t \in [0, T]$, the present formulation defaults to the time-invariant covariates setting.

The following assumptions are used throughout the article. The main assumptions are the first two. Assumption [A1] expresses the fact that the conditional hazard $\alpha(\cdot)$ only depends on the marker information and in particular not on time.

[A1] (High-quality marker information) The stochastic processes $(N_1, X_1, Z_1), \ldots, (N_n, X_n, Z_n)$ are independent and identically distributed with predictable intensity

$$\lambda_i(t)\mathrm{d}t = E\{dN_i(t) \mid \mathcal{F}_{t-}\} = \alpha\{X_i(t)\}Z_i(t)\mathrm{d}t,$$

where the conditional hazard $\alpha(\cdot)$ depends on time t only via the marker $X_i(t)$.

Our second main assumption is a Markov-type condition and concerns the dynamics of the marker processes $X_i(t)$. In what follows, s can be seen as landmark time, i.e., the time at which the prediction is to be computed.

[A2] (Markov-type assumption on marker process) The conditional distribution of $X_i(s+t)$, given $\mathcal{F}_s$, $T_i \geq s+t$, $Z_i(s+t) = Z_i(s) = 1$, depends only on $X_i(s)$ and t and, in particular, not on s.

Assumptions [A1] and [A2] define a setting where it is possible to use previous data for future prediction based only on marker information. A situation, as stated in [A1], where $\alpha(\cdot)$ only depends on the marker is described in Yong et al. (1997) who argues that time since infection of AIDS has little implication on its hazard. In this particular example, Assumption [A1] is covering the fact that the count of CD-4 cells in the blood is way more important than just the time since infection. Time is still part of the model through the time dependent covariates X_i, N_i and Z_i. Time invariance of the marker process transition probabilities stated in Assumption [A2] allows using already observed developments of patients in the past for future predictions. Methodology for assessing the validity of [A1] and [A2] before applying the proposed approach in practice is provided in Section B.2 of the supplementary material. The same section also contains numerical evidence on the sensitivity of the approach when either assumption is violated; the results indicate that the method is robust to mild violations of both assumptions.

For x_* in the interior of the support of $X_i(t)$ we make the following additional assumptions which are discussed in Section B.3 of the supplementary material.

[A3] The support A of $X_i(t)$ is a finite closed interval that does not depend on t. The conditional density $f(s, s+t, x, z)$ of $\{X(s), X(s+t)\}$ given $(Z(s), Z(s+t)) = (1,1)$ exists and is twice continuously differentiable with respect to x and z for $\delta_T \leq s \leq s+t \leq T-\delta_T$, $x, z \in A$. The conditional density $f(s,z)$ of $X(s)$ given $Z(s) = 1$ and the functions

$$\int_A f(s, s+t, x, z)\mathrm{d}z, \quad \int_A \alpha(z) f(s, s+t, x, z)\mathrm{d}z$$

exist and are twice continuously differentiable with respect to x for $0 \leq s \leq s+t \leq T$, $x \in A$. The conditional hazard $\alpha(x)$ is twice continuously differentiable for $x \in A$ and bounded away from 0. For some constant $C > 0$ we have that $|f(s, s+t_1, x, z) - f(s, s+t_2, x, z)| \leq C|t_1 - t_2|$ for $s, t_1, t_2 \geq 0$, $s+t_1, s+t_2 \leq T$, $\delta_T \leq t_1, t_2 \leq T - \delta_T$, $x, z \in A$.

[A4] The expectations $\gamma(s) = E\{Z_i(s)\}$ and $\gamma(t,s) = E\{Z_i(t+s)Z_i(s)\}$ exist and are continuous. For some constant $C > 0$ we have that $|\gamma(t_1, s) - \gamma(t_2, s)| \leq C|t_1 - t_2|$ for $s, t_1, t_2 \geq 0$, $s+t_1, s+t_2 \leq T$, $\delta_T \leq t_1, t_2 \leq T - \delta_T$. Furthermore, for $\delta_T \leq t \leq T - \delta_t$ the term $\Gamma(t, x_*)$ is bounded from below and for $x \in A$ the function $E(x)$ is bounded from below, where

$$\Gamma(t, x_*) = \int_0^{T-t} \gamma(t,s) f(s, t+s, x_*, z)\mathrm{d}s\mathrm{d}z, \; E(x) = \int_0^T \gamma(s) f(s,x)\mathrm{d}s.$$

[A5] The kernel K has bounded support, $[-1,1]$ say, and is continuously differentiable on $[-1,1]$. The bandwidths b_1 and b_2 are equal to $c_{b,1}n^{-\rho_1}$ or $c_{b,2}n^{-\rho_2}$ for some $c_{b,1}, c_{b,2} > 0$, $1/4 < \rho_1, \rho_2 < 1/3$.

[A6] It holds that

$$\text{pr}\{|X_i(s+t) - X_i(s)| \leq \delta, Z_i(s+t) = 1 \mid X_i(s) = z, Z_i(s) = 1\} \leq \delta\kappa(t)$$

for all $z \in A$, $0 \leq s < s+t \leq T$ and $\delta > 0$ small enough, where κ is a positive function with $\int_0^T \kappa(t)\mathrm{d}t \leq C_\kappa$ for a constant $C_\kappa > 0$.

[A7] For some constant $C > 0$ we have that $E\{|X_i(s+t_1) - X_i(s+t_2)|^4\} \leq C|t_1 - t_2|^2$ for $s, t_1, t_2 \geq 0$, $s+t_1, s+t_2 \leq T$, $\delta_T \leq t_1, t_2 \leq T - \delta_T$.

In the sequel it is shown that under these conditions, it is possible to estimate the marker conditional future hazard with starting point s

$$h_{x,s}(t) = \text{pr}\left\{T_i \in (s+t, s+t+\mathrm{d}t) \mid X_i(s) = x, Z_i(s+t) = Z_i(s) = 1\right\}/\mathrm{d}t, \tag{1}$$

where T_i is the survival time of individual i. Assumption [A2] allows writting $h_{x,s}(t) = h_x(t)$. We will use $h_{x,s}(t)$ if we want to underline when the prediction is made but keep the notation $h_x(t)$ most of the time. Our main results contain limiting distributions for estimators of the process $h_x(t)$ for fixed value of x and for the consistency of uniform confidence bands based on wild bootstrap. Trivially, we have

$$h_{x,s}(t) = h_x(t) = \text{pr}\left\{T_i \in (s+t, s+t+\mathrm{d}t) \mid X_i(s) = x, T_i > s+t\right\}/\mathrm{d}t,$$

under the additional assumption of non-informative censoring, i.e.

$$\begin{aligned}\text{pr}\left\{T_i \in (s+t, s+t+\mathrm{d}t) \mid X_i(s) = x, T_i > s+t\right\} \\ = \text{pr}\left\{T_i \in (s+t, s+t+\mathrm{d}t) \mid X_i(s) = x, Z_i(s+t) = Z_i(s) = 1\right\}.\end{aligned}$$

Under this assumption, our approach yields an estimator of the conditional survival function, defined by simply integrating the conditional hazard estimator appropriately.

Consequently the prediction of the future conditional hazard becomes an estimation problem. Thus, our approach can be considered as an in-sample forecasting method. In-sample forecasting has been introduced using structural models in Martínez-Miranda et al. (2013) and has been majorly used in reserving. See also Hiabu et al. (2016); Mammen et al. (2021) and Bischofberger et al. (2019). Our framework allows us to write the marker conditional future hazard as

$$h_x(t) = E[\alpha\{X_i(s+t)\} \mid X_i(s) = x, Z_i(s) = Z_i(s+t) = 1].$$

Let $K_b(\cdot) = b^{-1}K(\cdot/b)$ be a kernel with bandwidth b. By Assumption [A5] K is a continuously differentiable function with bounded support. Estimators for $\alpha(z)$ and $h_x(t)$ have been proposed in Nielsen (2000) and we will use them here as well. For bandwidths b_1 and b_2 define

$$\hat{\alpha}_{i,b_1}(z) = \frac{\sum_{k\neq i}\int_0^T K_{b_1}\{z - X_k(s)\}\mathrm{d}N_k(s)}{\sum_{k\neq i}\int_0^T K_{b_1}\{z - X_k(s)\}Z_k(s)\mathrm{d}s}, \tag{2}$$

and

$$\hat{h}_{x,b_1,b_2}(t) = \frac{\sum_{i=1}^n \int_0^T \hat{\alpha}_{i,b_1}\{X_i(t+s)\}Z_i(t+s)Z_i(s)K_{b_2}\{x - X_i(s)\}\mathrm{d}s}{\sum_{i=1}^n \int_0^T Z_i(t+s)Z_i(s)K_{b_2}\{x - X_i(s)\}\mathrm{d}s}. \tag{3}$$

To simplify notation we also write $\hat{\alpha}_i$ and $\hat{h}_x$ for $\hat{\alpha}_{i,b_1}$ and $\hat{h}_{x,b_1,b_2}$ if dependence on bandwidths does not need to be highlighted. Estimator $\hat{\alpha}_i$ is a usual leave-one-out estimator for the hazard.

Its use as an approximation of $\alpha(\cdot)$ is intuitive since $\hat{\alpha}_i$ can be thought as the natural extension to the present setting of Hjort (1994)'s local likelihood principle in modeling a constant hazard rate function, see also Nielsen & Linton (1995). Also $\hat{h}_x(t)$ arises from using a kernel estimator of the conditional density of $X_i(s)$ and combining it with $\hat{\alpha}_i$. For achieving parametric rates it is important that the bias terms arising in the smoothing are of order $o(n^{-1/2})$. This can be achieved by choosing the bandwidths b_1 and b_2 such that $b_1^2, b_2^2 = o(n^{-1/2})$, see again Assumption [A5]. Notice that our estimators also work when some data are right censored. This is controlled via the individual time dependent exposure measures $Z_i(\cdot)$.

3. ASYMPTOTIC THEORY

This section formulates the main result of the present research which states that the proposed hazard estimator yields a parametric rate of uniform convergence and is asymptotically normal. On first sight this fact seems very surprising since we estimate the hazard rate non-parametrically. For the pointwise case, this has been already observed in Nielsen (2000); see also related models studied in Castellana & Leadbetter (1986), Bosq (2012), Kutoyants (2004), and Aeckerle-Willems & Strauch (2022), where parametric rates show up in nonparametric settings. Thus, these models differ from other non-parametric models where estimators only allow a slower convergence rate compared to parametric estimation problems. The main difference lies in the fact that typically in nonparametric problems only local information can be used whereas in our model all individuals with markers $X_i(\cdot)$ having visited the neighbourhood of x at some time point $s \leq T - t$ add information for the estimation of $h_x(t)$.

THEOREM 1. *Suppose that Assumptions [A1]–[A7], apply for an $x = x_*$ in the interior of the support of $X_i(t)$. Then it holds for $\delta_T > 0$ that*

$$n^{-1/2}(\hat{h}_{x_*} - h_{x_*}) \to \mathbb{G}_{x_*}$$

in distribution, weakly in $\ell^\infty([\delta_T, T - \delta_T])$, where $\mathbb{G}_{x_*}$ is a tight Gaussian process taking values in $\ell^\infty([\delta_T, T - \delta_T])$ with mean 0 and covariance $\Sigma(t_1, t_2)$ stated in Section B.7 of the supplementary material.

For the proof of Theorem 1 see Sections B.4–B.6 in the supplementary material. There, we also argue that the result of the theorem also holds for the boundary points x_* of A if in the smoothing step (3) the convolution kernel $K_{b_2}(x - u)$ is replaced by a boundary corrected kernel or if we use local linear estimation instead of local constant smoothing. We conjecture that the result of Theorem 1 also holds if the interval $[\delta_T, T - \delta_T]$ is replaced by $[0, T - \delta_T]$. But this would require additional detailed assumptions on the speed of convergence of the joint density of $\{X_i(t_1), X_i(t_2)\}$ to infinity for $t_1 - t_2 \to 0$. Theorem 1 is used in the next Section as the basis for constructing pointwise and uniform confidence bands. Implementation of the proposed estimator in practice requires the development of a data driven, consistent bandwidth selection rule which is discussed in detail in Section A.2 of the supplementary material.

4. CONFIDENCE BANDS

We now introduce a bootstrap procedure for the construction of pointwise and uniform confidence sets. For a critical discussion of confidence sets based on Gaussian approximations, see Bie et al. (1987). Our approach is a slight modification of wild bootstrap procedures recently proposed in counting process models (Beyersmann et al., 2013; Bluhmki et al., 2019). Because in our case the estimation error is not a martingale, we need to adapt this approach to our setting. We will use the multiplier bootstrap, as presented in Chernozhukov et al. (2013). Fix x_* in the

interior of the support of $X_i(t)$. Define

$$W_i(t) = \int_0^T \hat{\alpha}_{i,b_1}\{X_i(t+s)\} Z_i(t+s) Z_i(s) K_{b_2}\{x_*, X_i(s)\} \mathrm{d}s.$$

Then

$$n^{-1} \sum_{i=1}^{n} \hat{\Gamma}(t, x_*)^{-1} W_i(t) = \hat{h}_{x_*}(t), \text{ where}$$
$$\hat{\Gamma}(t,x) = n^{-1} \sum_{i=1}^{n} \int_0^{T-t} Z_i(t+s) Z_i(s) K_{b_2}\{x, X_i(s)\} ds.$$

In the proof of Theorem 1 it is shown that $\hat{h}_{x_*}(t) = A(t) + B(t) + o_P(n^{-1/2})$, where

$$A(t) = n^{-1} \sum_{i=1}^{n} \int_0^T g_{t,x_*}\{X_i(s)\} dM_i(s),$$
$$B(t) = n^{-1} \sum_{i=1}^{n} \int_0^T \left[\alpha\{X_i(t+s)\} - h_{X_i(s)}(t)\right] Z_i(t+s) Z_i(s) K_{b_2}\{x_*, X_i(s)\} \mathrm{d}s$$

and

$$g_{t,x}(z) = \frac{\int_0^{T-t} \gamma(t,s) f(s, t+s, x, z) ds}{\int_0^T \gamma(s) f(s,z) ds}. \tag{4}$$

Our bootstrap estimator of the distribution of the process $\hat{h}_{x_*}(\cdot) - h_{x_*}(\cdot)$ is the conditional distribution of $\bar{h}_{x,B}(\cdot) = A_B(\cdot) + B_B(\cdot)$ given all observations $\mathbb{X}_n = \{N_i(t), Z_i(t), X_i(t) : 1 \le i \le n, 0 \le t \le T \text{ with } Z_i(t) = 1\}$, where

$$A_B(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^T \hat{g}_{i,t,x_*}\{X_i(s)\} V_i [dN_i(s) - \hat{\alpha}_i\{X_i(s)\} Z_i(s) ds],$$
$$B_B(t) = n^{-1/2} \sum_{i=1}^{n} V_i \{\hat{\Gamma}(t, x_*)^{-1} W_i(t, x_*) - \hat{h}_{x_*}(t)\},$$

with V_i i.i.d. normal random variables independent of (N_i, X_i, Z_i) with expectation 0 and variance 1. Furthermore, we have used the following leave-one-out estimator of $g_{t,x}(z)$

$$\hat{g}_{i,t,x}(z) =$$
$$n^{-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \int_0^{T-t} \hat{E}_j\{X_j(t+s)\}^{-1} K_{b_2}\{z, X_j(t+s)\} Z_j(t+s) Z_j(s) K_{b_2}\{x, X_j(s)\} ds. \tag{5}$$

We apply the bootstrap to get an approximation for the quantiles of the distribution of the random variables $\sigma^{-1}_{\mathbb{G}_{x_*}}(t)\{\hat{h}_{x_*}(t) - h_{x_*}(t)\}$ for fixed t and for the $\sup_{t \in [\delta_T, T-\delta_T]} \sigma^{-1}_{\mathbb{G}_{x_*}}(t) |\hat{h}_{x_*}(t) - h_{x_*}(t)|$. Denote the asymptotic distributions of the random variable by $\mathrm{pr}^{\mathbb{G}^*_x(t)}$ and of the supremum by $\mathrm{pr}^{\mathbb{G}^*_{x,M}}$. Here, $\sigma^2_{\mathbb{G}_{x_*}}(t)$ is the variance of $\hat{h}_{x_*}(t) - h_{x_*}(t)$. The bootstrap estimators of these quantities are given by the conditional distribution $\mathrm{pr}^{h^*_{x_*,B}(t)}$ of

$$h^*_{x_*,B}(t) = \hat{\sigma}_{\mathbb{G}_{x_*}}(t)^{-1}\{A_B(t) + B_B(t)\},$$

and of the conditional distribution $\text{pr}^{h^*_{x*,B,M}}$ of $\sup_{\delta_T \le t \le T-\delta_T} |h^*_{x*,B}(t)|$. Here $\hat{\sigma}^2_{\mathbb{G}_{x*}}(t)$ is an estimator of the variance of $\hat{h}_{x*}(t)$ and is equal to $E\{\bar{h}_{x*,B}(\cdot)^2 \mid \mathbb{X}_n\}$. In the proof of Theorem 2 below, which is provided in Section B.8 of the supplementary material, we argue that $\hat{\sigma}^2_{\mathbb{G}_{x*}}(t)$ converges to the variance $\sigma^2_{\mathbb{G}_{x*}}(t)$ of the limiting Gaussian process. The consistency of the bootstrap approach is established in the next theorem.

THEOREM 2. *Under the assumptions of Theorem* 1, $h^*_{x*,B}(t)$ *approximates the distribution of* $\mathbb{G}^*_{x*}(t)$ *pointwise and uniformly, i.e. in the pointwise case it holds that*

$$d_K\left(pr^{h^*_{x*,B}(t)}, pr^{\mathbb{G}^*_{x*}(t)}\right) \to 0,$$

in probability, for all $t \in (0, T)$ *and in the uniform case it holds that*

$$d_K\left(pr^{h^*_{x*,B,M}}, pr^{\mathbb{G}^*_{x*,M}}\right) \to 0,$$

in probability, where $d_K(\cdot,\cdot)$ *denotes the Kolmogorov distance.*
Implementation details for the $1-\alpha$ pointwise bootstrap confidence intervals and uniform bootstrap confidence bands are provided in Section A.1 of the supplementary material.

5. NUMERICAL EXAMPLES

5.1. *Simulation study*

The simulation study in this section assesses the performance of the proposed estimator and the accuracy level of the associated confidence sets. We generated data according to the High-Quality Marker (HQM) model described in Section 2 with a univariate marker. We considered 3 different functional shapes for the marker-only hazard $\alpha_1(x) = \exp(2x-2)/15$, $\alpha_2(x) = 4(x-0.3)^4$ and $\alpha_3(x) = 4[1+\exp\{-4(x-1)\}]^{-1}$, displayed in Fig. 5 in the supplement. These hazards are only dependent on a marker measurement x and thus satisfy Assumption [A1]. The marker X is simulated in the following way: first, we simulate a Gaussian random walk $G(t_g)$ with jump sizes that vary according to a normal distribution $N(0, 0.07^2)$ on the grid $\{t_d/10 : t_d = 0, \ldots, 100\} \ni t_g$ and with uniform randomly chosen starting points $G(0) = x_0 = 0.1, 0.2, \ldots, 0.9$. The resulting continuous process, marker $X(t)$ for all $t \in [0, 10]$, was derived via linear interpolation between the values $G(t_g)$. To mimic cohort follow-ups, we then considered that this marker was only observed at individual-specific discrete follow-up visits $\tilde{t} = a + D$, where $a = 1, \ldots, 9$ and $D \sim N(0, 0.07^2)$. Note that since the Gaussian random walk is a homogeneous Markov process, Assumption [A2] is also satisfied.

Given the hazard $\alpha_j, j = 1, 2, 3$ and the marker X, we then simulate n independent one-jump counting processes, each with intensity $\alpha_j\{X(t)\}$ before a jump and 0 after. If a jump did not happen before 10 the survival time T was right censored by 10.

We evaluated 6 settings based on samples of size $n = 300, 600$ and the 3 marker-only hazards. In each setting we calculated 1000 realizations of the estimator $\hat{h}_x(t)$ defined in (3) together with local and uniform confidence bands, see also (A1) and (A2) in the supplementary material. This was done for marker values x equal to $q_{0.1}, q_{0.25}, q_{0.5}, q_{0.75}, q_{0.9}$, where q_z is the empirical z quantile of all observed marker values. We compared $\hat{h}_x(t)$ with the true hazard $h_x(t)$ defined in (1). Since the true hazard $h_x(t)$ does not have a closed form, it was numerically approximated using the relation $h_x(t) = \partial/\partial t\{-\log S_x(t)\}$, where $S_x(t)$ was approximated by simulations. The bandwidths were set to $b_1 = b_2 = b$ with b chosen as minimizer of Mean Integrated Squared

Error (MISE)

$$\mathrm{MISE}(h_x, \hat{h}_{x,b,b}) = E\Big[\int_0^{10} \{\hat{h}_x(t) - h_x(t)\}^2 \mathrm{d}t\Big].$$

Different values of b are selected for each combination of x, n, and α. We did not simulate the cross-validation bandwidths as proposed in Section A.2 of the supplementary material because, due to the large number of settings, such a calculation would be computational too complex.

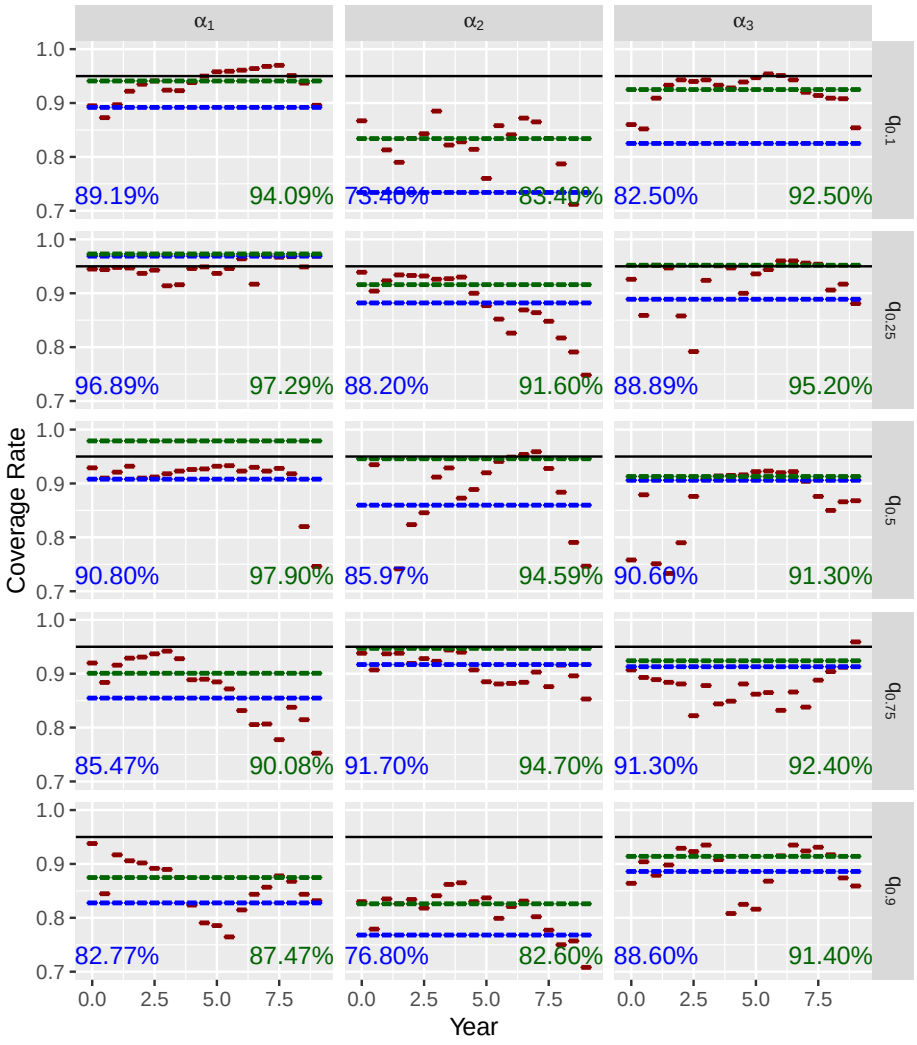


Fig. 1. Coverage rate of the pointwise (red) and uniform (blue) confidence bands for the marker-only hazards α_j, $j = 1, 2, 3$, for $n = 300$ individuals and marker values x at the 0.1, 0.25, 0.50, 0.75, and 0.90 quantile of the empirical marker distribution. The exact rates for the uniform confidence band computed on the total time window (blue) or computed while ignoring the first and last year (green) are also reported. The confidence bands are based on 1000 realizations and a bootstrap with 1000 repetitions.

Under the aforementioned settings we performed extensive simulations to quantify the relative bias of $\hat{h}_x(t)$ and the coverage rates of the associated pointwise and uniform confidence bands. Due to space restrictions, graphical illustrations of the results are deferred to Section A.3 of the supplementary material. In summary, the simulations indicate that the methodology generally provides a very small bias, which decreases as the sample size increases. Furthermore, the standard deviation is much larger compared to the bias, which aligns with the theoretical findings. Nevertheless, for marker-only hazards with a very low value the relative bias is substantial. This is not surprising, as a low marker-only hazard value leads to a smaller number of observed events which do not allow for accurate estimations. Increased bias is also observed at boundary values of time, e.g. at the first or last year which affect the coverage rates of the 95% pointwise and uniform confidence bands as can indicatively be seen in Fig.1. The figure suggests that boundary effects in combination with small marker-only hazards values increase the coverage error of the proposed technique. In Section A.3 of the supplementary material we argue that increasing the sample size slightly improves the coverage error (see Fig.10 there); nevertheless effective treatment of the issue necessitates the use of a local linear estimator, i.e., estimator $\tilde{h}_x(t)$ defined in Section B.1 of the supplement.

We note that additional simulations were performed for a smaller sample size of $n = 50$ patients. The estimator captured the key elements of the marker-only hazard and gave reasonable predictions but the quality of inference was not perfect (results not shown). In the perspective of proposing prediction tools, we encourage the use of large enough sample sizes as recommended in the literature (e.g., Riley et al. (2020)). Indeed, too small samples may provide inaccurate individual predictions in particular in the case of complex relationships with the marker.

5.2. *Comparison with the Joint Modeling methodology*

In biomedical research, the JM methodology of the longitudinal marker and the time-to-event have been proposed to compute dynamic individual predictions (Rizopoulos, 2012; Ferrer et al., 2019). Here we compare the HQM estimator and JM estimators when data are generated according to the framework described in Section 5.1. The comparison is based on the values of the model's estimated MISE, Area Under the Curve (AUC) and Brier Score (BS) metrics. The future conditional survival function (given the information $\mathcal{F}_s$ up to s) is estimated by

$$S_\star^{\text{TRUE}}(s+t \mid \mathcal{F}_s) = \text{pr}\{T_\star > s+t \mid \mathcal{X}_\star(s), T_\star > s\}, \tag{6}$$

where $\mathcal{X}_\star(s)$ is the marker history of an individual $\star$ before landmark time s.

The JM estimator consists of two parametric submodels which are linked by a shared latent structure. We consider here the classical specification with a linear mixed model for the longitudinal marker measurements and a proportional hazard model for the time-to-event:

$$\begin{cases} X_i(t) & = m_i(t) + \varepsilon_i(t) \\ & = F_i(t)^\top \beta + R_i(t)^\top B_i + \varepsilon_i(t) \\ \lambda_i(t) & = \lambda_0(t) \exp\{W_i^T \gamma + m_i(t)\eta\}, \end{cases} \tag{7}$$

where $F_i(t)$ and $R_i(t)$ are covariate vectors including functions of time, that are associated with the vector of fixed effects $\beta \in \mathbb{R}^p$ and the vector of individual random effects $B_i \sim N(0, D)$ with unknown D, respectively. The measurement errors $\varepsilon_i(t) \sim N(0, \sigma^2)$ and are independent of B_i. The instantaneous risk of event is defined according to the baseline hazard λ_0, and a linear predictor that includes covariates W_i and a function of the marker trajectory, in this example $m_i(t)$ the underlying true current marker level of individual i. In this specific setting, the baseline hazard λ_0 is approximated by cubic splines with 5 internal knots, and considered

a linear regression with time, i.e. $F_i(t) = t, R_i(t) = t$, and no adjustment for covariates, i.e., $W_i = 0$. The joint model is fitted on all the longitudinal and survival information within the maximum joint likelihood framework using the `JM` `R` package (Rizopoulos, 2010). The numerical integration involved in the log-likelihood computation is performed by a pseudo-adaptive Gauss Hermite quadrature with 9 knots, and optimization of the log-likelihood was achieved by an Expectation-Maximization algorithm. See Rizopoulos (2012), Ferrer et al. (2019) for further details.

The two methodologies were trained on samples of $n = 300, 600$ individuals and the marker-only hazards $\alpha_1, \alpha_2, \alpha_3$, corresponding to a total of 6 scenarios. Their predictive performance was then compared on an external sample of 100 new individuals. Once the JM was fitted, the future conditional survival function (6) was computed at different landmark times s and for different horizons t, using the parametric fit described above. For the HQM estimator we used:

$$\hat{S}_\star^{\text{HQM}}(s + t \mid \mathcal{F}_s) = \exp\left\{ - \int_0^t \hat{h}_{x_\star}(u) du \right\} \tag{8}$$

where $x_\star$ is the last marker measurement $X_\star(s)$ of an individual $\star$ before landmark time s. Note that since the marker is Markov and the marker-only hazards do not directly depend on time, the future conditional hazard $h_x(t)$ is independent of the landmark time s. For comparing the predictive performances of the two approaches, we considered 3 different landmark times 1.5, 3.5 and 5.5 leading to 18 cases. For each case, we computed 100 realizations of the conditional survival estimate of each individual $\star$ based on their marker data up to the landmark time s. Implementation details for all three metrics are provided in Section A.3 of the supplementary material, along with graphical illustration of all simulation results.

In summary, in all cases of the MISE and the AUC simulations (36 instances in total) the HQM model systematically outperformed the JM approach. This is probably due to the parametric specification of the JM which assumes a log linear form for the hazard that is not compatible with the α_2 and α_3 cases. For α_1, although both approaches yield close MISEs, the HQM model performs slightly better than JM in all instances. This might be due to the misspecification of parametric model for the marker which was generated as a random walk rather than a individual-specific linear trajectory. The HQM model is better than JM in the Brier score simulations too; however, in this case, their BS metric values are very close to each other and especially for α_2 they can be regarded as equivalent.

6. Application

The methodology developed in the previous Sections is applied for predicting the clinical progression of patients diagnosed with primary biliary cirrhosis (PBC) of the liver, based on the publicly available dataset pbc2 of the Mayo Clinic (Therneau & Grambsch, 2000). The data set contains information on a randomized clinical trial of the D-penicillamine versus placebo for a total of 312 patients who met certain eligibility criteria. The patients were followed-up for a maximum of 13 years between 1974 and 1984. Repeated measures of different biological markers were collected over time including albumin, bilirubin and alkaline phosphatase along with time to clinical progression defined as the minimum date of death or transplantation.

We consider two high-quality markers for the prediction of clinical progression in PBC. These are bilirubin (normal range 0.2 to 1.2 mg/dl) and albumin (normal range of 3.4 to 5.4 g/dl). Both are produced by the liver and according to Lammers et al. (2015); Hirschfield et al. (2018) are known to correlate with death. High concentration of bilirubin and low concentration of albumin indicates a liver dysfunction. Under assumption [A1] the risk of clinical progression is indepen-

dent of time given the current marker level. Assumption [A2] expresses the fact that the future level of either marker depends on its current level and the time elapsed in between the patient's evaluations. For our methodology to be applicable, we need (a) to assume the markers are observed without measurement error, and (b) to interpolate between the marker measurements and extrapolate linearly at the last measurement. This is necessary as we assume the marker X to be continuous. More details are discussed in Fusaro et al. (1993) and Nielsen (1999). Furthermore, we assume that the exposure Z for patient i is $Z_i(s) = I(s < t_i)$, where t_i is the time of clinical progression of patient i and $I(\cdot)$ is the indicator function.

Our model prediction based on each marker is illustrated in Figs. 2 and 3 with 4 different marker values: 1 mg/dl, 4 mg/dl, 7 mg/dl, 10 mg/dl for bilirubin, and 2 g/dl, 3 g/dl, 4 g/dl, 5 g/dl for albumin. The values are chosen evenly based on the range of around 90% of the respective marker data, which is 0.1 mg/dl - 10 mg/dl for bilirubin and 1.5 g/dl - 5 g/dl for albumin. Both, the future conditional hazard and the future conditional survival function are predicted for 10 years with the 95% pointwise and uniform confidence bands. In each example, $h_x(t)$ is estimated by $\hat{h}_{x,b_1,b_2}(t)$, given in (3) with bandwidths determined by cross-validation; full details on obtaining the bandwidths for each example are provided at the end of Section A.2 in the supplementary material. The predicted hazard of clinical progression and the progression-free survival does not seem to change with the bilirubin level. For albumin though, we highlight a higher risk of clinical progression in patients with lower concentrations of albumin. For instance, the predicted 5-year progression-free survival decreases from approximately 0.85 for an albumin concentration of 5 g/dL to 0.48 for an albumin concentration of 2 g/dL. Note, that because right censoring is present the study is not completely run off and thus, the survival function does not go all the way to zero.

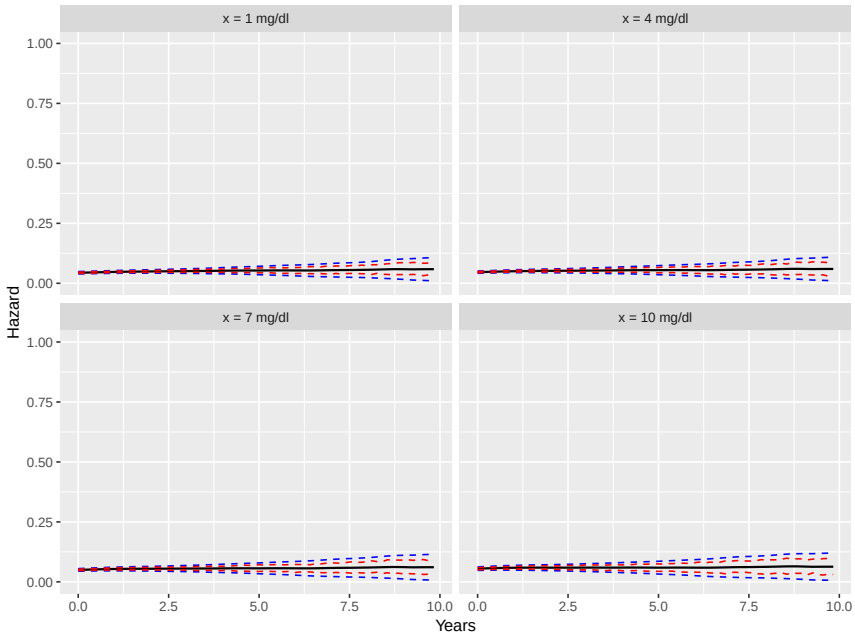
For the albumin marker, the first level considered, $x = 2$g/dl, corresponds to the first percentile of the marker's distribution and therefore lies in the boundary. Consequently, the confidence bands of the estimator are wider than those of the rest marker levels, due to the underperformance of the local constant estimator in the boundary. This issue has also been demonstrated in the coverage rate simulations of Section A.3 in the supplementary material where it is also exhibited that boundary correction remedies this deficiency. Finally, see Section A.4 in the supplement for a simultaneous application of the HQM and JM methodologies using the pbc2 dataset.

7. CONCLUSION

This paper is a first piece of work of integrating the superefficient approach into mainstream theory of mathematical statistics. The estimation procedure along with confidence bands and a bandwidth selection rule is implemented in the R package HQM, Bagkavos et al. (2022) together with an estimation example based on time-invariant covariates. We showed in an illustration and a comparative simulation study that our new method is already of practical use in some important applied problems such as the dynamic individual prediction of health events. However, further extensions are needed so that the method becomes fully operational in applied statistics.

First, an immediate and unproblematic extension would be to allow for categorical variables driving either the marker process or the marker dependent hazard or both. Second, extension of the present methodology to multidimensional markers may help in some contexts achieving the high-quality property, i.e., the assumption that the hazard only depends on the marker process. However this would require a quite different and much more complex mathematical theory which needs a very careful modeling of the marker process. An interesting alternative where a large part of the mathematical theory of this paper could be used would be based on one-dimensional marker indices. In such a approach one firstly reduces the markers into a one-dimensional composite indicator defined as a weighted sum of the markers where weights are

(a) Bilirubin - Hazard



(b) Bilirubin - Survival

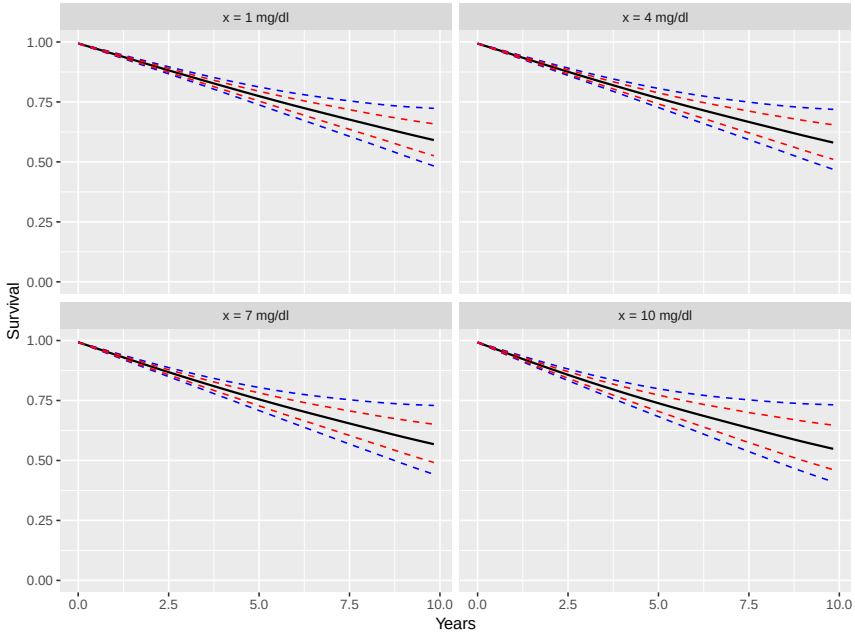
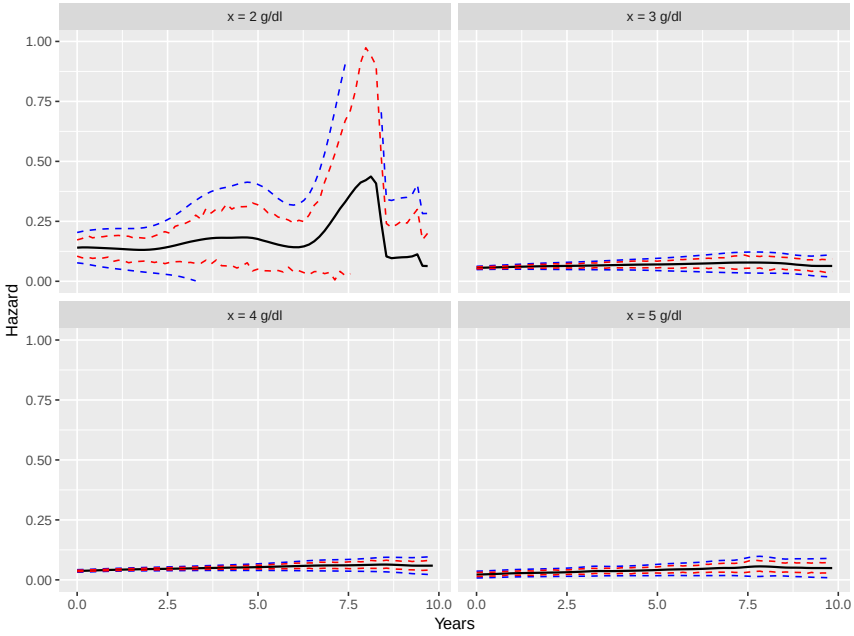


Fig. 2. HQM estimator (black) with pointwise (red) and uniform (blue) confidence bands for the future, conditional on bilirubin, (a) hazard and (b) survival function for the next 10 years.

internally estimated, and secondly apply the superefficient methodology as described here to the one-dimensional composite indicator. This would thus lead to a semi-parametric extension of

(a) Albumin - Hazard



(b) Albumin - Survival

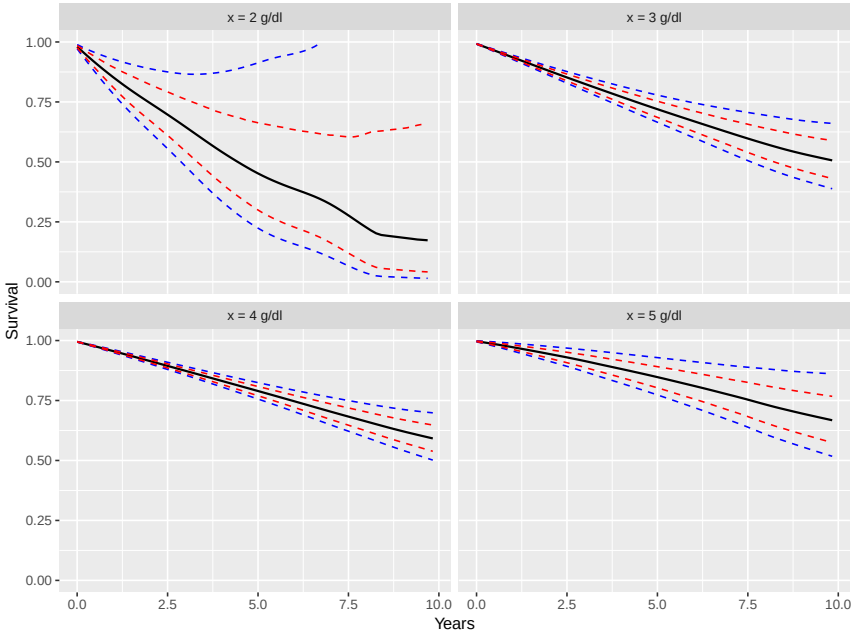


Fig. 3. HQM estimator (black) with pointwise (red) and uniform (blue) confidence bands for the future, conditional on albumin, (a) hazard and (b) survival function for the next 10 years.

445 the superefficient model for multidimensional marker. Another semiparametric extension multiplies the fully nonparametric marker only hazard model by a parametric dependency on time.

Extending our modelling framework to such a semiparametric marker dependent hazard would still result in a parametric rate of convergence allowing uniform confidence bands based on transformed Gaussian processes.

Finally, we assumed that the marker was measured without error and in continuous time, the latter assumption being achieved by interpolation. Extending the methodology to handle noisy and intermittently missing marker observations, as done in the parametric JM approach with an underlying latent marker process (Rizopoulos, 2012), is also one direction of future research.

SUPPLEMENTARY MATERIAL

The supplementary material available at Biometrika online includes additional simulation results and illustrations, discussion on the assumptions, auxiliary lemmas and the proofs of Theorems 1 and 2.

REFERENCES

AECKERLE-WILLEMS, C. & STRAUCH, C. (2022). Sup-norm adaptive simultaneous drift estimation for ergodic diffusions. *The Annals of Statistics* **50**, 3484–3509.

ANDERSON, J. R., CAIN, K. C. & GELBER, R. D. (1983). Analysis of survival by tumor response. *J Clin Oncol* **1**, 710–719.

BAGKAVOS, D., ISAKSON, A., PROUST-LIMA, P., MAMMEN, E. & NIELSEN, J. P. (2022). *HQM: Superefficient Estimation of Future Conditional Hazards Based on Marker Information*. R package version 0.1.1.

BISCHOFBERGER, S. M., HIABU, M., MAMMEN, E. & NIELSEN, J. P. (2019). A comparison of in-sample forecasting methods. *Computational Statistics & Data Analysis* **137**, 133–154.

BOSQ, D. (2012). *Nonparametric statistics for stochastic processes: estimation and prediction*, vol. 110. Springer Science & Business Media.

CASTELLANA, J. & LEADBETTER, M. (1986). On smoothed probability density estimation for stationary processes. *Stochastic Processes and their Applications* **21**, 179–193.

CUI, Y. et al. (2023). Instrumental variable estimation of the marginal structural cox model for time-varying treatments. *Biometrika* **110**, 101–118.

FERRER, L., PUTTER, H. & PROUST-LIMA, C. (2019). Individual dynamic predictions using landmarking and joint modelling: validation of estimators and robustness assessment. *Statistical methods in medical research* **28**, 3649–3666.

FUSARO, R. E., NIELSEN, J. P. & SCHEIKE, T. H. (1993). Marker-dependent hazard estimation: An application to AIDS. *Statistics in Medicine* **12**, 843–865.

HENDERSON, R., DIGGLE, P. & DOBSON, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480.

HIABU, M., MAMMEN, E., MARTINEZ-MIRANDA, M. D. & NIELSEN, J. P. (2016). In-sample forecasting with local linear survival densities. *Biometrika* **103**, 843–859.

HIRSCHFIELD, G. M., DYSON, J. K., ALEXANDER, G. J. et al. (2018). The British Society of Gastroenterology/UK-PBC primary biliary cholangitis treatment and management guidelines. *Gut* **67**, 1568–1594.

HJORT, N. L. (1994). Dynamic likelihood hazard rate estimation, *Preprint series. Statistical Research Report, University of Oslo. https://www.duo.uio.no/bitstream/handle/10852/47750/1993-4.pdf* .

JEWELL, N. P. & NIELSEN, J. P. (1993). A framework for consistent prediction rules based on markers. *Biometrika* **80**, 153–164.

KUTOYANTS, Y. A. (2004). *Statistical inference for ergodic diffusion processes*. Springer Science & Business Media.

LAMMERS, W. J., HIRSCHFIELD, G. M. et al. (2015). Development and validation of a scoring system to predict outcomes of patients with primary biliary cirrhosis receiving ursodeoxycholic acid therapy. *Gastroenterology* **149**, 1804–1812.

MAMMEN, E. (1992). Bootstrap, wild bootstrap, and asymptotic normality. *Probability Theory and Related Fields* **93**, 439–455.

MAMMEN, E., MARTÍNEZ-MIRANDA, M. D., NIELSEN, J. P. & VOGT, M. (2021). Calendar effect and in-sample forecasting. *Insurance: Mathematics and Economics* **96**, 31–52.

MAMMEN, E. & NIELSEN, J. P. (2007). A general approach to the predictability issue in survival analysis with applications. *Biometrika* **94**, 873–892.

MARTÍNEZ-MIRANDA, M. D., NIELSEN, J. P., SPERLICH, S. & VERRALL, R. J. (2013). Continuous chain ladder: Reformulating and generalising a classical insurance problem. *Expert Systems with Applications* **40**, 5588–5603.

NIELSEN, J. P. (1999). Super-efficient hazard estimation based on high-quality marker information. *Biometrika* **86**, 227–232.

NIELSEN, J. P. (2000). Super-efficient prediction based on high-quality marker information. *ASTIN Bulletin: The Journal of the IAA* **30**, 295–303.

NIELSEN, J. P. & LINTON, O. B. (1995). Kernel estimation in a nonparametric marker dependent hazard model. *The Annals of Statistics* **23**, 1735–1748.

PROUST-LIMA, C. & TAYLOR, J. M. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics* **10**, 535–549.

RILEY, R. D., ENSOR, J. et al. (2020). Calculating the sample size required for developing a clinical prediction model. *British Medical Journal* **368**, m441.

RIZOPOULOS, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software* **35**, 1–33.

RIZOPOULOS, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. Chapman and Hall/CRC.

SURESH, K., TAYLOR, J. M., SPRATT, D. E., DAIGNAULT, S. & TSODIKOV, A. (2017). Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model. *Biometrical Journal* **59**, 1277–1300.

THERNEAU, T. M. & GRAMBSCH, P. M. (2000). The Cox model. In *Modeling survival data: extending the Cox model*. Springer, pp. 39–77.

VAN HOUWELINGEN, H. C. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics* **34**, 70–85.

YONG, F. H., TAYLOR, J. M., BRYANT, J. L., CHMIEL, J. S., GANGE, S. J. & HOOVER, D. (1997). Dependence of the hazard of AIDS on markers. *AIDS* **11**, 217–228.

SUPPLEMENTARY MATERIAL FOR "SUPEREFFICIENT ESTIMATION OF FUTURE CONDITIONAL HAZARDS BASED ON TIME-HOMOGENEOUS HIGH-QUALITY MARKER INFORMATION"

BY D. BAGKAVOS, A. ISAKSON, E. MAMMEN, J.P. NIELSEN AND C. PROUST–LIMA

SUMMARY

The supplementary material in Section A includes description of the bootstrap construction of the pointwise and uniform confidence intervals in Section A.1 and description of the Cross-Validation bandwidth selector in Section A.2. Section A.3 contains additional graphical illustrations and simulation results. A comparison of the HQM and Joint Modeling approaches for the pbc2 dataset in contained in Section A.4. Section B of the supplementary material includes basic definitions in Section B.1, a discussion on the validity of Assumptions [A1] and [A2] in Section B.2 and a discussion on assumptions [A4]–[A7] in Section B.3. The proof of Theorem 1 starts with an outline of the proof in Section B.4 and proceeds with auxiliary results, definitions and the actual proof in Sections B.5–B.7. Finally, the proof of Theorem 2 is provided in Section B.8.

A. CONFIDENCE BANDS AND BANDWIDTH SELECTION

A.1. Confidence bands

The development and implementation of the pointwise bootstrap confidence intervals and the uniform bootstrap confidence bands of Section 4, relies on the fact that we can replace the quantiles of the distribution of $\mathbb{G}^*_{x_*}$ with the ones of the conditional distribution of $h^*_{x_*,B,M}$. The pointwise confidence band is a confidence interval for a given $t \in (0,T)$, while the uniform one requires that the true value $h(t)$ lies in the band for all $t \in [\delta_T, T-\delta_T]$. For the construction of the latter, the supremum is usually used. This is also the reason why uniform confidence bands tend to be a little bit wider than their pointwise counterparts. An outline of the bootstrap construction for both cases is as follows:

(a) Pointwise: for a given $t \in (0,T)$ generate ${h^*}^{(1)}_{x_*,B}(t), \ldots, {h^*}^{(N)}_{x_*,B}(t)$ for $N = 1000$ and order it ${h^*}^{[1]}_{x_*,B}(t) \leq \cdots \leq {h^*}^{[N]}_{x_*,B}(t)$. Then

$$\hat{I}^1_{n,N} = \left[\hat{h}_{x_*}(t) - \hat{\sigma}_{\mathbb{G}_{x_*}}(t) n^{-1/2} {h^*}^{[N(1-\frac{\alpha}{2})]}_{x_*,B}(t), \hat{h}_{x_*}(t) - \hat{\sigma}_{\mathbb{G}_x}(t) n^{-1/2} {h^*}^{[N\frac{\alpha}{2}]}_{x_*,B}(t)\right] \tag{A1}$$

is a $1-\alpha$ pointwise confidence band for $h_{x_*}(t)$.

(b) Uniform: generate $\bar{h}^{(1)}_{x_*,B}(t), \ldots, \bar{h}^{(N)}_{x_*,B}(t)$ for $N = 1000$ for all $t \in [\delta_T, T-\delta_T]$ and define $W^{(i)} = \sup_{t\in[0,T]} |\bar{h}^{(i)}_{x_*,B}(t)|$ for $i = 1, \ldots, N$. Order $W^{[1]} \leq \cdots \leq W^{[N]}$. Then

$$\hat{I}^2_{n,N} = \left[\hat{h}_{x_*}(t) \pm \hat{\sigma}_{\mathbb{G}_{x_*}}(t) n^{-1/2} W^{[N(1-\alpha)]}\right] \tag{A2}$$

is a $1-\alpha$ uniform confidence band for $h_{x_*}(t)$.

A.2. Bandwidth Selection

Bandwidth selection is based on cross-validation (CV) throughout. For the classical approach, see Härdle & Marron (1985) and for a bandwidth selection for hazards, see Gámiz et al. (2016). We use a similar idea here. Let

$$Q(b_1, b_2) = \sum_{i=1}^{N} \int_0^T \int_s^T Z_i(t) Z_i(s) \left\{\hat{h}_{X_i(s),b_1,b_2}(t-s) - h_{X_i(s)}(t-s)\right\}^2 dtds$$

be our measure of deviation. Let $Q = Q_1 + Q_2 + Q_3$, where

$$Q_1(b_1, b_2) = \sum_{i=1}^{N} \int_0^T \int_s^T Z_i(t) Z_i(s) \hat{h}^2_{X_i(s),b_1,b_2}(t-s) dt ds,$$
$$Q_2(b_1, b_2) = -2R(b_1, b_2),$$

with

$$R(b_1, b_2) = \sum_{i=1}^{N} \int_0^T \int_s^T Z_i(t) Z_i(s) \hat{h}_{X_i(s),b_1,b_2}(t-s) h_{X_i(s)}(t-s) dt ds,$$

and Q_3 is

$$Q_3 = \sum_{i=1}^{N} \int_0^T \int_s^T Z_i(t) Z_i(s) h^2_{X_i(s)}(t-s) dt ds.$$

Note that Q_3 does not depend on the bandwidths and thus, can be ignored, while Q_1 is known. To estimate R, we rewrite it using

$$h_{X_i(s)}(t-s) Z_i(t) dt = d\Lambda_i(t).$$

We get

$$\begin{aligned} R(b_1, b_2) &= \sum_{i=1}^{N} \int_0^T \int_s^T Z_i(s) \hat{h}_{X_i(s),b_1,b_2}(t-s) d\Lambda_i(t) ds \\ &= \sum_{i=1}^{N} \int_0^T \int_0^t Z_i(s) \hat{h}_{X_i(s),b_1,b_2}(t-s) ds d\Lambda_i(t) = \sum_{i=1}^{N} \int_0^T \hat{g}_{i,b_1,b_2}(t) d\Lambda_i(t), \end{aligned}$$

where

$$\hat{g}_{i,b_1,b_2}(t) = \int_0^t Z_i(s) \hat{h}_{X_i(s),b_1,b_2}(t-s) ds.$$

This motivates to estimate R by

$$\hat{R}(b_1, b_2) = \sum_{i=1}^{N} \int_0^T \hat{g}^{-i}_{i,b_1,b_2}(t) dN_i(t), \quad \hat{g}^{-i}_{i,b_1,b_2}(t) = \int_0^t Z_i(s) \hat{h}^{-i}_{X_i(s),b_1,b_2}(t-s) ds,$$

and $\hat{h}^{-i}$ is estimated without information from the ith counting process.

An alternative is K-fold cross-validation where I_j are sets of indices with $\cup_{j=1}^{K} I_j = \{1, \ldots, n\}$, $I_k \cap I_j = \emptyset$ and $|I_j| \approx |I_k|$ for all $j, k = 1, \ldots, K$. We use the estimator

$$\hat{R}_K(b_1, b_2) = \sum_{j=1}^{K} \sum_{i \in I_j} \int_0^T g^{-I_j}_{i,b_1,b_2} i(t) dN_i(t), \quad \hat{g}^{-I_j}_i(t) = \int_0^t Z_i(s) \hat{h}^{-I_j}_{X_i(s)}(t-s) ds,$$

and $\hat{h}^{-I_j}$ is estimated without information from all counting processes i with $i \in I_j$.

The proposed bandwidth selection rule implies that the bandwidths b_1 and b_2 required for calculating the hazard rate estimator $\hat{h}_{x,b_1,b_2}(t)$ via (3) (and subsequently $\hat{\alpha}_{i,b_1}(z)$ via (2)) are determined by solving

$$(b_1, b_2) = \arg\min_{b_1,b_2} CV(b_1, b_2),$$

where $CV(b_1, b_2) = Q_1(b_1, b_2) - 2\tilde{R}(b_1, b_2)$ with $\tilde{R} \in \{\hat{R}, \hat{R}_K\}$. In the examples presented in Figs. 2 and 3, implementation of the CV score function was performed with the K-fold cross-validation procedure with $K = 12$. Even though K-fold cross-validation is implemented with $K = 10$ as the default value, the

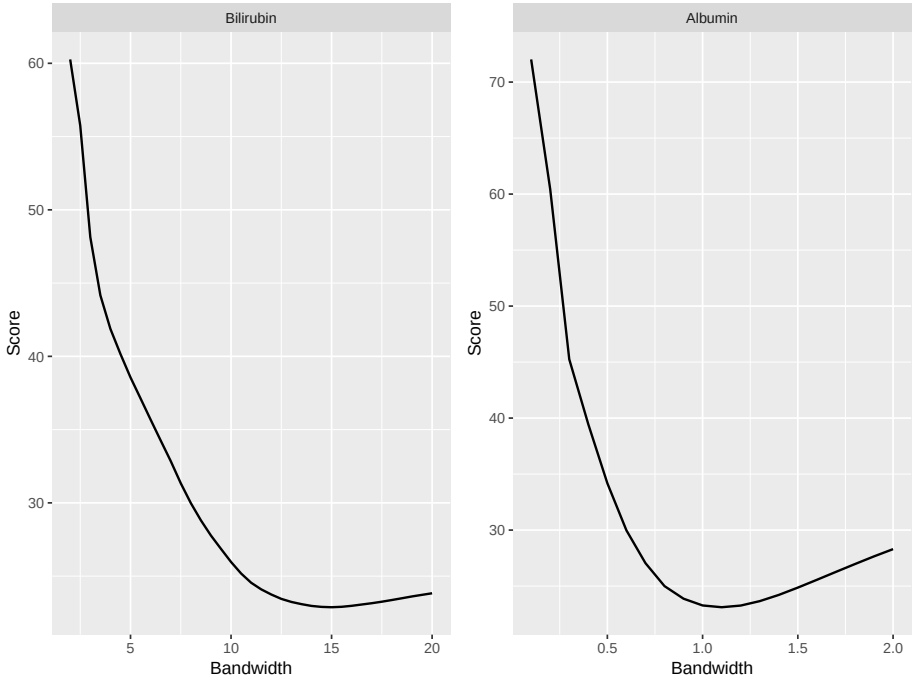


Fig. 4. Bandwidths scores $CV(b)$ based on the bandwidth b for albumin and bilirubin. The minimum is attained at $b = 15$ for bilirubin and $b = 1.1$ for albumin.

specific value is selected because the dataset contains 312 unique individuals, hence $K = 12$ corresponds to $312/12 = 26$ unique individuals (indices) in each fold with $I_k \cap I_j = \emptyset$ and $|I_j| = |I_k|$ for all $j, k = 1, \ldots, 12$. Because minimisation over a two-dimensional smoothing parameter may be unstable, we used $b_1 = b_2 = b$ with b determined by

$$b = \arg\min_b CV(b, b).$$

The $CV(b, b)$ scores for bilirubin and albumin are illustrated in Fig. 4 and suggest $b = 15$ for bilirubin and $b = 1.1$ for albumin. These are also the bandwidths employed in calculating the corresponding hazard rate (and hence survival function) estimators depicted in Figs. 2 and 3.

A.3. Additional graphical illustrations and simulation results

This section discusses in detail and illustrates graphically the simulation results that were outlined in Section 5.

Figs. 6–7 display the relative bias of our methodology. The boxplots are implemented with the default `R` options, i.e., the boxes are determined by the first and the third quartiles while the whiskers are implemented with `range=1.5`. Consequently they extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. As also noted in Section 5.1, the plots suggest that the methodology generally provides small amounts of bias which decreases as the sample size increases. Furthermore, the standard deviation is much larger compared to the bias, which aligns with the theory. However, the simulation results for $(\alpha_1, q_{0.1}), (\alpha_1, q_{0.25}), (\alpha_2, q_{0.5}), (\alpha_3, q_{0.1})$ and $(\alpha_3, q_{0.25})$ reveal a substantial relative bias. This is a consequence of the small marker-only hazards value for these cases, which results to a smaller number of observed events and hence to absence of enough information for accurate estimations. Larger bias values can also be seen at the boundaries of the time domain, i.e. for the first or last year in Figs. 6 and 7.

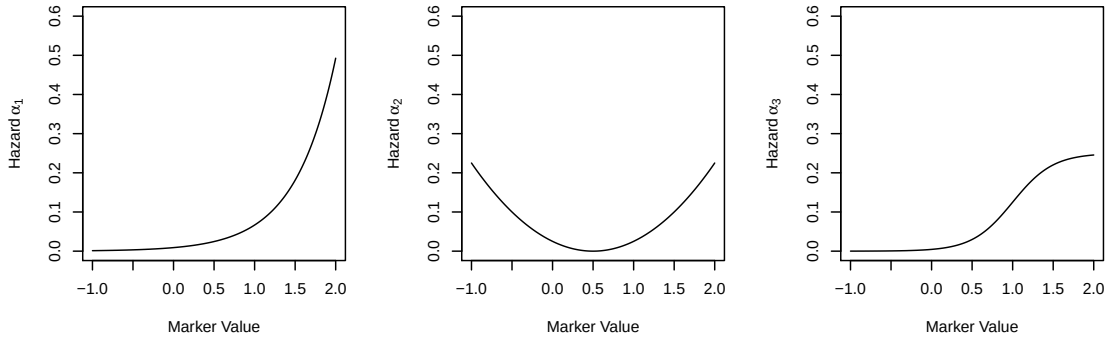


Fig. 5. True marker-only hazards generated in the simulation studies: $\alpha_1(x) = \exp(2x-2)/15$, $\alpha_2(x) = 4(x-0.3)^4$ and $\alpha_3(x) = 4[1+\exp\{-4(x-1)\}]^{-1}$.

Both findings are also reflected in the visualisation of the estimation of the future conditional hazard function based by the HQM estimator in Figs. 8 and 9 as well as in Figs. 1 and 10 which illustrate the coverage rates of the 95% pointwise and uniform confidence bands. The figures suggest that the unmodified implementation of (A2), in combination with small marker-only hazards values, increases the coverage error of the proposed technique across time and result to low coverage rates as it is evident particularly for $(\alpha_2, q_{0.1}), (\alpha_2, q_{0.25}), (\alpha_2, q_{0.5}), (\alpha_1, q_{0.75})$ for $n = 300$. This behaviour slightly improves as the sample sizes increases to $n = 600$ but the biggest improvement comes after excluding the first and last years in the calculation of the pointwise confidence bands. The resulting coverage level (green dashes) is constant across time and closer to the nominal 95% than the level based on the full time range (red dashes) as can be seen in Figs. 1 and 10. A more effective approach is the implementation of (A2) after replacing $\hat{h}_{x_*}(t)$ with a local linear estimator such as estimator $\tilde{h}_x(t)$ defined in Section B.1 below and implemented in the HQM package. Indeed, even though not reported here, replicating the coverage rate simulations after this change resulted in pointwise confidence bands that achieve a coverage rate of the same level or even closer to the nominal level than the green lines of Figs. 1 and 10 and which is constant across time. Thus we conclude that the automatic treatment of edge effects by $\tilde{h}_x(t)$ largely improves the coverage rates in comparison to the local constant estimator $\hat{h}_x$.

Regarding the simulations of Section 5.2, for each combination of marker only hazard $\alpha_1, \alpha_2, \alpha_3$ and landmark time $s = 1.5, 3.5, 5.5$, and for each sample size $n = 300, 600$, Figs. 11–12 summarize as boxplots the results of the MISE comparison between the HQM and JM approaches. Similarly Figs. 13–14 contain the corresponding AUC results and Figs.15–16 report the corresponding Brier score boxplots for each method.

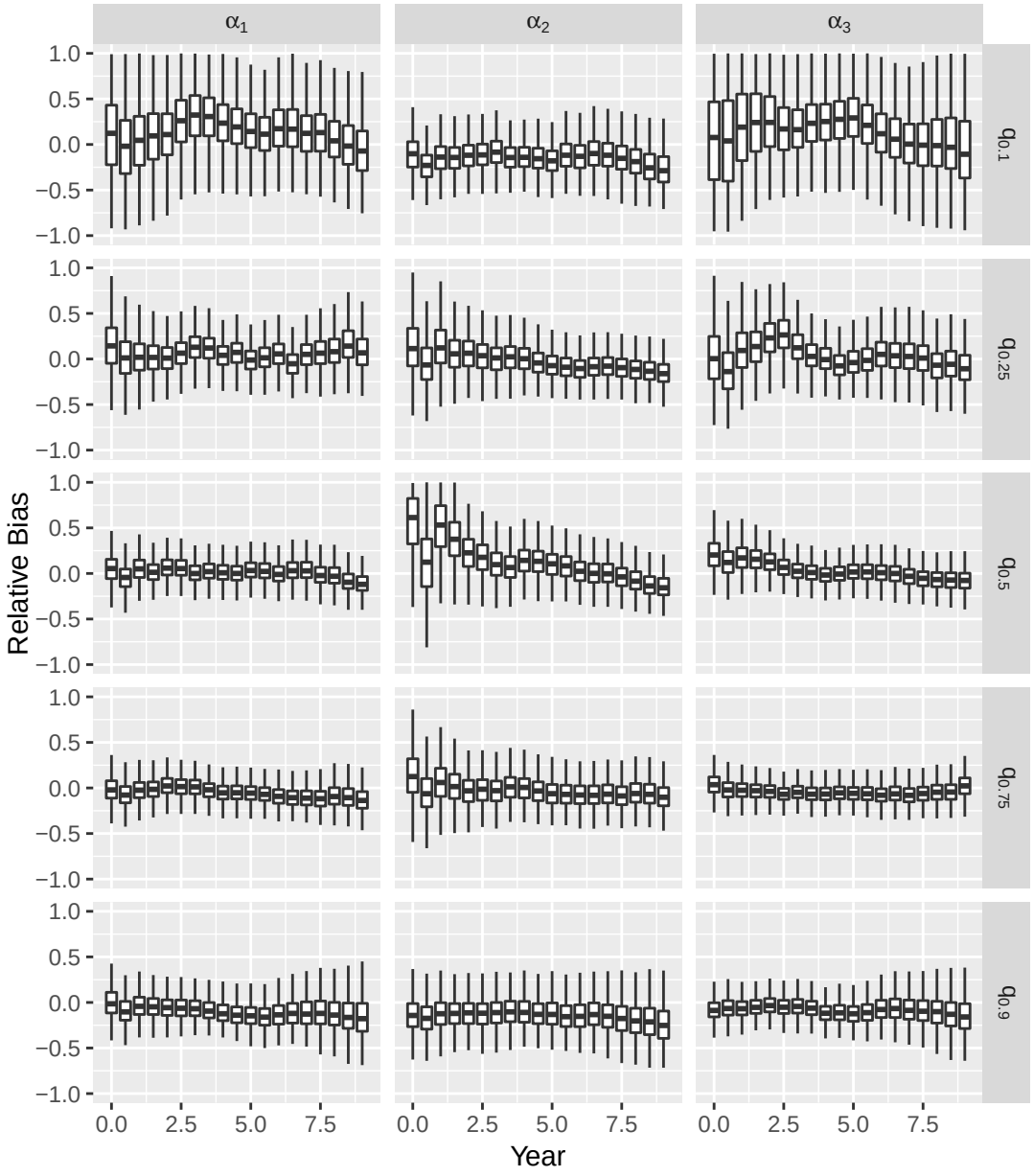


Fig. 6. Relative errors of $\hat{h}_x(t)$ based on 1000 realizations of $\{\hat{h}_x(t) - h_x(t)\}/h_x(t)$ at year $t = 0, 0.5, \ldots, 8.5, 9$ for the three different marker-only hazards α_j, $j = 1, 2, 3$, for $n = 300$ individuals and marker values x at the 0.1, 0.25, 0.50, 0.75, and 0.90 quantile of the empirical marker distribution.

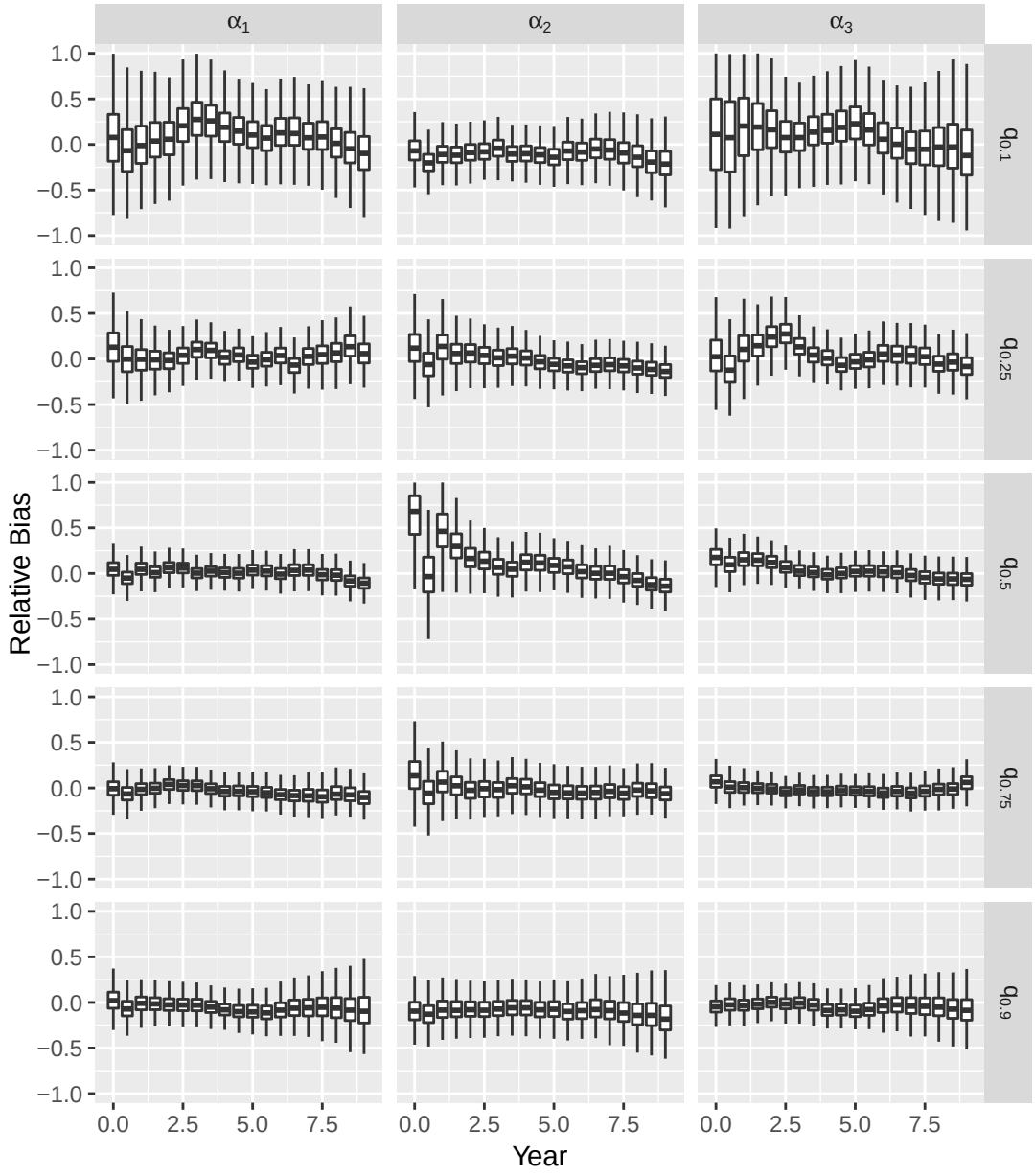


Fig. 7. Relative errors of $\hat{h}_x(t)$ based on 1000 realizations of $\{\hat{h}_x(t) - h_x(t)\}/h_x(t)$ at year $t = 0, 0.5, \ldots, 8.5, 9$ for the three different marker-only hazards α_j, $j = 1, 2, 3$, for $n = 600$ individuals and marker values x at the 0.1, 0.25, 0.50, 0.75, and 0.90 quantile of the empirical marker distribution.

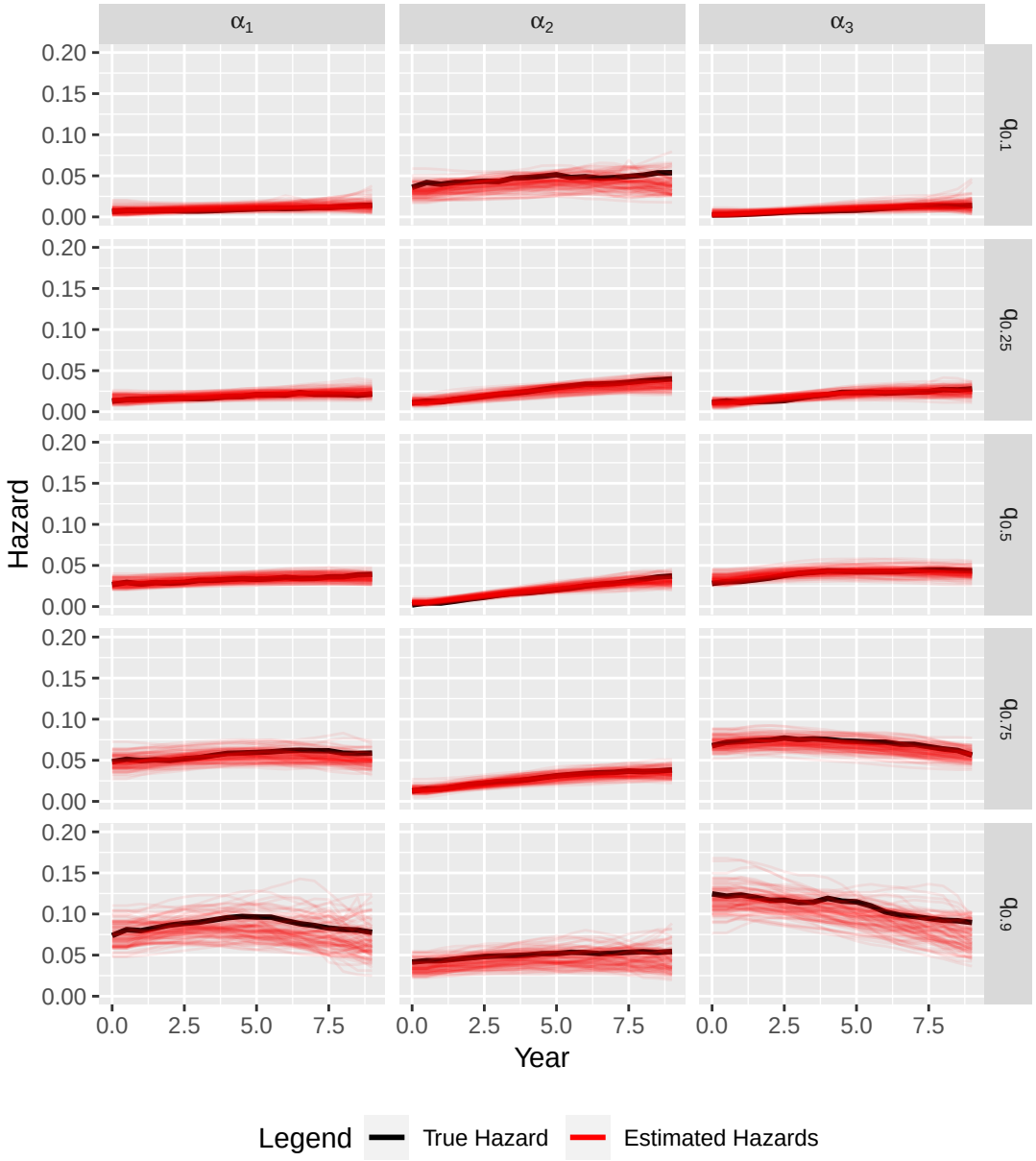


Fig. 8. Estimates of the future conditional hazard function based on 100 realizations of $\hat{h}_x(t)$ (red) together with the true future conditional hazard $h_x(t)$ (black) for the three different marker-only hazards α_j, $j = 1, 2, 3$, for $n = 300$ individuals and marker values x at the 0.1, 0.25, 0.50, 0.75, and 0.90 quantile of the empirical marker distribution.

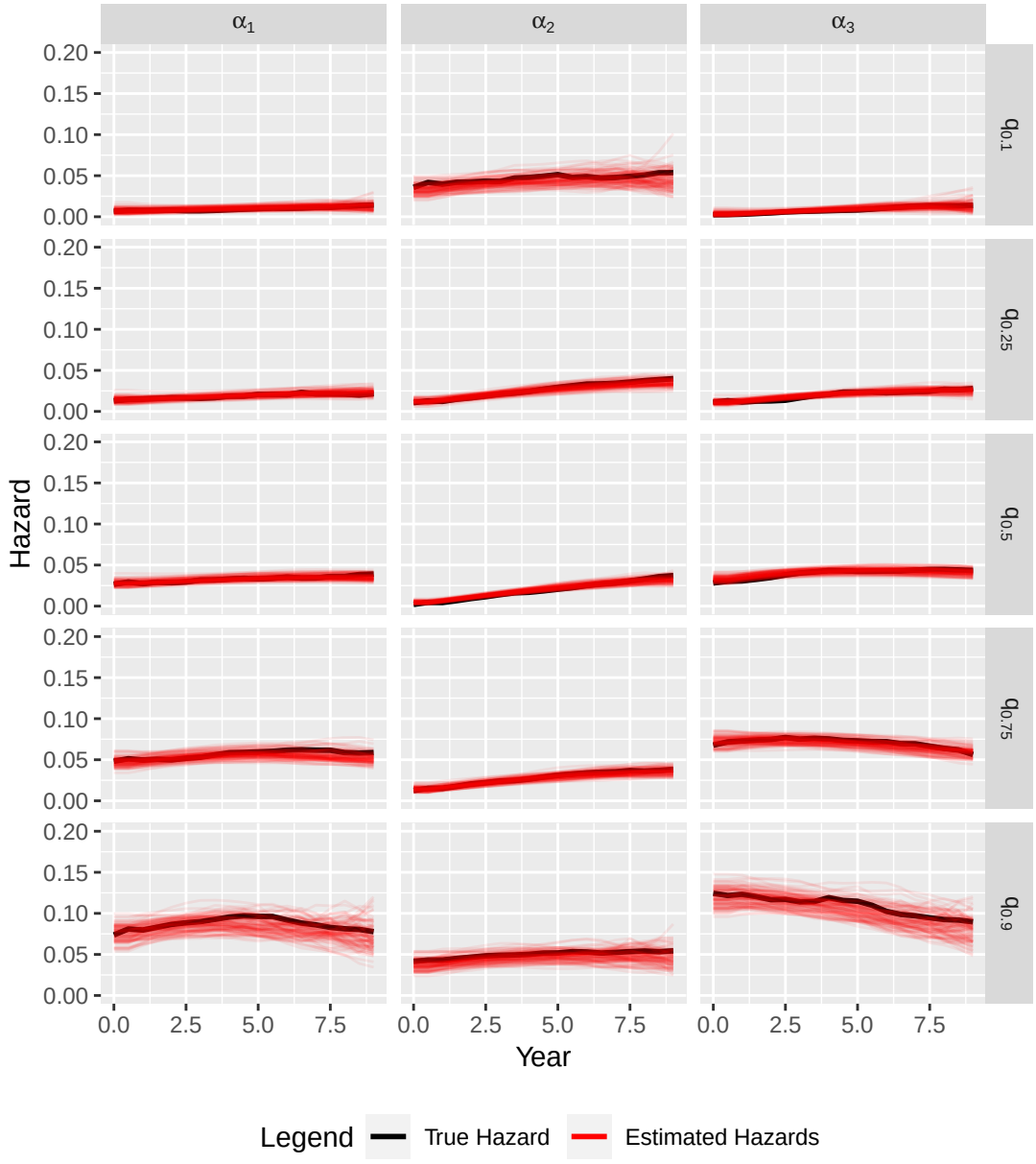


Fig. 9. Estimates of the future conditional hazard function based on 100 realizations of $\hat{h}_x(t)$ (red) together with the true future conditional hazard $h_x(t)$ (black) for the three different marker-only hazards α_j, $j = 1, 2, 3$, for $n = 600$ individuals and marker values x at the 0.1, 0.25, 0.50, 0.75, and 0.90 quantile of the empirical marker distribution.

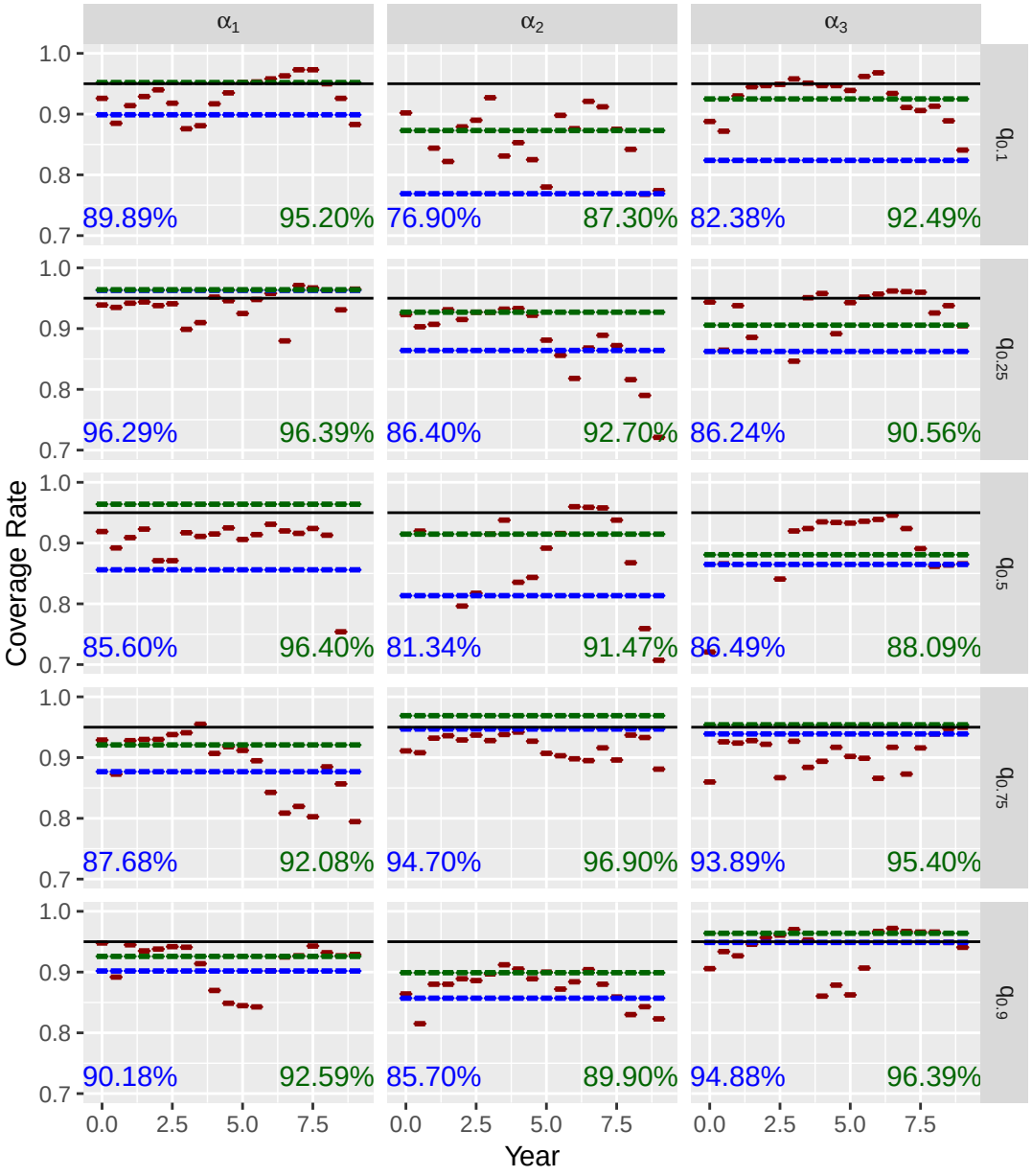


Fig. 10. Coverage rate of the pointwise (red) and uniform (blue) confidence bands for the marker-only hazards α_j, $j = 1, 2, 3$, for $n = 600$ individuals and marker values x at the 0.1, 0.25, 0.50, 0.75, and 0.90 quantile of the empirical marker distribution. The exact rates for the uniform confidence band computed on the total time window (blue) or computed while ignoring the first and last year (green) are also reported. The confidence bands are based on 1000 realizations and a bootstrap with 1000 repetitions.

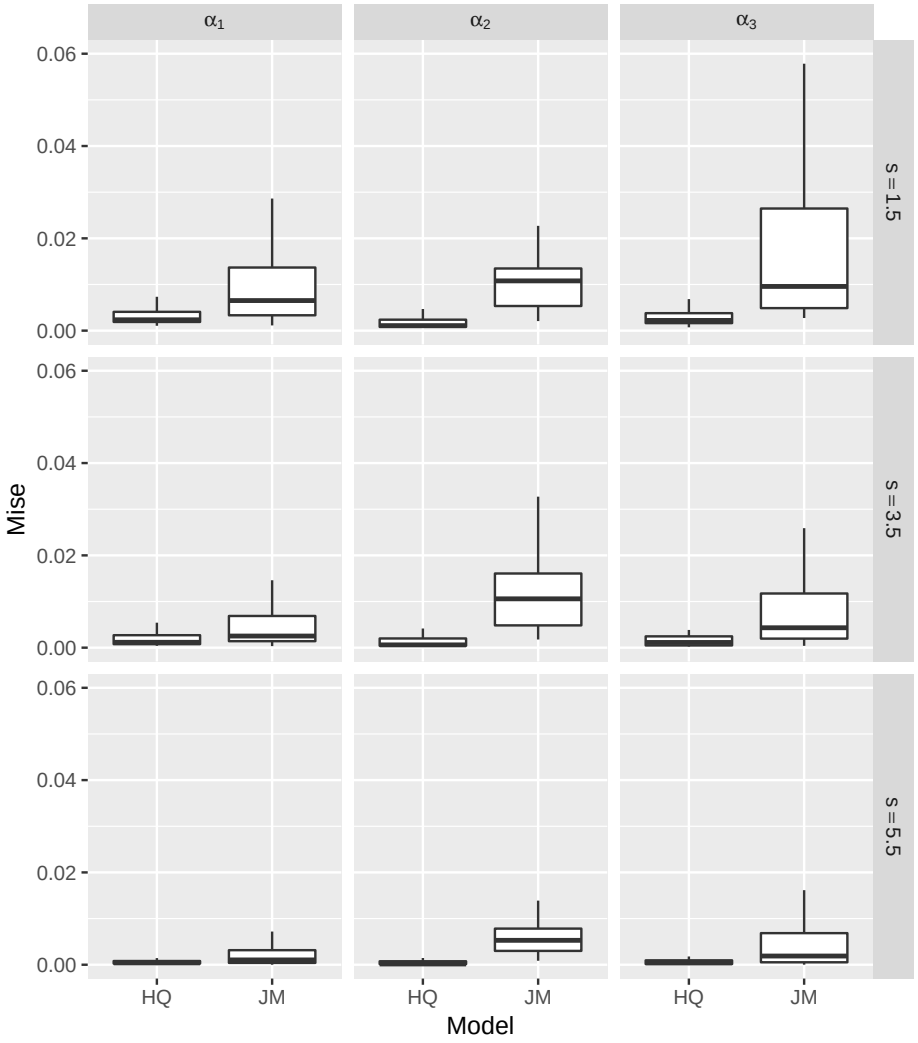


Fig. 11. Mean integrated squared error of the conditional survival estimate in a sample of 100 new simulated individuals. Conditional survival estimates $\hat{S}_\star^{\text{JM}}$ derived from the joint model and $\hat{S}_\star^{\text{HQM}}$ derived from the high-quality marker model are compared based on 100 realizations for 3 different marker-only hazards α_j, $j = 1, 2, 3$, a sample of $n = 300$ individuals and 3 landmark times $s = 1.5, 3.5, 5.5$.

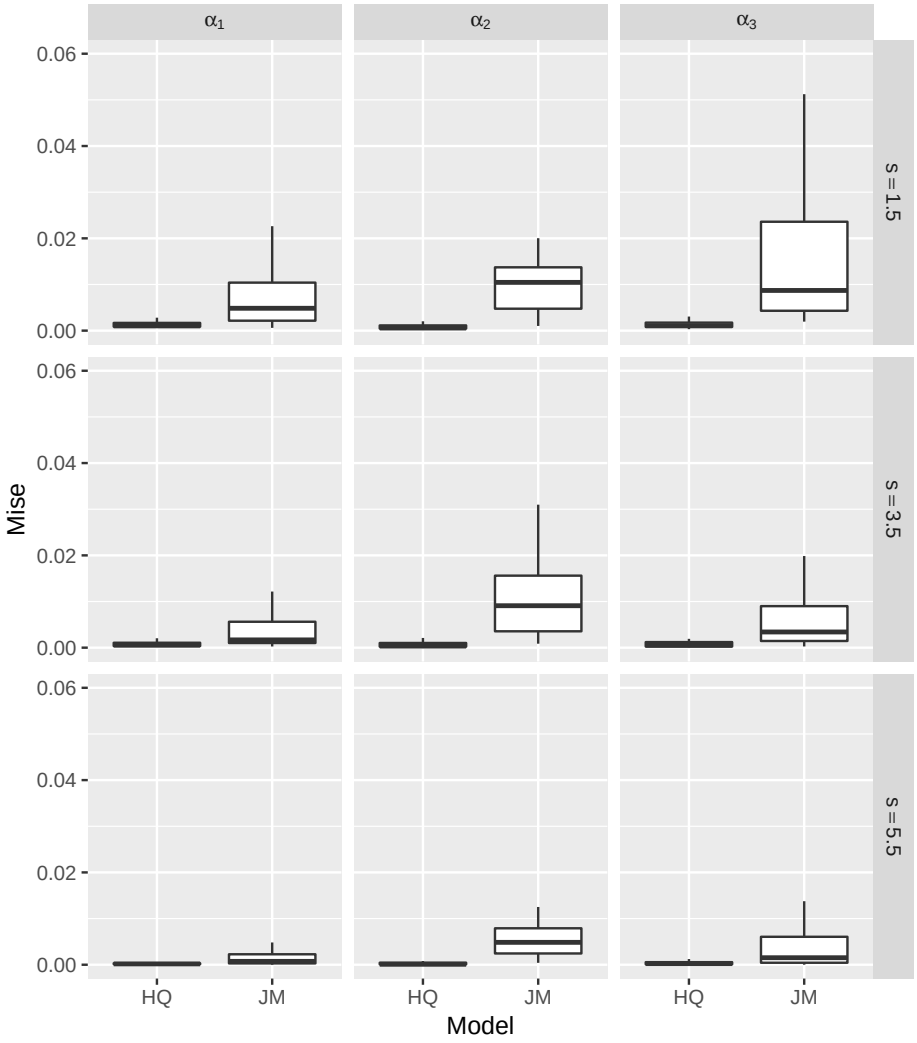


Fig. 12. Mean integrated squared error of the conditional survival estimate in a sample of 100 new simulated individuals. Conditional survival estimates $\hat{S}_\star^{\text{JM}}$ derived from the joint model and $\hat{S}_\star^{\text{HQM}}$ derived from the high-quality marker model are compared based on 100 realizations for 3 different marker-only hazards α_j, $j = 1, 2, 3$, a sample of $n = 300$ individuals and 3 landmark times $s = 1.5, 3.5, 5.5$.

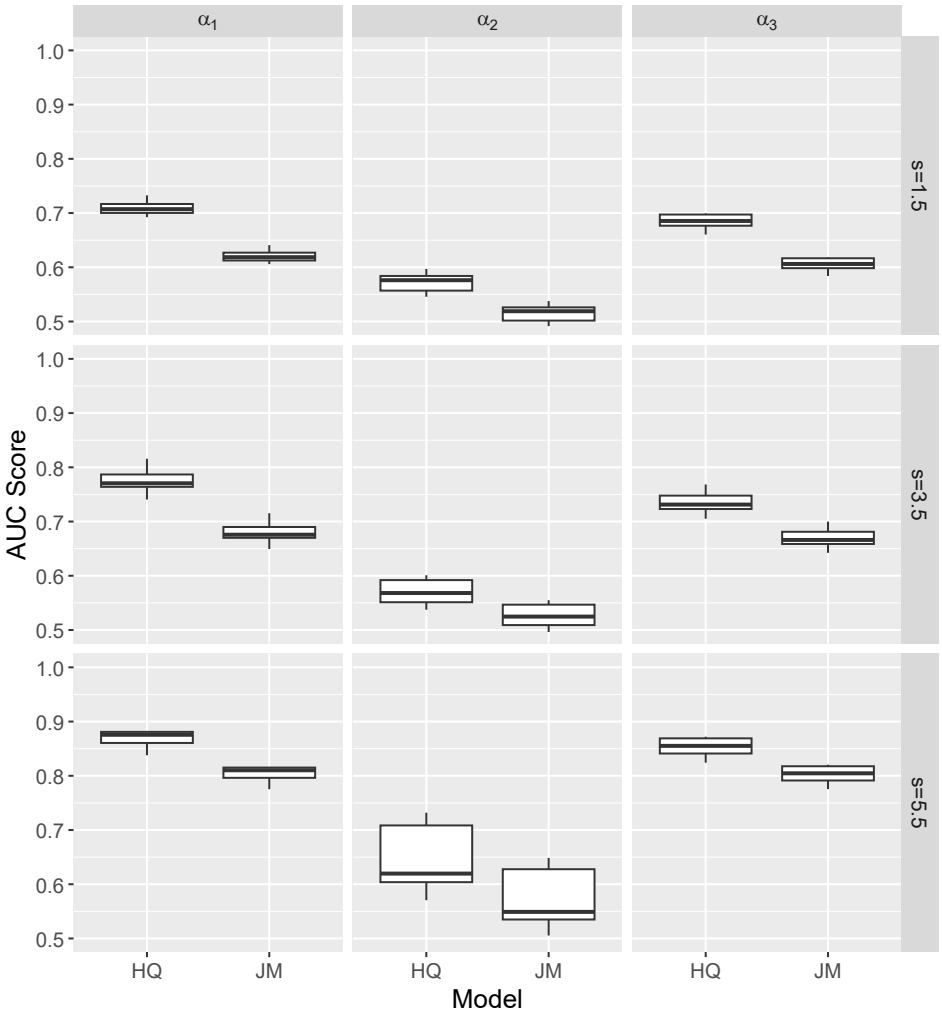


Fig. 13. AUC of the conditional survival estimate in a sample of 100 new simulated individuals. Conditional survival estimates $\hat{S}_\star^{\text{JM}}$ derived from the joint model and $\hat{S}_\star^{\text{HQM}}$ derived from the high-quality marker model are compared based on 100 realizations for 3 different marker-only hazards α_j, $j = 1, 2, 3$, a sample of $n = 300$ individuals and 3 landmark times $s = 1.5, 3.5, 5.5$.

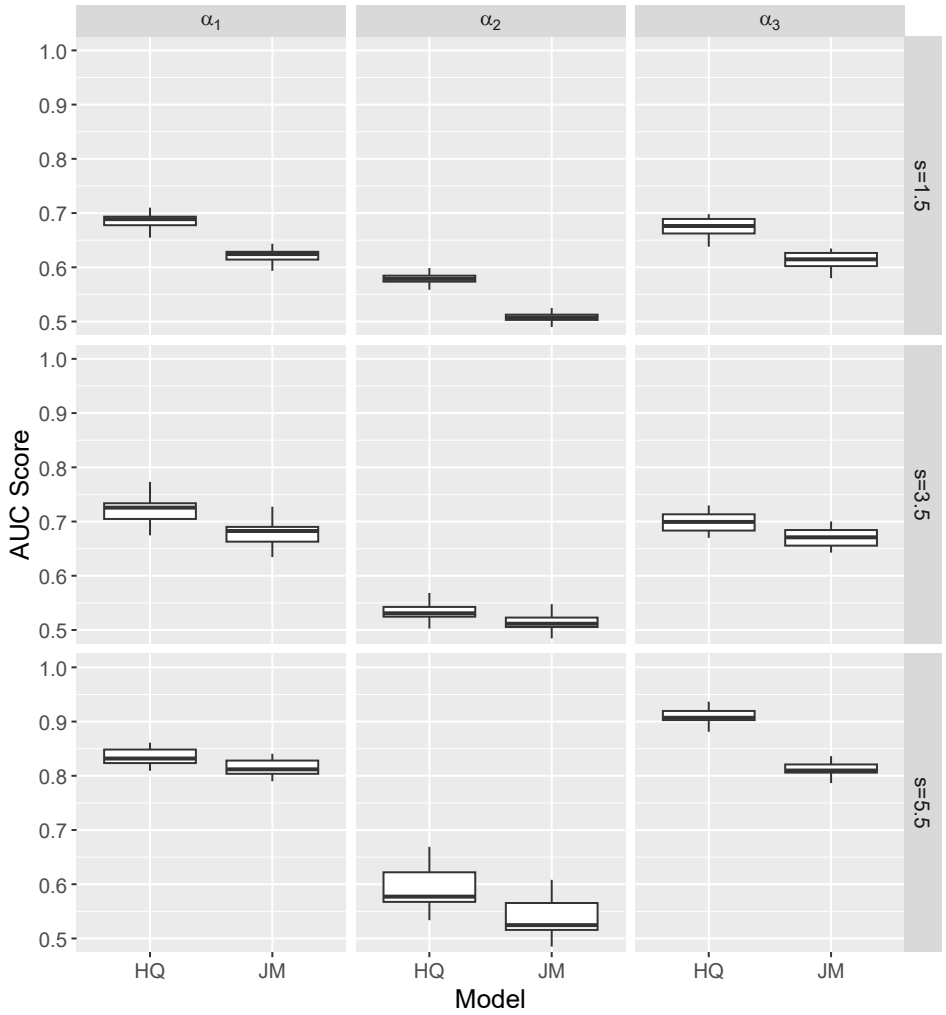


Fig. 14. AUC of the conditional survival estimate in a sample of 100 new simulated individuals. Conditional survival estimates $\hat{S}_\star^{\text{JM}}$ derived from the joint model and $\hat{S}_\star^{\text{HQM}}$ derived from the high-quality marker model are compared based on 100 realizations for 3 different marker-only hazards α_j, $j = 1, 2, 3$, a sample of $n = 600$ individuals and 3 landmark times $s = 1.5, 3.5, 5.5$.

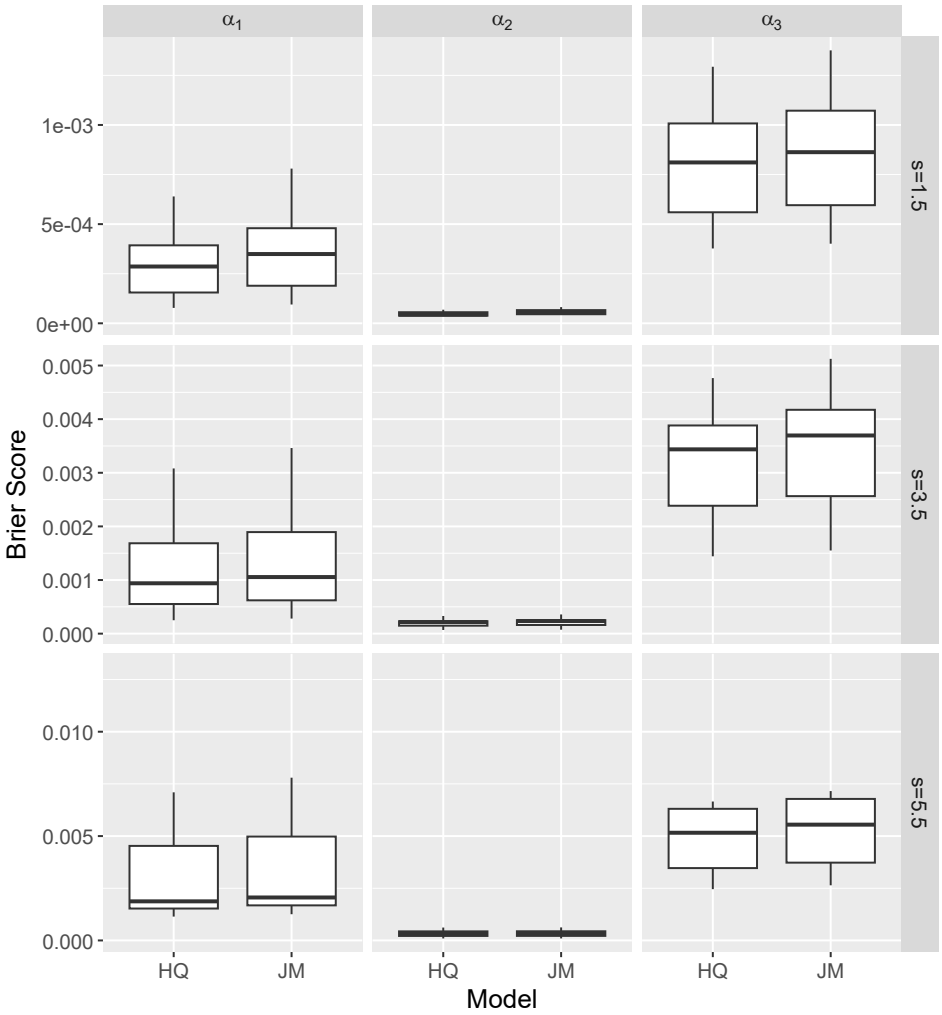


Fig. 15. Brier score of the conditional survival estimate in a sample of 100 new simulated individuals. Conditional survival estimates $\hat{S}_\star^{\text{JM}}$ derived from the joint model and $\hat{S}_\star^{\text{HQM}}$ derived from the high-quality marker model are compared based on 100 realizations for 3 different marker-only hazards α_j, $j = 1, 2, 3$, a sample of $n = 300$ individuals and 3 landmark times $s = 1.5, 3.5, 5.5$.

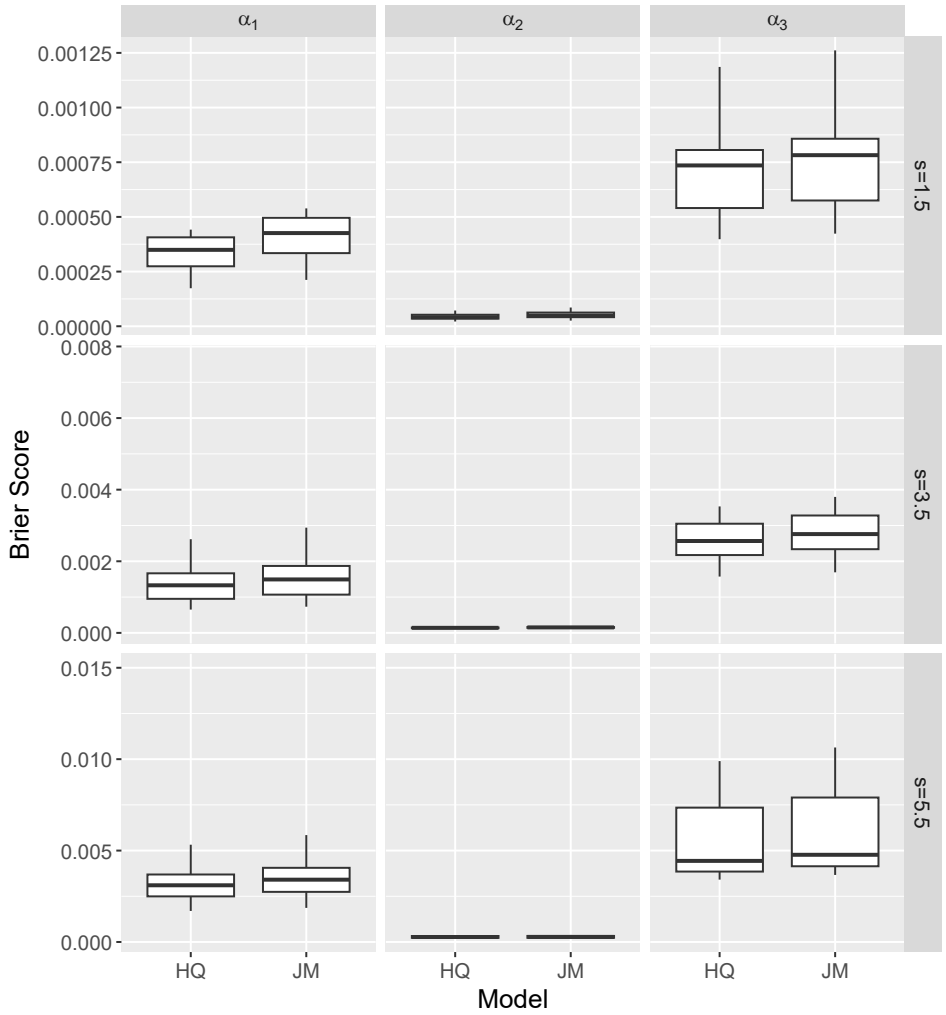


Fig. 16. Brier score of the conditional survival estimate in a sample of 100 new simulated individuals. Conditional survival estimates $\hat{S}_\star^{\text{JM}}$ derived from the joint model and $\hat{S}_\star^{\text{HQM}}$ derived from the high-quality marker model are compared based on 100 realizations for 3 different marker-only hazards α_j, $j = 1, 2, 3$, a sample of $n = 300$ individuals and 3 landmark times $s = 1.5, 3.5, 5.5$.

The AUC and Brier score simulations are implemented as replications of the MISE simulations, using exactly the same data (which were generated in all examples as described in Section 5.1) and simulation parameters (discussed in detail in Section 5.2). The obvious and only difference is that the AUC and Brier score measures are used instead of the MISE. Following Blanche et al. (2015), the AUC is defined by

$$\mathrm{AUC}(s,t) = \mathrm{pr}\left\{\pi_i(s,t) > \pi_j(s,t) \mid D_i(s,t) = 1, D_j(s,t) = 0, T_i > s, T_j > s\right\}$$

where $D_i(s,t) = I(s < T_i \leq s+t, \delta_i = 1)$; that is, for any subject i at risk at time s, $D_i(s,t) = 1$ when the individual i experiences the event of interest in the time interval $(s, s+t]$ and $D_i(s,t) = 0$ otherwise. Further, for $0 \leq s < T_i$,

$$\pi_i(s,t) = \mathrm{pr}\left\{s < T_i \leq s+t, \delta_i = 1 \mid T_i > s, X_i(s)\right\}.$$

The Brier score is defined as

$$\mathrm{BS}(s,t) = E\left[\left\{D(s,t) - \pi(s,t)\right\}^2 \mid T > s\right],$$

i.e. it is a mean squared error.

Throughout the simulation study, the $\mathrm{AUC}(s,t)$ is approximated by

$$\hat{\mathrm{AUC}}(s,t) = \frac{\sum_{i=1}^n \sum_{j=1}^n I\{\pi_i(s,t) > \pi_j(s,t)\}\hat{D}_i(s,t)\{1-\hat{D}_i(s,t)\}\hat{W}_i(s,t)\hat{W}_j(s,t)}{\sum_{i=1}^n \sum_{j=1}^n \hat{D}_i(s,t)\{1-\hat{D}_i(s,t)\}\hat{W}_i(s,t)\hat{W}_j(s,t)}. \tag{A3}$$

The weights $\hat{W}_\cdot$ are defined by

$$\hat{W}_i(s,t) = \frac{I(T_i > s+t)}{\hat{G}(s+t|s)} + \frac{I(s < T_i < s+t)\delta_i}{\hat{G}(T_i|s)},$$

where $\hat{G}(u)$ denotes the Kaplan-Meier estimator of the survival function of the censoring distribution at u and for each $u > s$, $\hat{G}(u|s) = \hat{G}(u)/\hat{G}(s)$ estimates the conditional probability of not being censored at time u conditionally on being uncensored at time s.

Note that implementation of (A3) for the JM method is readily provided in the `JM` package via the `aucJM` routine.

The Brier score is approximated by

$$\hat{\mathrm{BS}}(s,t) = \frac{1}{nS_T}\sum_{i=1}^n \hat{W}_i(s,t)\left\{\hat{D}_i(s,t) - \pi_i(s,t)\right\}^2.$$

For the JM method, this calculation is readily implemented via the routine `pec` in package `pec`.

A.4. Comparison of the HQM and Joint Modeling approaches for the pbc2 dataset

In addition to the illustration of the proposed HQM methodology in Figs. 2 and 3 of Section 6, here we expand the analysis of the PBC dataset by comparing the survival probabilities of patients diagnosed with primary biliary cirrhosis obtained by the HQM, the JM and the Kaplan–Meier estimator. Two sets of examples are presented, one based on the Bilirubin marker and one based on the Albumin marker using exactly the same marker values as in Section 6: 1 mg/dl, 4 mg/dl, 7 mg/dl, 10 mg/dl for Bilirubin, and 2 g/dl, 3 g/dl, 4 g/dl, 5 g/dl for Albumin.

As the JM methodology does not accommodate conditioning based on the values of a marker at a certain level, with purpose to produce comparable results across all three methodologies, here we compare the future mean survival probabilities from landmark times. That is, for each given landmark time value, we select from the dataset all the information before the given landmark time and compute the survival at different times after the landmark predicted from the JM approach. The resulting subset of the dataset is used to predict the survival function. For example, for Bilirubin=1 mg/dl, for landmark time $s = 1$, we use the full PBC dataset to simultaneously model the marker's evolution over time via a linear mixed effects model, and the instantaneous risk of death as a function of the underlying marker level. The performance

of the fitted JM model is then computed on the subset for which $\mathcal{X}_*(s) < 1$ and $\mathcal{X}_*(s) > 1$. In the present setting the JM model formulation given in (7) is written as

$$\begin{cases} X_i(t) & = m_i(t) + \varepsilon_i(t) \equiv Y_i(t)(\beta + B_i) + \varepsilon_i(t) \\ \lambda_i(t) & = \lambda_0(t)\exp\{m_i(t)\eta\}, \end{cases} \tag{A4}$$

with β, B_i and $\varepsilon_i(t)$ being exactly as in (7). In our case $X_i(t)$ is either the Bilirubin or Albumin markers and $Y_i(t)$ is the `years` variable, which is parameterized linearly, i.e. as $(1, t)$; this is in accordance to the typical JM applications. After the JM is fitted on all the data, posterior probability of the event for a new subject $\star$ from a landmark time s and up to an horizon t based on the history of the biomarker up to s, $\mathcal{X}_\star(s)$, is computed as:

$$\hat{S}^{\text{JM}}_\star(s+t) = P\left\{T_\star > s+t \mid \mathcal{X}_\star(s), T_\star > s; \hat{\theta}\right\}, \tag{A5}$$

where $\hat{\theta}$ is the vector of parameters of the joint model previously estimated in an independent sample using all the available information during the follow-up. Exactly the same methodology, with obvious adjustments, is employed for both markers and across all marker levels used for conditioning. A shorter range of landmark times, $s = (2, 3, 4, 5)$, is employed for albumin and a wider range, $s = (1, 4, 7, 10)$, for bilirubin. This is because in the context of PBC progression, reduced albumin levels (which indicate liver dysfunction) typically occur earlier than increased bilirubin levels, which are usually observed at advanced stages of the disease.

Implementation of (A4) and (A5) is readily available in the JM package in `R`, see also Rizopoulos (2012). In the examples implemented in this section, the joint model is implemented as a time-dependent relative risk model with a piecewise constant baseline risk function. It is fitted via the `jointModel` function using as input the two submodels fitted in (A4), setting `timeVar=year` (i.e. the time variable in the linear mixed model is the `year` variable) and specifying as fitting method the `piecewise-PH-aGH` option. Implementation of (A5) is performed via the function `survfitJM` of the `JM` package using as input the fitted joint model, specifying as the `newdata` argument the resulting subset (i.e. the longitudinal and covariate information for the subjects for which prediction of survival probabilities is required) and as the `survTimes` argument the equidistant, with step 0.5, gridpoints $t \in (s, 14]$. As this implementation produces separate survival curves for each individual in the subset, the survival curve reported in each example is obtained by averaging the survival probabilities of all individuals across each gridpoint t.

As in Section 6, estimation of the survival function with the HQM methodology is performed, using the same subset as in the JM estimator, via (8) using in each example K-fold CV bandwidth (determined as described in Section A.2). Finally, estimation of the survival function with the Kaplan–Meier estimator is performed via the `survfit` routine in `R`.

The results are contained in Fig. 17 where it is seen that all three approaches produce probabilities which are generally close to each other, with the HQM and Kaplan–Meier curves being slightly closer than the JM estimator. Additionally a comparison of the AUC and Brier score metrics for the HQM and JM estimators reported in Table 1, suggests that for the examples presented in Fig. 17, HQM dominates the JM method in the context of the PBC dataset. The only exception is landmark time $s = 10$ in the Bilirubin example, where while strictly speaking HQM performs worse than JM, nevertheless both metrics indicate that both estimators perform rather equivalently.

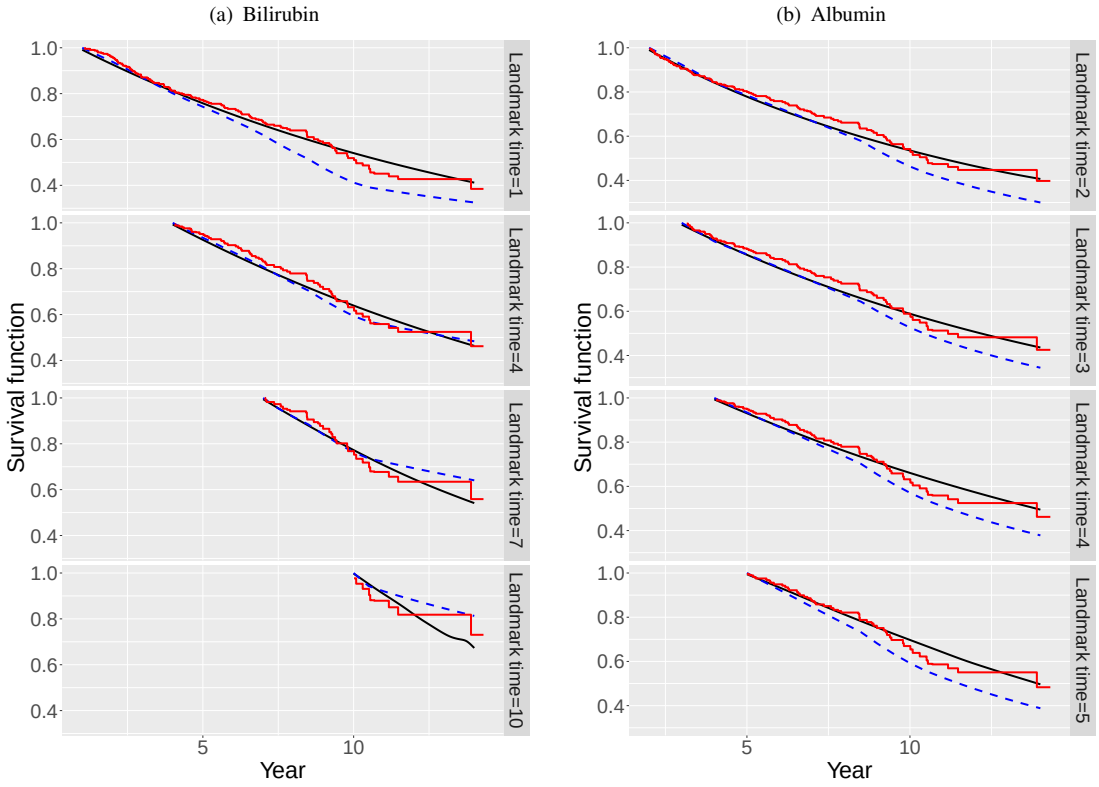


Fig. 17. Comparison of HQM (black solid line) and JM (blue dashed line) estimators based on landmarking using as reference the Kaplan–Meier (red stairstep line), for future conditional on (a) Bilirubin and (b) Albumin survival functions.

Table 1. *AUCs and Brier scores (BS) for the HQM and JM estimators of Fig.17.*

	Bilirubin					Albumin			
	AUC		BS			AUC		BS	
Landmark	HQM	JM	HQM	JM	Landmark	HQM	JM	HQM	JM
1	0.910	0.820	0.065	0.072	2	0.921	0.776	0.078	0.092
4	0.889	0.832	0.062	0.071	3	0.862	0.728	0.065	0.072
7	0.812	0.749	0.083	0.093	4	0.846	0.803	0.053	0.067
10	0.882	0.904	0.161	0.159	5	0.874	0.803	0.055	0.064

B. DISCUSSION ON ASSUMPTIONS, AUXILIARY LEMMAS AND PROOFS OF THEOREMS 1 AND 2

B.1. Basic definitions and local linear estimation

We will develop our theory for the case that the marker takes values in a fixed interval $A \subset \mathbb{R}$. This could be always achieved by transforming the marker values monotonically into this interval. Typically, the density of the marker will decrease to zero at the boundaries of A. We will deviate from this assumption by assuming that the density is bounded away from 0 on the whole interval. Our theory could be extended to a more general setting but at the cost of additional technical difficulties in the proofs. Notice that in the boundary points of A, $\hat{h}_x$ does not achieve square-root consistent estimation of the underlying hazard

rate h_x as its bias there is of order b_2. To take care of boundary effects we will assume that the kernel is appropriately modified at the boundary to avoid large bias terms at the boundary. Clearly, boundary corrections may be superfluous in case of densities that vanish at the boundaries fast enough. To define the boundary corrected kernel, set $b = b_1, b_2$ and for a density function K^*, let $K_b^*(u) = b^{-1}K^*(b^{-1}u)$. For $u, v \in A$ set

$$K_b(u,v) = \frac{c_{2,u} - c_{1,u}(u-v)}{c_{0,u}c_{2,u} - c_{1,u}^2} K_b^*(u-v), \quad c_{k,u} = \int_A (u-w)^k K_b^*(u-w)\mathrm{d}w.$$

Note that for the modified kernel we have that

$$\int_A K_b(u,w)\mathrm{d}w = 1, \quad \int_A (u-w)K_b(u,w)\mathrm{d}w = 0.$$

Other boundary corrected kernels with these properties will also work in our asymptotic approach.

Replacing the convolution kernel $K_{b_2}(x-u)$ in the definition of $\hat{h}_x(t)$ in (3) by the boundary corrected kernel $K_{b_2}(x,u)$ yields a local linear estimator, denoted by $\tilde{h}_x(t)$ and defined by

$$\tilde{h}_x(t) = \frac{\sum_{i=1}^n \int_0^T \tilde{\alpha}_{i,b_1}\{X_i(t+s)\} Z_i(t+s) Z_i(s) K_{b_2}\{x, X_i(s)\}\mathrm{d}s}{\sum_{i=1}^n \int_0^T Z_i(t+s) Z_i(s) K_{b_2}\{x, X_i(s)\}\mathrm{d}s},$$

where

$$\tilde{\alpha}_{i,b_1}(z) = \frac{\sum_{k\neq i} \int_0^T K_{b_1}\{z, X_k(s)\}\mathrm{d}N_k(s)}{\sum_{k\neq i} \int_0^T K_{b_1}\{z, X_k(s)\} Z_k(s)\mathrm{d}s}.$$

Similarly to $\tilde{h}_x(t)$, $\tilde{\alpha}_{i,b_1}(z)$ results by replacing the convolution kernel in (2) with the boundary corrected kernel $K_{b_2}(x,u)$.

In contrast to $\hat{h}_x$, $\tilde{h}_x$ is a square-root consistent estimator of h_x as its bias is of order $b_1 b_2 + b_2^2$ for all points of A (including boundary points) which under our assumptions is $o(n^{-1/2})$. It should be noted here that the use of the boundary corrected estimator of $\tilde{\alpha}_{i,b_1}$ in $\tilde{h}_x(t)$ is not necessary for achieving square-root consistent estimation of h_x. Another approach for obtaining a local linear estimator is to replace the smoothing step in (3) by local linear smoothing. That is, estimate h_x by μ where $(\mu, \mu^\star)$ minimizes

$$\sum_{i=1}^n \int_0^T \left[\hat{\alpha}_{i,b_1}\{X_i(t+s)\} - \mu - \mu^\star\{X_i(s) - x\}\right]^2 Z_i(t+s) Z_i(s) K_{b_2}\{x - X_i(s)\} ds.$$

This would again result in a square-root consistent estimator of the underlying hazard for all $x \in A$, irrespectively of whether the marker-only hazard is estimated as in (2), by a boundary corrected kernel or by local linear smoothing as in Nielsen (1998).

B.2. Discussion on the validity of Assumptions [A1] and [A2]

This Section demonstrates the validity of Assumptions [A1] and [A2] in the context of the pbc2 dataset and discusses the sensitivity of the method when the assumptions are violated. First note that Assumption [A2] consists of two parts:

(i) The first part is the Markov property, that is,

$$\mathcal{L}\left\{X_i(s+t) | \mathcal{F}_s, T_i \geq s+t, Z_i(s+t) = Z_i(s) = 1\right\}$$

depends only on $X_i(s)$ and $T_i \geq s+t, Z_i(s+t) = Z_i(s) = 1$ and s.

(ii) The second part is time homogeneity, i.e.

$$\mathcal{L}\left\{X_i(s+t) | X_i(s), T_i \geq s+t, Z_i(s+t) = Z_i(s) = 1\right\}$$

depends only on $X_i(s)$ and $T_i \geq s+t, Z_i(s+t) = Z_i(s) = 1$ (and not on s).

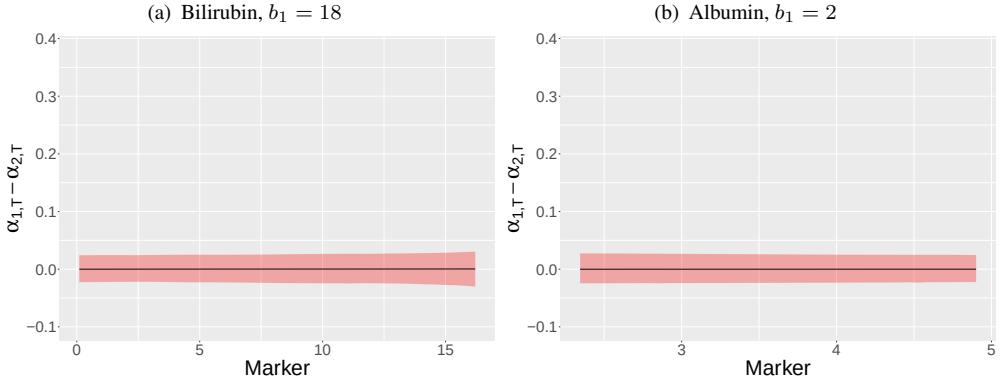


Fig. 18. $\alpha_T(z)$ (solid black line) and 95% confidence intervals (shaded red area) for Bilirubin and Albumin.

Condition (i) is rather redundant as we estimate

$$h_{x,s}(t) = P\left\{T_i \in (s+t, s+t+\ dt) | X_i(s) = x, Z_i(s+t) = Z_i(s) = 1\right\} / \, dt$$

and not

$$P\left\{T_i \in (s+t, s+t+\ dt) | X_i(s) = x, \mathcal{F}_s, T_i \geq s+t, Z_i(s+t) = Z_i(s) = 1\right\} / \, dt.$$

That is, we estimate the conditional hazard given $X_i(s)$ and forgetting the past values $X_i(u), u < s$. Consequently the whole theory developed herein would hold without condition (i) but it is more natural to estimate what is the conditional hazard for individual i given all information up today $(= s)$. As such we proceed with testing [A1] and [A2](ii).

Testing [A1] in the context of the pbc2 dataset amounts to calculating and constructing point-wise confidence intervals for the quantity $\alpha_T(z) = \alpha_{1,T}(z) - \alpha_{2,T}(z)$ for both the bilirubin and albumin markers, where

$$\alpha_{1,T}(z) = \frac{\sum_{i=1}^n \int_0^{T/2} K_{b_1}\left\{z - X_k(s)\right\} \, dN_k(s)}{\sum_{i=1}^n \int_0^{T/2} K_{b_1}\left\{z - X_k(s)\right\} Z_k(s) \, d(s)},$$
$$\alpha_{2,T}(z) = \frac{\sum_{i=1}^n \int_{T/2}^{T} K_{b_1}\left\{z - X_k(s)\right\} \, dN_k(s)}{\sum_{i=1}^n \int_{T/2}^{T} K_{b_1}\left\{z - X_k(s)\right\} Z_k(s) \, d(s)}.$$

As $\alpha_{1,T}(z)$ and $\alpha_{2,T}(z)$ are asymptotically independent it is expected that when [A1] holds $\alpha_T(z)$ will be equal to zero. For significance level $\alpha = 5\%$, each confidence interval is calculated as $(\bar{\alpha}(z) - \alpha^b_{1-\alpha/2}(z), \bar{\alpha}(z) - \alpha^b_{\alpha/2}(z))$ where $\alpha^b_{1-\alpha/2}(z), \alpha^b_{\alpha/2}(z)$ are the empirical quantiles of the distribution of $\alpha_T(z)$, obtained after 100 bootstrap replications and $\bar{\alpha}(z)$ is the expectation of $\alpha_T(z)$ calculated as the sample average of $\alpha_T(z)$ across the same 100 bootstrap replications. In Fig.18 we see that 0 is contained in the bootstrap confidence intervals for $\alpha_T(z)$ for both markers and for all values of z. Thus, [A1] is not rejected for all pointwise tests.

For testing [A2](ii), consider the test statistic

$$S_g(t,x) = n^{-1} \sum_{i=1}^n W_{1,i}(t,x)/\Gamma_1(t,x) - n^{-1} \sum_{i=1}^n W_{2,i}(t,x)/\Gamma_2(t,x)$$

where

$$W_{1,i}(t,x) = \int_0^{(T-t)/2} g\left\{X_i(t+s)\right\} Z_i(t+s) Z_i(s) K_b\left\{x, X_i(s)\right\} \, ds$$

$$\Gamma_1(t,x) = n^{-1}\sum_{i=1}^{n}\int_0^{(T-t)/2} Z_i(t+s)Z_i(s)K_b\left\{x, X_i(s)\right\}\, ds$$

$$W_{2,i}(t,x) = \int_{(T-t)/2}^{T-t} g\left\{X_i(t+s)\right\} Z_i(t+s)Z_i(s)K_b\left\{x, X_i(s)\right\}\, ds$$

$$\Gamma_2(t,x) = n^{-1}\sum_{i=1}^{n}\int_{(T-t)/2}^{T-t} Z_i(t+s)Z_i(s)K_b\left\{x, X_i(s)\right\}\, ds.$$

Here $g(z) = 1$ for $z \leq z^*$ and 0 otherwise. Under assumption [A2](ii), it is expected that $S_g(t,x)$ will be close to zero across both time (t) and the marker values (x). For each combination of t and x (i.e. for each implementation) the critical values of $S_g(t,x)$ are calculated by wild bootstrap as follows

(i) Generate $V_1, \ldots, V_n \sim N(0,1)$.
(ii) Calculate

$$S_g^*(t,x) = n^{-1}\sum_{i=1}^{n} V_i W_{1,i}(t,x)/\Gamma_1(t,x) - n^{-1}\sum_{i=1}^{n} V_i W_{2,i}(t,x)/\Gamma_2(t,x).$$

(iii) Repeat (i) and (ii) for 100 times to obtain $S^*_{1,g}(t,x), \ldots, S^*_{100,g}(t,x)$.
(iv) The critical value for fixed t is then obtained as the empirical 95th quantile of $S^*_{1,g}(t,x), \ldots, S^*_{100,g}(t,x)$.

The above bootstrap procedure is employed in testing [A2](ii) in the context of the pbc2 dataset. For both the bilirubin and albumin markers we consider the same conditioning values (x) as those depicted in Figs. 2 and 3. Then, for each combination of x and $t = 0, 1, 2, \ldots, 10$ we calculate $S_g(t,x)$ (depicted as +) and apply steps (i)–(iv) above to calculate the corresponding critical values (depicted as red triangles) in Figs. 19(a) and (b). In both plots the test statistic values are always smaller than the corresponding critical values across all years and marker value combinations. Thus [A2](ii) is not rejected for all pointwise tests.

Next we present a simulation study which investigates the sensitivity of the proposed methodology when either of the assumptions [A1] or [A2] is violated. We first consider the scenario where assumption [A1] does not hold. Let $g_1(s) = s/2, g_2(s) = s, g_3(s) = s^2, g_4(s) = s^3$ and define the corresponding marker only hazard functions

$$\alpha_1\{x, g_r(s)\} = \exp\{2x + g_r(s) - 2\}/15,\ \alpha_2\{x, g_r(s)\} = 4\{x + g_r(s) - 0.3\}^4,$$
$$\alpha_3\{x, g_r(s)\} = 4[-4\{x + g_r(s) - 1\}]^{-1}, r = 1, \ldots, 4.$$

Then, an indication on the accuracy of the HQM estimator when the marker only hazard depends on s is obtained by replicating the MISE simulation of Section 5.1 with $\alpha_j(x)$ there replaced by $\alpha_j\{x, g_r(s)\}, j = 1, 2, 3, r = 1, 2, 3, 4$ respectively. All other settings are identical to the settings described in Section 5.1 with the only difference that for each combination of n, s and $\alpha_j(x;\cdot)$, instead of boxplots, the simulation results reported on Table 2 contain the average MISEs of $\hat{h}_{x,b,b}$, calculated across 100 realizations of the process. For comparison purposes the average MISEs of the HQM estimator under assumption [A1], for each combination of n, s and $\alpha_j(x), j = 1, 2, 3$ are also provided on Table 2. The simulation results suggest that mild departures from Assumption [A1] do not significantly affect the MISE of the HQM estimator. This can be seen by the fact that, in comparison to the MISE of $\hat{h}_{x,b,b}$ under [A1], for $g_1(s) = s/2$ there is an average MISE increase of approximately 10-14% across both the sample size and all landmark levels. For $g_2(s) = s$ the corresponding increase is between 25-32%; we feel that the corresponding MISE figures are still acceptable as estimates of the underlying survival function. However for $g_3(s) = s^2$ the MISE of $\hat{h}_{x,b,b}$ is more than double (approximately 2.1-2.3 times more) than the corresponding MISE figures under [A1]. Finally, the effect is even more visible for $g_4(s) = s^3$ where the MISE figures are tripled compared to the MISE of $\hat{h}_{x,b,b}$ under assumption [A1].

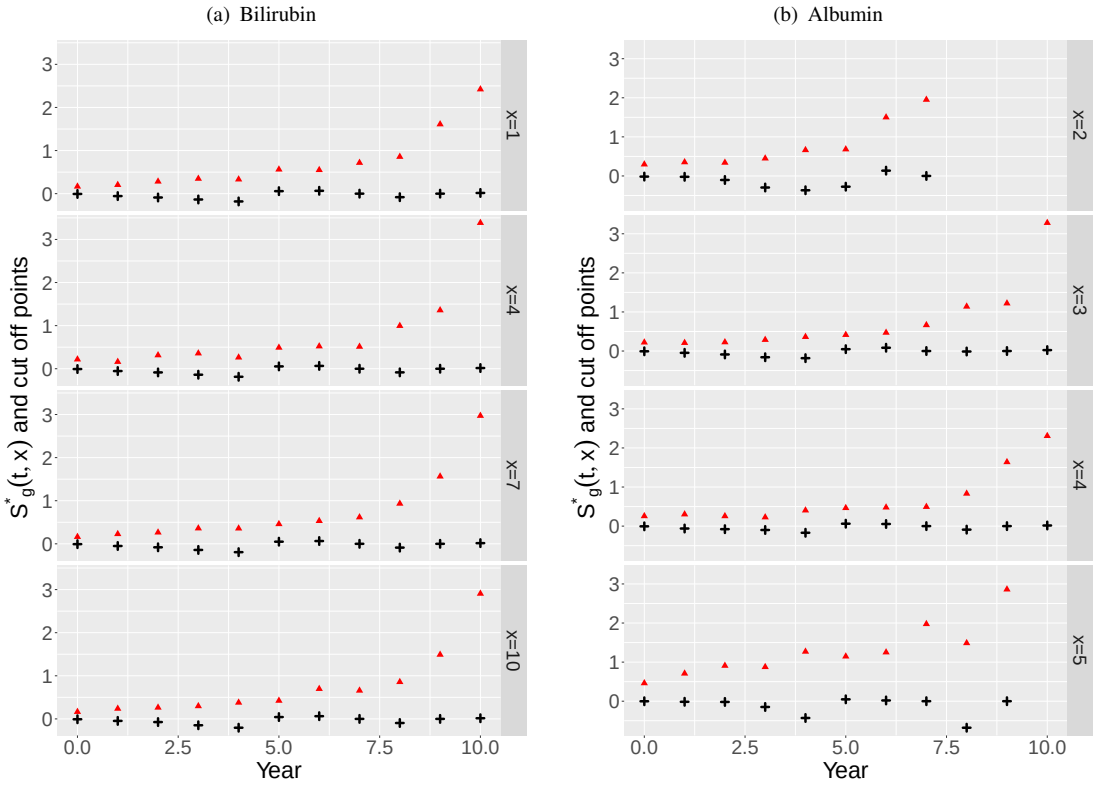


Fig. 19. Test statistic $S_g(t,x)$ values (+) and corresponding 95% cut-off points (red triangles) by year (horizontal axis), conditional on selected marker values (x panels).

Finally we examine the performance of the HQM estimator when assumption [A2] does not hold. This is achieved by replicating the MISE simulation of Section 5.1 with the difference that the marker values are now generated as follows: for each individual i the initial marker value, $X_i(s)$ is drawn from $N(0, 0.07^2)$. Then for each subsequent time point $t > s$ the value of the marker will be $X_i(t) = X_i(t-1) + \varepsilon_i(t)$ with $\varepsilon_i(t) \sim N\{\mu(t,s), \sigma^2(t,s)\}$. As a result the distribution of $X_i(t+s)|X_i(s)$ is Normal and depends on t, s and $X_i(s)$, i.e. it is a time-inhomogeneous Gaussian process. The process is implemented on the same grid values, t_g, as those used in the simulation of Section 5.1 and using the same (random, uniformly distributed) starting points, for all $T_i \geq t + s$. The magnitude of departure from the validity of assumption [A2] is controlled via the $\mu(t,s)$ and $\sigma^2(t,s)$ functions, while all other simulation settings are identical to those of Section 5.1. In total we repeat the simulation across all sample sizes, landmark times and marker-only hazards four times. Each repetition returns the average MISE of $\hat{h}_{x,b,b}$, across 100 replications of the process, using different a functional form for $\mu(t,s)$. Specifically, in place of $\mu(t,s)$, we consider the following four different functional forms, denoted by $\mu_j(t,s), j = 1, \ldots, 4$ and given by

$$\mu_1(t,s) = (s+t)/500, \mu_2(t,s) = (s+t)/100, \mu_3(t,s) = (s+t)/50, \mu_4(t,s) = (s+t)/20,$$

while in all cases the variance function is $\sigma^2(t,s) = (s+t)/20$. The results are reported on Table 3. For comparison purposes the average MISEs of the HQM estimator under assumption [A2], for each combination of n, s and $\alpha_j(x), j = 1, 2, 3$ are also provided, under the column titled (μ_0, σ_0). These figures correspond to the average MISE values reported on Tables 11 and 12, i.e. they represent the estimator's MISEs for a homogeneous process based on $N(0, 0.07^2)$. The average MISE figures for each of the four scenarios examined here are provided in columns $(\mu_1, \sigma), \ldots, (\mu_4, \sigma)$. As in the simulation for assumption [A1], here the results also suggest that mild departures from Assumption [A2], as these correspond

Table 2. *MISEs of HQM estimator when Assumption [A1] is violated*

sample size	s	$\alpha_1(x)$	$\alpha_1(x,2s)$	$\alpha_1(x,s^2)$	$\alpha_1(x,2s^2)$	$\alpha_1(x,s^3)$
	1.5	0.00243	0.00246	0.00292	0.00704	0.01982
	3.5	0.00181	0.00206	0.00258	0.00553	0.01739
	5.5	0.00133	0.00154	0.00198	0.00510	0.01522
	s	$\alpha_2(x)$	$\alpha_2(x,2s)$	$\alpha_2(x,s^2)$	$\alpha_2(x,2s^2)$	$\alpha_2(x,s^3)$
	1.5	0.00173	0.00189	0.00242	0.00531	0.01926
$n=300$	3.5	0.00141	0.00147	0.00222	0.00479	0.01575
	5.5	0.00142	0.00164	0.00184	0.00442	0.01473
	s	$\alpha_3(x)$	$\alpha_3(x,2s)$	$\alpha_3(x,s^2)$	$\alpha_3(x,2s^2)$	$\alpha_3(x,s^3)$
	1.5	0.00247	0.00304	0.00452	0.01052	0.03614
	3.5	0.00178	0.0022	0.00333	0.00803	0.02764
	5.5	0.00136	0.00138	0.00218	0.00499	0.01569
	1.5	0.00153	0.00182	0.0027	0.00684	0.01982
	3.5	0.00138	0.00161	0.00238	0.00547	0.01739
	5.5	0.00112	0.00138	0.00192	0.00451	0.01522
	s	$\alpha_2(x)$	$\alpha_2(x,2s)$	$\alpha_2(x,s^2)$	$\alpha_2(x,2s^2)$	$\alpha_2(x,s^3)$
	1.5	0.00128	0.00146	0.00216	0.00549	0.01731
$n=600$	3.5	0.00131	0.00148	0.0023	0.00535	0.01792
	5.5	0.00117	0.00130	0.00206	0.00511	0.01677
	s	$\alpha_3(x)$	$\alpha_3(x,2s)$	$\alpha_3(x,s^2)$	$\alpha_3(x,2s^2)$	$\alpha_3(x,s^3)$
	1.5	0.00143	0.00178	0.00197	0.00428	0.01235
	3.5	0.00136	0.00169	0.00223	0.00576	0.01559
	5.5	0.00108	0.00128	0.00171	0.00434	0.01481

to the first two inhomogeneous Gaussian processes based on $\mu_1(t,s)$ and $\mu_2(t,s)$, do not significantly affect the MISE of the HQM estimator. This can be seen by a simple comparison with the corresponding homogeneous process MISEs. For $\mu_1(t,s)$ the MISE increase is on average approximately 13% across all n, s and α_j combinations; the corresponding figure for $\mu_2(t,s)$ is approximately 35%. We feel that the actual MISE figures are still acceptable as estimates of the underlying survival function. However, the difference becomes substantial for $\mu_3(t,s)$ where the MISE of $\hat{h}_{x,b,b}$ is on average 5.4 times larger than the estimate's MISE under assumption [A2]; this corresponds to an average 81% increase in MISE across all n, s and α_j's. The effect is even more visible for $\mu_4(t,s)$ where the MISE figures are on average 26 times larger than the homogeneous process MISEs, or in other words an to an approximate 96% MISE increase.

Table 3. *MISEs of HQM estimator when Assumption [A2] is violated*

sample size	s	(μ_0, σ_0)	(μ_1, σ)	(μ_2, σ)	(μ_3, σ)	(μ_3, σ)
$n = 300$	1.5	0.00257	0.00261	0.00315	0.01126	0.04554
	3.5	0.00194	0.00213	0.00275	0.00829	0.03999
	5.5	0.00135	0.00156	0.00203	0.00781	0.03568
	1.5	0.00189	0.00206	0.00244	0.00819	0.04248
	3.5	0.00141	0.00148	0.00237	0.00775	0.03476
	5.5	0.00146	0.00173	0.00189	0.00663	0.03465
	1.5	0.00251	0.00312	0.00462	0.01647	0.08676
	3.5	0.00192	0.00231	0.00345	0.01291	0.06611
	5.5	0.00142	0.00143	0.00226	0.00756	0.03748
$n = 600$	1.5	0.00162	0.00192	0.00296	0.01090	0.04369
	3.5	0.00151	0.00176	0.00243	0.00846	0.03949
	5.5	0.00116	0.00145	0.00212	0.00688	0.03623
	1.5	0.00131	0.00162	0.00222	0.00844	0.03993
	3.5	0.00136	0.00157	0.00244	0.00842	0.03977
	5.5	0.00122	0.00143	0.00213	0.00784	0.03944
	1.5	0.00155	0.00185	0.00197	0.00646	0.02985
	3.5	0.00146	0.00175	0.00235	0.00898	0.03440
	5.5	0.00113	0.00132	0.00172	0.00682	0.03517

B.3. Discussion on assumptions [A3]–[A7]

Assumption [A3] contains standard smoothness conditions of several functions. It also implies that $h_x(s)$ is twice continuously differentiable with respect to $x \in A$ for $0 \leq s \leq T$ and bounded away from 0. We conjecture that partially the assumptions of a second derivative in [A3] can be replaced by only requiring a continuous first order derivative but do not follow this line for getting simplifications in the proof. Assumption [A4] is a weak smoothness assumption on γ and the first part of [A5] is a standard assumption on the smoothing kernel. Furthermore, the assumption that $\rho_1, \rho_2 < 1/3$ is only made for technical reasons to allow for some bounds in the proofs. On the other side the assumption that $\rho_1, \rho_2 > 1/4$ is essential for getting that the bias terms of both smoothing steps are of order $o(n^{-1/2})$. Note that this is implied because under our smoothness assumptions we have that the bias terms are of order b_1^2 and b_2^2. One gets that with choices of bandwidths that fulfill $\rho_1, \rho_2 > 1/4$ we have that the limiting distribution of our hazard estimator does not depend on the bandwidths. The choice of bandwidths only affects second order terms. Finally, [A6] will be used in the proof of Lemma B1 below. It is fulfilled if $\kappa(t)^{-1}\{X_i(s+t) - X_i(s)\}$ has a bounded density for a function κ with

$$\int_0^T \kappa(t)\mathrm{d}t \leq C_\kappa.$$

This is for example the case if $X_i(s)$ is a Brownian motion. Then one gets the assumption with a choice of κ that is proportional to $t^{-1/2}$. Assumption [A7] will be used in the proof of tightness of the process $\hat{h}_{x_*}(t)$, see the proof of (B14). It can be checked that this assumption is in particular fulfilled if $X_i(s)$ is a Brownian motion.

For simplification of notation from now on we assume that $b_1 = b_2$ and denote the bandwidth by b with $b = c_b n^{-\rho}$ for some $c_b > 0$, $1/4 < \rho < 1/3$.

B.4. Outline of the proof of Theorem 1. Basic decomposition of the estimation error

We will show now the validity of the statement of Theorem 1 for an estimator of the marker conditional future hazard where the convolution kernels in steps (2) and (3) are replaced by boundary corrected kernels as described in Section B.1. For simplicity we assume that x^* lies in the interior of the support of $X_i(t)$. The latter assumption is not needed as argued in Section B.1 but it essentially simplifies the proofs by avoiding additional discussions of arguments near the boundary.

First we will state an expansion of $\hat{h}_{x_*}(t) - h_{x_*}(t)$. For this expansion we have to introduce some additional notation. We define

$$g_{t,x}(z) = \frac{\int_0^{T-t} \gamma(t,s) f(s, t+s, x, z) ds}{\int_0^T \gamma(s) f(s,z) ds},$$

$$\hat{g}_{i,t,x}(z) = n^{-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \int_0^{T-t} \hat{E}_j\{X_j(t+s)\}^{-1} K_b\{X_j(t+s), z\} Z_j(t+s) Z_j(s) K_b\{x, X_j(s)\} ds,$$

$$\hat{E}_i(x) = n^{-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \int_0^{T} K_b\{x, X_j(s)\} Z_j(s) ds.$$

We write for the denominator of $\hat{h}_x(t)$

$$\hat{\Gamma}(t,x) = n^{-1} \sum_{i=1}^{n} \int_0^{T-t} Z_i(t+s) Z_i(s) K_b\{x, X_i(s)\} ds,$$

respectively. We also define the compensator $\alpha_i^*(t)$ of $\hat{\alpha}_i(t)$ as

$$\alpha_i^*(z) = \hat{E}_i(z)^{-1} n^{-1} \sum_{k \neq i} \int K_b\{z, X_k(s)\} Z_k(s) \alpha\{X_k(s)\} \mathrm{d}s.$$

We arrive at the following expansion of $\hat{h}_{x_*}(t) - h_{x_*}(t)$ which can be verified by straight forward but lengthy calculations.

$$\hat{h}_{x_*}(t) - h_{x_*}(t) = \hat{\Gamma}(t, x_*)^{-1} \{A(t) + B(t) + R_1(t) + R_2(t) + R_3(t)\},$$

where

$$A(t) = n^{-1} \sum_{i=1}^{n} \int_0^T g_{t,x_*}\{X_i(s)\} dM_i(s),$$

$$B(t) = n^{-1} \sum_{i=1}^{n} \int_0^T [\alpha\{X_i(t+s)\} - h_{X_i(s)}(t)] Z_i(t+s) Z_i(s) K_b\{x_*, X_i(s)\} \mathrm{d}s,$$

$$R_1(t) = n^{-1} \sum_{i=1}^{n} \int_0^T [\hat{g}_{i,t,x_*}\{X_i(s)\} - g_{t,x_*}\{X_i(s)\}] \, dM_i(s),$$

$$R_2(t) = n^{-1} \sum_{i=1}^{n} \int_0^T [\alpha_i^*\{X_i(t+s)\} - \alpha_i\{X_i(t+s)\}] \, Z_i(t+s) Z_i(s) K_b\{x_*, X_i(s)\} \mathrm{d}s,$$

$$R_3(t) = n^{-1} \sum_{i=1}^{n} \int_0^T \{h_{X_i(s)}(t) - h_{x_*}(t)\} Z_i(t+s) Z_i(s) K_b\{x_*, X_i(s)\} \mathrm{d}s.$$

The function $g_{t,x}(z)$ and its estimator $\hat{g}_{i,t,x}(z)$ were introduced in (4) and (5).

This decomposition allows us to outline the critical terms: the sum of the components $A(\cdot)$ and $B(\cdot)$ will converge weakly to a Gaussian process while the terms $R_i(\cdot), i = 1, 2, 3$ are error terms of lower order. Here, the first error term $R_1(x, \cdot)$ needs special attention, since the function $\hat{g}_{i,t,x_*}(X_i(s))$ is not

predictable with respect to $\mathcal{F}_i^s$. In particular, for this reason, $R_1(x,\cdot)$ is not a sum of martingales. This problem has also been called the predictability issue. For the pointwise case a solution can be found in Mammen & Nielsen (2007). Their approach used in this paper cannot be applied in our settings which require a uniform approximation. We will make use of methods from empirical process theory to treat our case.

B.5. Some preliminary results for the proof of Theorem 1

First, in this and all subsequent sections C denotes always a strictly positive constant with meaning changing at each occurrence, even also in the same formula. We start with some results on the convergence of the quantities $\hat{E}_i(x)$, $\hat{\Gamma}(t,x)$, $\alpha_i^*(z)$ and $\hat{g}_{i,t,x_*}(z)$. We first compare these quantities with the following random variables.

$$\hat{E}(x) = n^{-1}\sum_{j=1}^{n}\int_0^T K_b\{x, X_j(s)\}Z_j(s)ds,$$

$$\alpha^*(z) = \hat{E}(z)^{-1}n^{-1}\sum_{k=1}^{n}\int K_b\{z, X_k(s)\}Z_k(s)\alpha\{X_k(s)\}\mathrm{d}s.$$

By using bounds for the integrands one can easily verify that uniformly for $1 \le i \le n$, $x, z \in A$

$$\hat{E}_i(x) = \hat{E}(x) + o_P(n^{-1/2}/\log n), \tag{B1}$$

$$\alpha_i^*(z) = \alpha^*(z) + o_P(n^{-1/2}/\log n). \tag{B2}$$

Note that $n^{1/2}b/\log n$ converges to infinity. We now define

$$\hat{g}_{t,x_*}(z) = n^{-1}\sum_{j=1}^{n}\int_0^{T-t}\hat{E}\{X_j(t+s)\}^{-1}K_b\{z, X_j(t+s)\}Z_j(t+s)Z_j(s)K_b\{x_*, X_j(s)\}ds.$$

We now argue that uniformly for $\delta_T \le t \le T-\delta_T$, $z \in A$

$$\hat{g}_{t,x_*}(z) = \hat{g}_{i,t,x_*}(z) + o_P(n^{-1/2}/\log n). \tag{B3}$$

This claim does not follow as easily as (B1) and (B2) because now the integrands in the definition of $\hat{g}_{i,t,x_*}(z)$ cannot be bounded by a term that is of order $o_P(n^{-1/2})$ using the crude methods leading to (B1) and (B2). The bound now follows by application of the following lemma.

LEMMA B1. *For all $z \in A$, $1 \le i \le n$ the variable*

$$\eta_i(z) = \int_0^T Z_i(s)|K_b\{z, X_i(s)\}|\mathrm{d}s$$

fulfills

$$E\{|\eta_i(z)|^k\} \le (k)!c_V^k, \tag{B4}$$

with $c_V = 2C_KC_\kappa$ and $C_K = \sup\{hK_h(x,z) : h > 0, x, z \in A\}$. Furthermore, we have that

$$E[\exp\{\lambda|\eta_i(z)|\}] \le C < \infty \tag{B5}$$

for $\lambda < c_V^{-1}$.

Proof. For the proof of the lemma note that

$$E\{|\eta_i(z)|^k\} = E\left(\int_{[0,T]^k}\prod_{j=1}^{k}[Z_i(s_j)|K_b\{z, X_i(s_j)\}|\mathrm{d}s_j]\right)$$

$$= k!E\left(\int_{0<s_1<\cdots<s_k<T}\prod_{j=1}^{k}\{Z_i(s_j)|K_b\{z, X_i(s_j)\}|\mathrm{d}s_j\}\right)$$

$$\leq k!E\bigg(\int_{0<s_1<\cdots<s_k<T}\prod_{j=1}^{k-1}\left(Z_i(s_j)\,|K_b\{z,X_i(s_j)\}|\right)C_Kb^{-1}\mathrm{pr}\Big\{|X_i(s_k)-z|\leq b,$$
$$Z_i(s_k)=1\mid X_i(s_{k-1}),Z_i(s_{k-1})=1\Big\}\prod_{j=1}^{k}\mathrm{d}s_j\bigg)$$
$$\leq k!E\bigg(\int_{0<s_1<\cdots<s_k<T}\prod_{j=1}^{k-1}\left[Z_i(s_j)|K_b\{z,X_i(s_j)\}|\right]C_Kb^{-1}\mathrm{pr}\Big\{|X_i(s_k)-X_i(s_{k-1})|\leq 2b,$$
$$Z_i(s_k)=1\mid X_i(s_{k-1}),Z_i(s_{k-1})=1\Big\}\prod_{j=1}^{k}\mathrm{d}s_j\bigg)$$
$$\leq k!E\bigg(\int_{0<s_1<\cdots<s_k<T}\prod_{j=1}^{k-1}\left[Z_i(s_j)K_b\{z,X_i(s_j)\}\right]C_Kb^{-1}2b\kappa(s_k-s_{k-1})\prod_{j=1}^{k}\mathrm{d}s_j\bigg)$$
$$\leq k!E\bigg(\int_{0<s_1<\cdots<s_{k-1}<T}\prod_{j=1}^{k-1}\left[Z_i(s_j)|K_b\{z,X_i(s_j)\}|\mathrm{d}s_j\right]\bigg)2C_KC_\kappa$$
$$=2C_KC_\kappa kE\{|\eta_i(z)|^{k-1}\},$$

where [A6] has been used in the last two inequalities. Equation (B4) now follows by an iterative application of this inequality. For a proof of (B5) note that

$$E[\exp\{\lambda|\eta_i(z)|\}]=1+\sum_{j=1}^{\infty}\frac{\lambda^kE\{|\eta_i(z)|^k\}}{k!}.$$

From (B5) we get that uniformly for $1\leq i\leq n$, $z\in A$

$$\eta_i(z)=O_P(\log n). \tag{B6}$$

This shows that uniformly for $1\leq i\leq n$, $z\in A$, $\delta_T\leq t\leq T-\delta_T$

$$n^{-1}\int_0^{T-t}\hat{E}\{X_i(t+s)\}^{-1}K_b\{z,X_i(t+s)\}Z_i(t+s)Z_i(s)K_b\{x_*,X_i(s)\}ds$$
$$=O_P\left(\frac{\log n}{nb}\right)=o_P(n^{-1/2}),$$

which implies (B3).

The following lemma results by standard properties of one and two dimensional kernel density estimators.

LEMMA B2. *It holds that*

$$\sup_{x\in A,\delta_T\leq t\leq T-\delta_T}\left|\hat{\Gamma}(t,x)-E\{\hat{\Gamma}(t,x)\}\right|=O_p\left\{\left(\frac{\log n}{nb}\right)^{\frac{1}{2}}\right\},$$
$$\sup_{x\in A}\left|\hat{E}(x)-E\{\hat{E}(x)\}\right|=O_p\left\{\left(\frac{\log n}{nb}\right)^{\frac{1}{2}}\right\}.$$

Furthermore, the expectations $E\{\hat{E}(x)\}$ and $E\{\hat{\Gamma}(t,x)\}$ differ from $E(x)$ or $\Gamma(t,x)$, respectively, uniformly over $x\in A,\delta_T\leq t\leq T-\delta_T$ by a term of order $O(b^2)$. For $\hat{g}_{t,x_*}(z)$ we have that for $\Delta_{g,t}(z)=\hat{g}_{t,x_*}(z)-g_{t,x_*}(z)$ uniformly for $\delta_T\leq t\leq T-\delta_T$ and $z\in A$

$$\sup_{z\in A,\delta_T\leq t\leq T-\delta_T}|\Delta_{g,t}(z)|=O_p\left\{\left(\frac{\log n}{nb^2}\right)^{\frac{1}{2}}\right\},$$

$$\sup_{z,x\in A,|z-x|\le\delta,\delta_T\le t\le T-\delta_T}\frac{|\Delta_{g,t}(z)-\Delta_{g,t}(x)|}{|z-x|^\kappa}=O_p\left\{\left(\frac{\log n}{nb^{2+2\kappa}}\right)^{\frac{1}{2}}\right\}$$

for $0<\kappa<1$.

LEMMA B3. *It holds that*

$$\sup_{z\in A}|\alpha^*(z)-\alpha(z)|=o_P(n^{-1/2}/\log n).$$

Proof. Note that

$$\begin{aligned}\hat{E}(z)\{\alpha^*(z)-\alpha(z)\}&=n^{-1}\sum_{k=1}^n\int K_b\{z,X_k(s)\}Z_k(s)[\alpha\{X_k(s)\}-\alpha(z)]\mathrm{d}s\\&=n^{-1}\sum_{k=1}^n W_k^1(z)+W_k^2(z),\end{aligned}$$

with

$$\begin{aligned}W_k^1(z)&=\int K_b\{z,X_k(s)\}Z_k(s)\alpha'(z)\{X_k(s)-z\}\mathrm{d}s,\\W_k^2(z)&=2^{-1}\int K_b\{z,X_k(s)\}Z_k(s)\alpha''\{X_k^*(s,z)\}\{X_k(s)-z\}^2\mathrm{d}s,\end{aligned}$$

where $X_k^*(s,z)$ lies between z and $X_k(s)$. Note that $|W_k^2(z)|\le CV_k(z)b^2$. Because of (B6) this shows

$$n^{-1}\sum_{k=1}^n W_k^2(z)=O_p(\log nb^2)=o_P(n^{-1/2}/\log n).$$

Furthermore we have that $|b^{-1}W_k^1(z)|\le CV_k(z)$ for some other constant C. Thus we get the moment bounds of Lemma B1 which hold uniformly for $1\le k\le n$, $z\in A$. By application of Bernstein's inequality, see Lemma 5.7 in van de Geer (2000), we get that

$$\mathrm{pr}\left(n^{-1}\left|\sum_{k=1}^n b^{-1}\left[W_k^1(z)-E\{W_k^1(z)\}\right]\right|\ge\xi\right)\le 2\exp\left(-\frac{n\xi^2}{C_1n^{-1/2}\xi+C_2}\right),\tag{B7}$$

for all $\xi>0$ with some constants C_1 and C_2 not depending on z. We conclude that

$$\sup_{z\in A}\left|n^{-1}\sum_{k=1}^n W_k^1(z)-E\{W_k^1(z)\}\right|=O_p\{(\log n/n)^{1/2}b\}=o_P(n^{-1/2}/\log n).$$

It remains to check the order of $n^{-1}\sum_{k=1}^n E\{W_k^1(z)\}$. It can be easily checked that this sum is of order $O(b^2)=o(n^{-1/2}/\log n)$, uniformly for $z\in A$. □

B.6. Proof of Theorem 1

For simplicity, we only give the proof of Theorem 1 for the case where x_* lies in the interior of A and boundary corrected kernels are used in both smoothing steps. The proof is divided into two parts. Lemma B4 implies bounds for the error terms $R_1(t),\dots,R_3(t)$ and Lemma B5 shows weak convergence of the leading terms.

LEMMA B4. *For* $k=1,2,3$ *it holds that*

$$\sup_{t\in[\delta_T,T-\delta_T]}|R_k^*(t)|=o_p(n^{-1/2}),\tag{B8}$$

where

$$R_1^*(t)=\left|n^{-1}\sum_{i=1}^n\int_0^T\left[\hat{g}_{t,x_*}\{X_i(s)\}-g_{t,x}\{X_i(s)\}\right]dM_i(s)\right|,$$

$$R_2^*(t) = \sup_{t\in[0,T]} \left| n^{-1} \sum_{i=1}^{n} \int_0^T \left[\alpha^*\{X_i(t+s)\} - \alpha\{X_i(t+s)\}\right] Z_i(t+s) Z_i(t) K_b\{x_*, X_i(s)\} \mathrm{d}s \right|,$$

$$R_3^*(t) = \sup_{t\in[0,T]} \left| n^{-1} \sum_{i=1}^{n} \int_0^T \left\{h_{X_i(s)}(t) - h_{x_*}(t)\right\} Z_i(t+s) Z_i(t) K_b\{x_*, X_i(s)\} \mathrm{d}s \right|.$$

Proof. We start by studying $R_1^*(t)$. Choose $\kappa > 1/2$ such that $\rho(2+2\kappa) < 1$. Note that

$$R_1^*(t) = \left| n^{-1} \sum_{i=1}^{n} \int_0^T \Delta_{g,t}\{X_i(s)\} \mathrm{d}M_i(s) \right|,$$

where for $\delta_T \le t \le T - \delta_T$ the $\Delta_{g,t}$ are functions that, with probability tending to one, all lie in the class $\mathcal{G}_n$ of functions that are Holder continuous with exponent κ and a constant C and that are absolutely uniformly bounded by $Cn^{-1/6}$, see Lemma B2. For the proof of (B8) for $k = 1$ it remains to show that

$$\sup_{h\in\mathcal{G}_n} \left| n^{-1} \sum_{i=1}^{n} \int_0^T h\{X_i(s)\} \mathrm{d}M_i(s) \right| = o_p(n^{-1/2}). \tag{B9}$$

We now remark that

$$\int_0^T h\{X_i(s)\} \mathrm{d}M_i(s) = h\{X_i(T_i)\} Z_i(T_i) - E[h\{X_i(T_i)\} Z_i(T_i)]$$
$$- \int_0^T h\{X_i(s)\} \alpha\{X_i(s)\} Z_i(s) \mathrm{d}s + \int_0^T E[h\{X_i(s)\} \alpha\{X_i(s)\} Z_i(s)] \mathrm{d}s,$$

where T_i are the jumps of N_i. We now get (B9) from Lemma 5.2 in van de Geer (2000). Note that the δ-entropy of $\mathcal{G}_n$ with respect to the sup norm can be bounded by $C\delta^{-1/\kappa}$ with $\kappa > 1/2$ and that the functions in $\mathcal{G}_n$ are bounded by $Cn^{-1/6}$.

Proof for $R_2^*(t)$: With the help of Lemma B3 and (B6) we get that

$$\begin{aligned}
\sup_{\delta_T\le t\le T-\delta_T} |R_2^*(t)| &= \sup_{\delta_T\le t\le T-\delta_T} \left| n^{-1} \sum_{i=1}^{n} \int_0^{T-t} \left[\alpha^*\{X_i(t+s)\} - \alpha\{X_i(t+s)\}\right] \right. \\
&\qquad \left. \times Z_i(t+s) Z_i(t) K_b\{x, X_i(s)\} \mathrm{d}s \right| \\
&\le \sup_{x\in A} |\alpha^*(x) - \alpha(x)| \sup_{0\le t\le T-\delta_T} \left| n^{-1} \sum_{i=1}^{n} \int_0^T Z_i(s) K_b\{x, X_i(s)\} \mathrm{d}s \right| \\
&= o_p(n^{-1/2}/\log n) O_P(\log n) = o_p(n^{-1/2}),
\end{aligned}$$

which shows the claim of Lemma B4 for $k = 2$.

Proof for $R_3^*(t)$: For the proof of the claim of Lemma B4 for $k = 3$ one uses a second order Taylor expansion of $h_{x_*}(t)$ with respect to x and proceeds then as in the proof of the lemma for $k = 2$. □

Note that the lemma implies that uniformly for $\delta_T \le t \le T - \delta_T$

$$\hat{h}_{x_*}(t) - h_{x_*}(t) = \hat{\Gamma}(t, x_*)^{-1}\{A(t) + B(t)\} + o_p(n^{-1/2}).$$

This follows by application of (B2), (B3) and (B6).

LEMMA B5. *Under the assumptions of Theorem 1 it holds that*

$$n^{-1/2}\Gamma(t, x_*)^{-1}\{A(\cdot) + B(\cdot)\} \to \mathbb{G}_x$$

in distribution in $\ell^\infty([0,T])$*, where* $\mathbb{G}_{x_*}$ *is the Gaussian process as described in Theorem 1.*

Proof. Let $C[\delta_T, T-\delta_T] = \{f : [\delta_T, T-\delta_T] \to \mathbb{R} \mid f \text{ continuous}\}$. As $A(\cdot)$ and $B(\cdot)$ are continuous functions on $[\delta_T, T-\delta_T]$ a.s., we view them as elements of $C[\delta_T, T-\delta_T]$. Weak convergence then

follows from an application of a uniform central limit theorem for $C[\delta_T, T-\delta_T]$, see Theorem 7.5 of Billinglsey (1968). Specifically, we have to show that for all $t_1, \ldots, t_k$

$$n^{-1/2}\left(\Gamma(t_1,x_*)^{-1}\left\{A(t_1)+B(t_1)\right\},\ldots,\Gamma(t_k,x_*)^{-1}\left\{A(t_k)+B(t_k)\right\}\right) \to (\mathbb{G}_{x_*}(t_1),\ldots,\mathbb{G}_{x_*}(t_k))$$

in distribution. This can be easily established by application of the Lindeberg–Feller Central Limit Theorem, see Proposition 2.27 of van der Vaart (1998). Further we have to show that

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathrm{pr}\left[\sup_{|t-s|\le\delta} n^{-1/2}\left|A(t)+B(t)-\{A(s)+B(s)\}\right|\ge\varepsilon\right]=0, \tag{B10}$$

for each $\varepsilon>0$. The transition to the space $\ell^\infty([\delta_T, T-\delta_T])$ then follows as in van der Vaart (1998), Lemma 18.13. In order to prove (B10), note that this is the definition that our process is equicontinuous in t which implies the tightness of the process. We use a handy tool to prove equicontinuity in this case, see (12.51) after Theorem 12.3 in Billinglsey (1968). Define:

$$A_i(t)=\int_0^T g_{t,x_*}\{X_i(s)\}dM_i(s), \tag{B11}$$

$$B_i(t,x)=\int_0^{T-t}[\alpha\{X_i(s+t)\}-h_{X_i(s)}(t)]Z_i(s+t)Z_i(s)K_b\{x_*-X_i(s)\}ds. \tag{B12}$$

We will show that:

$$E\left(\left[n^{-1/2}\sum_{i=1}^n\{A_i(t_1)-A_i(t_2)\}\right]^2\right)\le C(t_2-t_1)^2, \tag{B13}$$

$$E\left(\left[n^{-1/2}\sum_{i=1}^n\{B_i(t_1)-B_i(t_2)\}\right]^4\right)\le C(t_2-t_1)^2, \tag{B14}$$

for $\delta_T\le t_1,t_2\le T-\delta_T$. Claim (B13) can be verified by showing that

$$\sup_{x\in A,\delta_T\le t_1,t_2\le T-\delta_T}|g_{t_1,x_*}(x)-g_{t_2,x_*}(x)|\le C|t_1-t_2|,$$

which can be easily done by making use of Assumptions [A3] and [A4]. For the proof of (B14) one shows that

$$E\left[\{B_i(t_1)-B_i(t_2)\}^2\right]\le C|t_2-t_1|,$$
$$E\left[\{B_i(t_1)-B_i(t_2)\}^4\right]\le C|t_2-t_1|^2,$$

which can be proved by applying [A3] and [A7]. This concludes the proof of the lemma. □
Combining Lemmas B4 and B5 completes the proof of Theorem 1.

B.7. Asymptotic covariance operator

In this section we will define the covariance operator of the limiting Gaussian process in Theorem 1.

First of all it can be checked that $n^{-1/2}A(\cdot)$ and $n^{-1/2}B(\cdot)$ are asymptotically independent. This can be seen by showing that $E\{n^{-1/2}A(t_1)n^{-1/2}B(t_2)\}$ converges to 0 for $\delta_T\le t_1,t_2\le T-\delta_T$. Furthermore we have that $E\{n^{-1/2}A(t_1)n^{-1/2}A(t_2)\}$ converges to

$$S_1(t_1,t_2)=E\left[\int_0^T g_{t_1,x_*}\{X_i(v)\}g_{t_2,x_*}\{X_i(v)\}\alpha\{X_i(v)\}Z_i(v)\mathrm{d}v\right].$$

Then, we have that $E\{n^{-1/2}B(t_1)n^{-1/2}B(t_2)\}$ converges to

$$S_2(t_1,t_2)=\int_0^{T-t_2}\int_0^{T-t_1}E\big([\alpha\{X_i(t_1+s_1)\}-h_{x_*}(t_1)]\,[\alpha\{X_i(t_2+s_2)\}-h_{x_*}(t_2)]$$

$$\times Z_i(t_1+s_1)Z_i(t_2+s_2) \mid X_i(s_1)=X_i(s_2)=x_*, Z_i(s_1)=Z_i(s_2)=1\big)$$
$$\times \text{pr}\{Z_i(s_1)=Z_i(s_2)=1\}\mathrm{d}s_1\mathrm{d}s_2.$$

Thus we get for the covariance operator of the limiting Gaussian process

$$\Sigma(t_1,t_2)=\Gamma(t_1,x_*)^{-1}\Gamma(t_2,x_*)^{-1}\{S_1(t_1,t_2)+S_2(t_1,t_2)\}.$$

B.8. Proof of Theorem 2

Note first that given the sample $\mathbb{X}_n$, $\bar{h}_{x_*,B}(t)$ is a mean zero Gaussian process. First we argue that the (conditional) covariance of this process converges to the covariance Σ of the limiting Gaussian process $\mathbb{G}_{x_*}$. We start by noting that

$$\bar{h}_{x_*,B}(t)=\tilde{h}_{x_*,B}(t)+o_p(1),$$

where

$$\tilde{h}_{x_*,B}(t)=\hat{\Gamma}(t,x_*)^{-1}n^{-1/2}\sum_{i=1}^{n}V_i\{A_i(t)+B_i(t)\},$$

with A_i and B_i as defined in (B11) and (B12). For a proof of this claim one makes use of arguments similar to the ones used in the proof of Lemma B4 to bound the remainder terms $R_k^*(t)$ for $k=1,2,3$. Now one can show that

$$\hat{\Gamma}(t_1,x_*)^{-1}\hat{\Gamma}(t_2,x_*)^{-1}n^{-1}\sum_{i=1}^{n}\{A_i(t_1)+B_i(t_1)\}\{A_i(t_2)+B_i(t_2)\}\to\Sigma(t_1,t_2), \tag{B15}$$

in probability, uniformly for $0\le t_1,t_2\le T-\delta_T$. We conclude that for all $\delta>0$ on an event with probability greater equal $1-\delta$ the conditional distribution of the Gaussian process $\bar{h}_{x_*,B}(\cdot)$ converges weakly to the distribution of $\mathbb{G}_{x_*}$.

Then, using slightly stronger arguments one can show that

$$E\left[\{\tilde{h}_{x,B}(t)-\bar{h}_{x_*,B}(t)\}^2 \mid \mathbb{X}_n\right]$$

converges in probability to 0, uniformly for $0\le t\le T-\delta_T$. With (B15) we conclude that

$$\hat{\sigma}^2_{\mathbb{G}_{x_*}}(t)=E\{\bar{h}_{x_*,B}(\cdot)^2 \mid \mathbb{X}_n\}$$

converges in probability to $\sigma^2_{\mathbb{G}_{x_*}}(t)=\Sigma(t,t)$. Using this convergence and the convergence of the bootstrap Gaussian process to $\mathbb{G}_{x_*}$ we get the statements of the theorem. Note that taking the maximum of function and the pointwise evaluation of a function is a continuous operation.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of probability measures*, John Wiley and Sons, New York.

BEYERSMANN, J., TERMINI, S. D. & PAULY, M. (2013). Weak convergence of the wild bootstrap for the Aalen–Johansen estimator of the cumulative incidence function of a competing risk. *Scandinavian Journal of Statistics* **40**, 387–402.

BLANCHE, P., PROUST-LIMA, C., LOUBÉRE, L., BERR, C., DARTIGUES, J.-F. & JACQMIN-GADDA, H. (2015). Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics* **71**, 102–113.

BIE, O., BORGAN, Ø. & LIESTØL, K. (1987). Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties. *Scandinavian Journal of Statistics* **14**, 221–233.

BLUHMKI, T., DOBLER, D., BEYERSMANN, J. & PAULY, M. (2019). The wild bootstrap for multivariate Nelson–Aalen estimators. *Lifetime data analysis* **25**, 97–127.

CHERNOZHUKOV, V., CHETVERIKOV, D. & KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics* **41**, 2786–2819.

GÁMIZ, M. L., MAMMEN, E., MARTÍNEZ-MIRANDA, M. D. & NIELSEN, J. P. (2016). Double one-sided cross-validation of local linear hazards. *J. Royal Stat. Society: Ser. B* **78**, 755–779.

HÄRDLE, W. & MARRON, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics* , 1465–1481.

NIELSEN, J. P. (1999). Marker dependent kernel estimation from local linear estimation. *Scandinavian Actuarial Journal* **2**, 113–224.

NIELSEN, J. P. (1999). Super-efficient hazard estimation based on high-quality marker information. *Biometrika* **86**, 227–232.

RIZOPOULOS, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. Chapman and Hall/CRC.

VAN DE GEER, S. (2000). *Empirical Processes in M-estimation*. Cambridge University Press.

VAN DER VAART, A.M. (1998). *Asymptotic Statistics*. Cambridge University Press.