City Research Online

City, University of London Institutional Repository

Citation: Bishopberger, S., Hiabu, M., Mammen, E. & Perch Nielsen, J. (2025). Smooth Backfitting for Additive Hazard Rates. Scandinavian Journal of Statistics, sjos.70004. doi: 10.1111/sjos.70004

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/35380/

Link to published version: https://doi.org/10.1111/sjos.70004

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

Check for updates

Received: 20 August 2024 | Revised: 29 May 2025 | Accepted: 16 June 2025

DOI: 10.1111/sjos.70004

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Smooth backfitting for additive hazard rates

Stephan M. Bischofberger[1] | **Munir Hiabu**[2] | **Enno Mammen**[3] | **Jens Perch Nielsen**[4]

[1]Staburo GmbH, Munich, Germany

[2]Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark

[3]Institute for Mathematics, Heidelberg University, Heidelberg, Germany

[4]Bayes Business School, City St George's, University of London, London, United Kingdom

Correspondence
Stephan M. Bischofberger, Staburo GmbH, Aschauer Str. 26a, 81549 Munich, Germany.
Email: bischofberger@staburo.de

Funding information
Deutsche Forschungsgemeinschaft

Abstract

Smooth backfitting was first introduced in an additive regression setting via a direct projection alternative to the classic backfitting method by Buja, Hastie, and Tibshirani. This paper translates the original smooth backfitting concept to a survival model considering an additively structured hazard. The model allows for censoring and truncation patterns occurring in many applications, such as medical studies or actuarial reserving. Our estimators are shown to be a projection of the data into the space of multivariate hazard functions with smooth additive components. Hence, our hazard estimator is the closest nonparametric additive fit, even if the actual hazard rate is not additive. This is different from other additive structure estimators, where it is not clear what is being estimated if the model is not true. We provide full asymptotic theory for our estimators. We propose an implementation of estimators that shows good performance in practice.

KEYWORDS

additive hazard model, local linear kernel estimation, smooth backfitting, survival analysis

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Scandinavian Journal of Statistics* published by John Wiley & Sons Ltd on behalf of The Board of the Foundation of the Scandinavian Journal of Statistics.

1 INTRODUCTION

This paper introduces a fundamental model and estimator for structured multivariate marker-dependent hazards: The smooth backfitting of additive hazards. In structured nonparametric regression, Mammen et al. (1999) modeled and estimated the additive structure by projecting data onto the appropriate additive subspace. The resulting projection estimator is known as the smooth backfitting estimator. The name comes from the fact that when calculating the projection estimator iteratively, one must not only smooth the component that is being updated, but all components. This is different from classical backfitting (Buja et al., 1989), where only the component that is being updated is smoothed. It has been shown that smooth backfitting performs much better than previous comparable smoothing kernel-based backfitting approaches, in particular in high-dimensional problems and with correlated covariates, see Nielsen and Sperlich (2005). A theoretical comparison between classical and smooth backfitting for additive regression models was recently conducted in (Huang & Yu, 2019), explaining why smoothing of all components leads to better adaptation. Since the initial smooth backfitting paper many variations and extensions have been developed using smooth backfitting to tackle more sophisticated problems in mathematical statistics, Mammen and Nielsen (2003), Yu et al. (2008), Mammen and Yu (2009), Mammen et al. (2014), Han and Park (2018), Mammen and Sperlich (2022), Bissantz et al. (2016), Han et al. (2020), Jeon and Park (2020), Hiabu, Mammen, et al. (2021) and Gregory et al. (2021).

The aim of the current paper is to transfer the original approach of additive nonparametric structures to marker-dependent hazard estimation and to allow for a potentially high number of covariates with possibly correlated markers. It turns out that when the original estimation problem is phrased as a minimization problem in the correct way via a counting process formulation, then our smooth backfitting additive hazard approach can be implemented and analyzed in a very similar way to smooth backfitting in regression. We see this as a necessary step to understand more complicated structures in marker-dependent hazards. The additive subspace is closed, making analysis more accessible, and the additive structure allows for a more immediate interpretation than more complicated models of structured hazards. One important alternative structure is the multiplicative or proportional hazard model. Survival analysis practitioners often work with such multiplicative marker-dependent hazard models, including the Cox model. Smooth backfitting for the multiplicative model was recently analyzed in Hiabu, Mammen, et al. (2021), where the analysis was challenged by the shape of the multiplicative subspace that is not closed like the additive subspace is and where some tricks had to be developed, for example, a solution weighted optimization, to arrive at a tractable estimation method and analysis. The additive approach developed in this paper does not face these two latter challenges, and it might perhaps have been more natural to develop this current paper first and then Hiabu, Mammen, et al. (2021) afterwards. Both this current paper and Hiabu, Mammen, et al. (2021) arrive at the same conclusion for smooth backfitting of marker-dependent hazard estimators as the authors in Nielsen and Sperlich (2005) did for smooth backfitting of nonparametric regression: Smoothing all components in every iteration step and not only smoothing the component that is being updated is important. Otherwise, the estimator breaks down in many cases—in particular in high dimensions—where smooth backfitting still works. Smooth backfitting seems more reliable than classical backfitting of kernel estimators, and we expect that the additive marker-dependent hazard model and estimator of this paper can be an important starting point for further developments of structured marker-dependent hazard approaches in survival analysis, just like the many developments we have seen in nonparametric regression. Code to replicate our simulation and application can be

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

found on github[1]. In the next section, we provide some insight into the additive model itself and its role in marker-dependent hazard models as a practical tool for survival analysis.

2 | ADDITIVE STRUCTURED HAZARDS AND RELATED LITERATURE

One well-known model in hazard regression is the proportional hazards model of Cox (1972), which has been seen as the natural equivalent to additive regression functions in linear and nonparametric regression. As pointed out in (Martinussen & Scheike, 2006, p. 103), additive hazard models have been "somewhat overlooked in practice" although they share the same advantages of additive regression models concerning both theoretical properties and implementation. To the best of the author's knowledge, this is still the case, with some exceptions (Aalen et al., 2019; Dukes et al., 2019; Tchetgen Tchetgen et al., 2015). However, in certain applications, an additive relationship in the hazard function is indeed more plausible than a proportional one (Beslow & Day, 1987; Kravdal, 1997; Lin & Ying, 1994; McDaniel et al., 2019). Moreover, (Aalen et al., 2008, pp. 155f) provides a variety of reasons for additive risk factors.

In the original additive hazards model (Aalen, 1980), the intensity of a counting process $\{N(t) : t \in [0, 1]\}$, conditional on the d-dimensional covariate $Z(t) = (Z_1(t), \dots, Z_d(t))^T$, satisfies

$$\lambda(t) = Z^T(t)\beta(t)Y(t) \tag{1}$$

at time t with a regression coefficient $\beta(t) = (\beta_1(t), \dots, \beta_d(t))^T$ and exposure Y which is equal to unity when an individual is at risk. An overview about this model is given in Martinussen and Scheike (2006) in which the authors praise it as a simple nondistributional model that is easy to implement. Nonparametric estimators of the cumulative regression coefficient $B(t) = \int_0^t \beta(s)\mathrm{d}s$ in model (1) have been examined in McKeague (1988) and Huffer and McKeague (1991) among others.

Model (1) imposes a linear relationship between the intensity and the value of the covariates through $Z^T(t)\beta(t)$. We loosen the assumption of linearity. Before introducing the model we investigate in this article, we describe the most general model and its disadvantages, and explain why we assume certain additive constraints. The completely nonparametric conditional intensity model

$$\lambda(t) = \alpha(t|Z)Y(t) \tag{2}$$

for a conditional hazard function α generalizes model (1) making it the most flexible model. As is common, we assume $\alpha(t|Z) = \alpha(t, Z(t))$ in this paper, that is, that the conditional hazard at time t given the covariates only depends on the values of the covariates at time t and not on the values of the past.

Model (2) has first been introduced for time-constant covariates in Beran (1981). Time dependent covariates were considered in McKeague and Utikal (1990) and Nielsen and Linton (1995). Other examples from the vast literature on nonparametric hazard estimators for this model include Van Keilegom and Veraverbeke (2001) or Spierdijk (2008). Without further structural restrictions, estimators of (2) suffer from the curse of dimensionality: The rate of convergence decreases exponentially. This is a well-known issue for unstructured nonparametric estimators, making them impractical in many cases, even in dimensions higher than, say, three. That one can not do better in the unstructured nonparametric case is known at least since Stone (1980) who provided formulas for the best possible rate of convergence for nonparametric estimators. Accordingly, the aforementioned nonparametric hazard estimators were only illustrated for the case with one-dimensional covariate Z.

To overcome this issue, one has to focus on a model that is more restrictive than the unstructured nonparametric hazard model (2). We restrict our assumptions to an additive model, which is nested within Equation (2). However, instead of the original additive Aalen model (1), we assume that the hazard rate consists of additive nonparametric components,

$$\alpha(t,z) = \alpha^* + \alpha_0(t) + \alpha_1(z_1) + \cdots + \alpha_d(z_d), \tag{3}$$

with smooth, but not further restricted, components $\alpha_k, k = 1,\dots,d$, depending on covariate values $z_1,\dots,z_d$. The constant α^* is a norming constant making the decomposition unique, as will later be further specified.

The additive model (3) is both more general but also more restrictive than the additive Aalen model (1). It is more restrictive because it does not allow the effect of covariates on the hazard to change with time. It is more general because the effect of the covariates on the hazard does not need to be linear. A very interesting model that generalizes both models is to replace each component $\alpha_k(z_k), k \geq 1$, in Equation (3) by a two-dimensional components $\alpha_k(t, z_k)$ capable of capturing a covariate effect that changes with time. While we do not consider this more general setting in this paper, we see the work done in this paper as a crucial step towards developing methods of such a more general kind. Another possible generalization is to consider multiple time scales, see, for example, Hiabu, Nielsen, and Scheike (2021).

To estimate the components in Equation (3), we propose a local polynomial least squares minimization under the constraint (3). The solution can be identified with the projection of the observation into the space of local polynomial additive hazard functions and can be calculated through a simple iterative procedure. We call the resulting estimator an additive smooth backfitting hazard estimator.

When estimating the hazard function $\alpha(t,z)$, by the nature of equation (3), it can happen that the estimate is negative at certain points. This is especially expected to happen if the underlying hazard function is far from being additive. However, it is reassuring that the smooth backfitting components $\hat{\alpha}_k$ will still have a clear interpretation as an approximation of the closest additive fit. In practice, if probabilities need to be calculated, one ad-hoc solution is to use the non-additive adjusted hazard

$$\alpha^{adj}(t,z) = \max(\alpha(t,z), \varepsilon), \varepsilon \geq 0.$$

Indeed, this is also what we do in the application Section 6.1.1 for $\varepsilon = 0$ with satisfying results.

3 | THE ADDITIVE HAZARD MODEL

Let $\mathcal{T} > 0$. We observe n i.i.d. copies of the stochastic processes $\{(N(t), Y(t), Z(t)) : t \in [0, \mathcal{T}]\}$ where N is a right-continuous counting process which is zero at time zero and which has jumps of size one. We assume that Y is a left-continuous stochastic process with values in $\{0, 1\}$ and which equals unity if the observed individual is at risk. Moreover, let Z be a d-dimensional left-continuous stochastic process with $Z(t) \in [0, R]^d$, $t \in [0, \mathcal{T}]$, for some $R > 0$. The multivariate

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

process $((N_1, Y_1, Z_1), \ldots, (N_n, Y_n, Z_n))$ is assumed to be adapted to the filtration $\{\mathcal{F}_t : t \in [0, \mathcal{T}]\}$ which satisfies the *usual conditions* (Andersen et al., 1993, p. 60).

In the following, we assume that for each $i = 1, \ldots, n$, the process N_i satisfies Aalen's multiplicative intensity model, that is, that its intensity λ_i satisfies

$$\lambda_i(t) = \lim_{h \downarrow 0} h^{-1} \mathbb{E}[N_i((t+h)-) - N_i(t-) | \mathcal{F}_{t-}] = \alpha(t, Z_i(t)) Y_i(t), \tag{4}$$

where $Y_i(t)$ is indicating if individual i is at risk at time t. The function $\alpha(t, Z(t))$ is the conditional hazard rate given the covariates Z at time t. Furthermore, we assume that α satisfies the additive structure of model (3), which we write as

$$\alpha(t, Z_i(t)) = \alpha^* + \sum_{j=0}^{d} \alpha_j(X_{ij}(t))$$

with the notation $X_i(t) = (t, Z_{i1}(t), \ldots, Z_{id}(t)) \in \mathcal{X}$ for $\mathcal{X} = [0, \mathcal{T}] \times [0, R]^d$. In the sequel, we will also write $x = (t, z_1, \ldots, z_d) \in \mathcal{X}$ and henceforth $\alpha(x) = \alpha(t, z)$ for short.

Each component of the additive hazard α is only identifiable up to an additive shift. Later, we will give conditions under which each component is uniquely identified.

Model (4) allows for different kinds of filtered data, making it very flexible. These filterings include left-truncation and right-censoring, which occur in many applications of survival analysis (Martinussen & Scheike, 2006). We now illustrate how to embed left-truncated covariates and right-censored survival time into model (4). Let T denote the survival time. Left-truncation means that we observe copies of (T, Z) only on a compact subset $\mathcal{I} \subseteq \mathcal{X}$ with the property that $(t_1, Z(t_1)) \in \mathcal{I}$ and $t_2 \geq t_1$ imply $(t_2, Z(t_2)) \in \mathcal{I}$ almost surely. We allow $\mathcal{I}$ to be random but assume it is independent from T given Z. The survival time T can also be subject to right censoring with censoring time C as long as C is conditionally independent from T given the covariate process Z. This condition holds in particular if the censoring time equals one of the components of Z. Hence, under this filtering scheme, we observe n i.i.d. copies of $(\widetilde{T}, Z^*, \mathcal{I}, \delta)$, where $\delta = \mathbb{1}(T^* < C)$, $\widetilde{T} = \min(T^*, C)$, and (T^*, Z^*) is the truncated version of (T, Z), that is, (T^*, Z^*) arises from (T, Z) by conditioning on the event $\{(T, Z(T)) \in \mathcal{I}\}$.

We can now define a counting process N_i for each individual $i = 1, \ldots, n$, via

$$N_i(t) = \mathbb{1}\left\{ \widetilde{T}_i \leq t,\ \delta_i = 1 \right\},$$

with respect to the filtration $\mathcal{F}_{i,t} = \sigma\left(\left\{ \widetilde{T}_i \leq s,\ Z_i^*(s),\ \mathcal{I}_i,\ \delta_i\ :\ s \leq t \right\} \cup \mathcal{N} \right)$, for a class of null-sets $\mathcal{N}$, which completes the filtration. In this setting, it can be easily shown that, under the above assumption of $\alpha(t|Z) = \alpha(t, Z(t))$, Aalen's multiplicative intensity model (4) is satisfied with hazard rate

$$\alpha(t, z) = \lim_{h \downarrow 0} h^{-1} \mathbb{P}(T_i \in [t, t+h) |\ T_i \geq t,\ Z_i(t) = z),$$

and exposure

$$Y_i(t) = \mathbb{1}\left\{ (t, Z_i^*(t)) \in \mathcal{I}_i,\ t \leq \widetilde{T}_i \right\},$$

for individual i. The sets $\mathcal{I}_i$ are allowed to be independent random copies of $\mathcal{I}$.

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4 | THE SMOOTH BACKFITTING ESTIMATOR OF ADDITIVE HAZARDS

4.1 | Smooth backfitting hazard estimator as projection

In this and the next section, we illustrate the equivalence of projections and estimators that minimize squared errors following the line of Mammen et al. (1999) where smooth backfitting was first introduced for nonparametric regression. The idea of describing smoothing estimators as projections in a regression setting is explained in great detail in Mammen et al. (2001). In the following, we introduce this projection principle for a counting process framework.

We will introduce our estimators as a projection from a functional space $\mathcal{H}$ onto a certain subspace. The choice of the subspace implies the class of functions that can be estimated and also the class of estimators to be considered. We now specify these functional spaces as well as (semi-) norms.

We define the unrestricted functional space as

$$\mathcal{H} = \{(f^{i,j})_{i=1,\dots,n, j=0,\dots,d+1}; f^{i,j} : \mathbb{R}^{d+2} \to \mathbb{R}\},$$

and subsets $\mathcal{H}_{full}^{LC} \subseteq \mathcal{H}_{full}^{LL} \subseteq \mathcal{H}$ via

$$\mathcal{H}_{full}^{LL} = \{f \in \mathcal{H} : f^{i,j}(s,x) \text{ does not depend on } i, s\},$$

$$\mathcal{H}_{full}^{LC} = \{f \in \mathcal{H} : f^{i,j}(s,x) \text{ does not depend on } i, s;$$
$$f^{i,j}(s,x) \equiv 0 \text{ for } j = 1, \dots, d+1\}.$$

Furthermore, for additive hazard functions, we define additive subsets

$$\mathcal{H}_{add}^{LL} = \left\{ f \in \mathcal{H}_{full}^{LL} : f^{i,0}(s,x) = \sum_{j=0}^{d} g_j(x_j); f^{i,j}(s,x) = h_j(x_j),\ j = 1, \dots, d+1, \right.$$
$$\left. \text{for some functions } g_j, h_j : \mathbb{R} \to \mathbb{R} \right\},$$

$$\mathcal{H}_{add}^{LC} = \left\{ f \in \mathcal{H}_{full}^{LC} : f^{i,0}(s,x) = \sum_{j=0}^{d} g_j(x_j) \text{ for some functions } g_j : \mathbb{R} \to \mathbb{R} \right\},$$

that contain the class of local linear and local constant hazard estimators, respectively. Moreover, we define a semi-norm $\|\cdot\|$ on $\mathcal{H}$ through

$$\|f\|^2 = \int\int \frac{1}{n}\sum_{i=1}^{n} \left[f^{i,0}(s,x) + \sum_{j=0}^{d} f^{i,j+1}(s,x) \left(\frac{x_j - X_{i,j}(s)}{h} \right) \right]^2$$
$$\times Y_i(s) K_h(x - X_i(s)) \mathrm{d}s\, \mathrm{d}\nu(x),$$

for $f \in \mathcal{H}$ and where ν is a measure with strictly positive density. This semi-norm will be used to define the projection in the sequel.

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Next, we will illustrate how $\mathcal{H}$ contains both hazard functions and the observations (N_i), $i = 1,\ldots,n$. For every $\varepsilon > 0$, the data can be identified with an element $\Delta_\varepsilon N \in \mathcal{H}$ via

$$\Delta_\varepsilon N^{i,0}(s,x) = \frac{1}{\varepsilon}\int_s^{s+\varepsilon} \mathrm{d}N_i(s), \qquad \Delta_\varepsilon N^{i,j}(s,x) \equiv 0, \qquad j = 1,\ldots,d+1.$$

We define the unstructured local constant and local linear hazard estimator as

$$\lim_{\varepsilon\to 0} \arg\min_{\theta\in\mathcal{H}^{LC}_{full}} \|\Delta_\varepsilon N - \theta\|, \qquad \lim_{\varepsilon\to 0} \arg\min_{\theta\in\mathcal{H}^{LL}_{full}} \|\Delta_\varepsilon N - \theta\|, \tag{5}$$

respectively. One can easily verify that these estimators coincide with the well-known local constant and local linear hazard marker dependent hazard estimators introduced in Nielsen and Linton (1995) and Nielsen (1998). Our estimator could be understood as a projection of unrestricted local linear smoothing introduced in Nielsen (1998).

For $\varepsilon \to 0$, each element $\Delta_\varepsilon N^{i,0}$ converges to a Dirac delta function. Hence, we write

$$\min_{\theta\in\mathcal{G}} \|\Delta N - \theta\| := \lim_{\varepsilon\to 0} \min_{\theta\in\mathcal{G}} \|\Delta_\varepsilon N - \theta\|,$$

for $\mathcal{G} \subseteq \mathcal{H}$.

We define the local constant and local linear nonparametric additive hazard estimator, respectively, as

$$\arg\min_{\theta\in\mathcal{H}^{LC}_{add}} \|\Delta N - \theta\|, \quad \arg\min_{\theta\in\mathcal{H}^{LL}_{add}} \|\Delta N - \theta\|. \tag{6}$$

For the minimization over all additive hazard functions, we can either use a direct projection into $\mathcal{H}^P_{add}$, $P \in \{LC, LL\}$ which is given by $\min_{\theta\in\mathcal{H}^P_{add}} \|\Delta N - \theta\|$ or we use a Pythagorean argument to project in two steps: For $\hat{\alpha} \in \mathcal{H}^P_{add}$, it holds $\|\Delta N - \hat{\alpha}\|^2 = \|\Delta N - \tilde{\alpha}\|^2 + \|\tilde{\alpha} - \hat{\alpha}\|^2$ with $\tilde{\alpha} \in \mathcal{H}^P_{full}$. The last identity holds because the elements $\Delta N - \tilde{\alpha}$ and $\tilde{\alpha} - \hat{\alpha}$ are orthogonal (Mammen et al., 2001). In additive marker-dependent hazard estimation, the unrestricted marker-dependent hazard estimators can be understood as intermediate in an iterative projection procedure that first projects to the unrestricted space and then to the additive space.

4.2 | Smooth backfitting hazard estimator via least squares

In the previous section, we introduced the local constant estimator as a projection from $\mathcal{H}$. In this section, we show how this connects to the better-known least squares criterion, and thereby also state the estimator in a way that is more directly mathematically tractable. We first consider the unstructured local polynomial hazard estimators. For a general understanding, we write down the general formulation for polynomials of order p, but in this paper we will only consider the local constant and the local linear case, $p = 0, 1$.

We will estimate the additive components of the hazard function via kernel smoothers. Let $k : \mathbb{R} \to \mathbb{R}$ be a symmetric and continuous kernel function such that $\int k(u)\mathrm{d}u = 1$. We define $K(u_0,\ldots,u_d) = \prod_{j=0}^d k(u_j)$. For a smoothing parameter $h > 0$, $K_h(u) = \prod_{j=0}^d k_h(u_j) = \prod_{j=0}^d h^{-1}k(h^{-1}u_j)$. In the sequel, we will use a modification of the kernel function to ensure that the kernel always integrates to unity. We replace $k_h(u - v)$ by

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$k_h(u,v) = I_{(u,v\in[0,1])}\left(\int k_h(s-v)\mathrm{d}s\right)^{-1} k_h(u-v) \tag{7}$$

for every $h > 0$ to correct for normalization at the boundaries from now on. Furthermore, we define the multivariate kernel

$$K_h(u,v) = \prod_{j=0}^{d} k_h(u_j, v_j),$$

for $u = (u_0, \ldots, u_d)$ and $v = (v_0, \ldots, v_d)$.

The unstructured pth order local polynomial estimator of the hazard function in x is defined as the first component of

$$\begin{aligned}\lim_{\varepsilon\to 0} \operatorname*{arg\,min}_{\substack{\theta_0\,:\,\mathbb{R}^{d+1}\to\mathbb{R}\\ \theta_j\,:\,\mathbb{R}^{d+1}\to\mathbb{R}^{d+1}\\ j=1,\ldots,p}} \sum_{i=1}^{n} \int\int \Bigg\{ \frac{1}{\varepsilon}\int_s^{s+\varepsilon} \mathrm{d}N_i(u) - \theta_0(x) \\ - \theta_1^T(x)\left(\frac{x_0 - X_{i0}(s)}{h}, \ldots, \frac{x_d - X_{id}(s)}{h}\right)^T - \cdots \\ -\theta_p^T(x)\left(\left(\frac{x_d - X_{id}(s)}{h}\right)^p, \ldots, \left(\frac{x_d - X_{id}(s)}{h}\right)^p\right)^T \Bigg\}^2 \\ \times K_h(x, X_i(s))Y_i(s)\mathrm{d}s\ )d\nu(x),\end{aligned} \tag{8}$$

The cases $p = 0, 1$ are exactly the local constant and local linear projection estimator defined in Equation (5).

For the rest of this paper, we limit ourselves to the same kernel k and bandwidth h for each dimension to keep the notation simple. Henceforth, if there is no confusion about the boundaries of the integrals, $\int$ denotes integration over the whole support $[0, \mathcal{T}] \times [0, R]^d$. The measure ν has to have a strictly positive density, but the estimator does not depend on the specific choice of ν if we don't have restrictions on the functions θ_j. We will specify a weighting function w such that $\mathrm{d}\nu(x) = w(x)\mathrm{d}x$. Note that this estimator allows for local polynomial approximation at degree p, but it is not additive yet.

The nonparametric additive hazard estimator we investigate in this paper is defined by the minimization in Equation (8) under the following constraints on the structural form of θ. For $p = 0$, the constraint $\theta_0(x) = \overline{\alpha}^* + \sum_{j=0}^{d} \overline{\alpha}_j(x_j)$ for some functions $\overline{\alpha}_0, \ldots, \overline{\alpha}_d$ and a constant $\overline{\alpha}^*$, leads to the local constant estimator as introduced in Equation (6):

$$\begin{aligned}\lim_{\varepsilon\to 0} \operatorname*{arg\,min}_{\substack{\overline{\alpha}^*\in\mathbb{R},\\ \overline{\alpha}_j\,:\mathbb{R}\to\mathbb{R},\\ j=0,\ldots,d}} \sum_{i=1}^{n} \int\int \left\{ \frac{1}{\varepsilon}\int_s^{s+\varepsilon} \mathrm{d}N_i(u) - \left[\overline{\alpha}^* + \overline{\alpha}_0(t) + \overline{\alpha}_1(z_1) + \cdots \overline{\alpha}_d(z_d)\right] \right\}^2 \\ \times K_h(x, X_i(s))Y_i(s)\mathrm{d}s\, \mathrm{d}\nu(x).\end{aligned} \tag{9}$$

For the unique identification of the constant component α^* and the components $\alpha_j, j = 0, \ldots, d$, we will set further constraints in Equation (13).

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The local linear additive hazard estimator as defined in Equation (6) arises by setting $\theta_0(x) = \overline{\alpha}^* + \sum_{j=0}^{d} \overline{\alpha}_j(x_j)$ and $\theta_1(x) = (\partial/\partial x_0 \theta_0(x), \ldots, \partial/\partial x_d \theta_0(x))$.

$$\lim_{\varepsilon\to 0} \underset{\substack{\overline{\alpha}^* \in \mathbb{R}, \\ \overline{\alpha}_j : \mathbb{R}\to\mathbb{R}, \\ \overline{\alpha}_j' : \mathbb{R}\to\mathbb{R}, \\ j=0,\ldots,d}}{\arg\min} \sum_{i=1}^{n} \int \int \left\{ \frac{1}{\varepsilon} \int_s^{s+\varepsilon} \mathrm{d}N_i(u) - \left[\overline{\alpha}^* + \overline{\alpha}_0(t) + \overline{\alpha}_1(z_1) + \cdots \overline{\alpha}_d(z_d) \right. \right.$$

$$\left. \left. + \ \overline{\alpha}_0'(x_0)\left(\frac{x_0 - X_{i0}(s)}{h}\right) + \cdots + \overline{\alpha}_d'(x_d)\left(\frac{x_d - X_{id}(s)}{h}\right)\right]\right\}^2$$
$$\times K_h(x, X_i(s))Y_i(s)\mathrm{d}s \ \mathrm{d}\nu(x). \tag{10}$$

Existence and uniqueness of the minimizers of (9) and (10) will be established later.

4.3 | The local constant smooth backfitting additive kernel hazard estimator

The minimization in Equation (8) for $p = 0$ leads to the unstructured local constant estimator $\hat{\alpha}^{LC}$ defined via $\hat{\alpha}^{LC}(x) = \hat{O}(x)/\hat{E}(x)$ with

$$\hat{O}(x) = \frac{1}{n}\sum_{i=1}^{n} \int K_h(x, X_i(s))\mathrm{d}N_i(s),$$
$$\hat{E}(x) = \frac{1}{n}\sum_{i=1}^{n} \int K_h(x, X_i(s))Y_i(s)\mathrm{d}s.$$

for $x \in \mathcal{X}$. The estimators $\hat{O}$ and $\hat{E}$ estimate the occurrence and exposure of the observations. The exposure E is defined via $E(x) = f_t(z)\mathbb{E}[Y(t)]$ where $f_t(z)$ is the conditional density of $(Z_1(t), \ldots, Z_d(t))$ given $Y(t) = 1$. The occurrence is defined as $O(x) = \alpha(x)E(x)$ for $x = (t, z) \in \mathcal{X}$. The structure of a hazard estimator as an estimator of occurrence divided by an estimator of exposure is in line with piece-wise constant hazard estimators in Martinussen and Scheike (2002).

To define the local constant smooth backfitting additive hazard estimators, we proceed as follows. Following the derivation in Section 4.2, the estimator is defined through Equation (9). The solution $\overline{\alpha} = (\overline{\alpha}^*, \overline{\alpha}_0, \ldots, \overline{\alpha}_d)$ satisfies the first-order conditions

$$\overline{\alpha}^* = \frac{\int_{\mathcal{X}} [\hat{\alpha}^{LC}(x) - \sum_{j=0}^{d} \overline{\alpha}_j(x_j)] w(x)\mathrm{d}x}{\int_{\mathcal{X}} w(x)\mathrm{d}x} \tag{11}$$

and

$$\overline{\alpha}_k(x_k) = \int_{\mathcal{X}_{x_k}} \hat{\alpha}^{LC}(x)\frac{w(x)}{w_k(x_k)}\mathrm{d}x_{-k} - \sum_{j\neq k} \int_{\mathcal{X}_{x_k}} \overline{\alpha}_j(x_j)\frac{w(x)}{w_k(x_k)}\mathrm{d}x_{-k} - \overline{\alpha}^*, \tag{12}$$

for $k = 0, \ldots, d$, where we write $w_k(x_k) = \int_{\mathcal{X}_{x_k}} w(x)\mathrm{d}x_{-k}$ for the marginals of w using the notation $\mathcal{X}_{x_k} = \{y \in \mathcal{X} : y_x = x_k\}$ and dx_{-k} denoting integration over all components except for k. For the

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

unique identification of the solution, we also set the conditions

$$\int_{\mathcal{X}_k} \overline{\alpha}_k(x_k) w_k(x_k) \mathrm{d}x_k = 0, \qquad k = 0, \dots, d. \tag{13}$$

These identification conditions enable us to further get

$$\overline{\alpha}^* = \frac{\int_{\mathcal{X}} \hat{\alpha}^{LC}(x) w(x) \mathrm{d}x}{\int_{\mathcal{X}} w(x) \mathrm{d}x} = \frac{\int_{\mathcal{X}} \hat{O}(x) \mathrm{d}x}{\int_{\mathcal{X}} \hat{E}(x) \mathrm{d}x}$$

from Equation (11), where the second equality arises from the definition of $\hat{\alpha}$ and if we set the weighting to $w(x) = \hat{E}(x)$. One can further reduce the estimator to

$$\overline{\alpha}^* = \frac{\sum_{i=1}^n \int \mathrm{d}N_i(s)}{\sum_{i=1}^n \int Y_i(s) \mathrm{d}s}. \tag{14}$$

This simplification is due to the normalization $\int K_h(x, X_i(s)) \mathrm{d}x = 1$ of the kernel function K_h in Equation (7). The estimator $\overline{\alpha}^*$ is the additive hazard equivalent of the intercept in non-parametric regression. Note that in backfitting of the regression function m in Mammen et al. (1999), the estimator for the additive constant m_0 of the conditional mean m is given as $\tilde{m}_0 = \overline{Y}_n$. Our result for $\overline{\alpha}^*$ is the total number of occurrences divided by the average exposure time. In the case of non-filtered data, $\int \mathrm{d}N_i(s)$ equals unity for every i and thus $\overline{\alpha}^* = \left(\frac{1}{n}\sum_{i=1}^n \int Y_i(s)\mathrm{d}s\right)^{-1}$. This term is the natural survival analysis equivalent of the empirical mean in regression.

The constant component α^* and all components α_j of the unknown underlying hazard α are uniquely identified through

$$\int \alpha_j(x_j) E_j(x_j) \mathrm{d}x_j = 0 \tag{15}$$

with $E_j(x_j) = \int E(x) \mathrm{d}x_{-j}$ for all j. This motivates the choice $w(x) = \hat{E}(x)$ in Equation (13) and the notation $\hat{E}_k(x_k)$ instead of $w_k(x_k)$ for this choice of weighting from now on.

For the same data-adaptive weighting, we simplify the terms in Equation (12) with some new notation. Analogously to the one-dimensional marginals, we write $\hat{E}_{k,j}(x_k, x_j) = \int_{\mathcal{X}_{x_k x_j}} \hat{E}(x) \mathrm{d}x_{-(k,j)}$ for $x_{-(k,j)} = (x_0, \dots, x_{j-1}, x_{j+1}, \dots, x_{k-1}, x_{k+1}, \dots, x_d)$ and $\mathcal{X}_{x_k, x_j} = \{(x'_0, \dots, x'_d) \in \mathcal{X} : x'_k = x_k, x'_j = x_j\}$, that is, we integrate over all components except for x_j and x_k which are fixed values. Analogously, we define the marginal occurrence estimator $\hat{O}_k(x_k) = \int_{\mathcal{X}_{x_k}} \hat{O}(x) \mathrm{d}x_{-k}$.

In the local constant case investigated here, it can be easily shown that it holds

$$\hat{O}_k(x_k) = \frac{1}{n} \sum_{i=1}^n \int k_h(x_k, X_{ik}(s)) \mathrm{d}N_i(s), \tag{16}$$

$$\hat{E}_k(x_k) = \frac{1}{n} \sum_{i=1}^n \int k_h(x_k, X_{ik}(s)) Y_i(s) \mathrm{d}s, \tag{17}$$

$$\hat{E}_{j,k}(x_j, x_k) = \frac{1}{n} \sum_{i=1}^n \int k_h(x_j, X_{ij}(s)) k_h(x_k, X_{ik}(s)) Y_i(s) \mathrm{d}s, \tag{18}$$

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for $j \neq k$ if each pair of covariates has a rectangular support. Thus, these estimators are indeed just one- and two-dimensional marginal estimators and can be computed efficiently for high dimensions $d > 2$.

Now, Equation (12) implies the backfitting equation

$$\overline{\alpha}_k(x_k) = \hat{\alpha}_k(x_k) - \sum_{j\neq k} \int_{\mathcal{X}_j} \overline{\alpha}_j(x_j)\frac{\hat{E}_{k,j}(x_k, x_j)}{\hat{E}_k(x_k)}\mathrm{d}x_j - \overline{\alpha}^*, \tag{19}$$

for the notation $\hat{\alpha}_k(x_k) = \hat{O}_k(x_k)/\hat{E}_k(x_k)$.

Using the last expression, we can get estimators for $\alpha_0,\ldots, \alpha_d$ through iterative backfitting via

$$\overline{m}_k^{[r+1]}(x_k) = \hat{\alpha}_k(x_k) - \sum_{j<k} \int \overline{\alpha}_j^{[r+1]}(x_j)\frac{\hat{E}_{k,j}(x_k, x_j)}{\hat{E}_k(x_k)}\mathrm{d}x_j - \sum_{j>k} \int \overline{\alpha}_j^{[r]}(x_j)\frac{\hat{E}_{k,j}(x_k, x_j)}{\hat{E}_k(x_k)}\mathrm{d}x_j,$$

$$\overline{\alpha}_k^{[r+1]}(x_k) = \overline{m}_k^{[r+1]}(x_k) - \left(\int \hat{E}_k(x_k)\mathrm{d}x_k\right)^{-1} \int \overline{m}_k^{[r+1]}(x_k)\hat{E}_k(x_k)\mathrm{d}x_k, \tag{20}$$

for $k = 1,\ldots, d$ in step $r + 1$. Recall that $\hat{\alpha}_k$, $k = 0,\ldots, d$, are the (non-additive) estimators which were defined via $\hat{\alpha}_k(x_k) = \hat{O}_k(x_k)/\hat{E}_k(x_k)$. We suggest to start with the initialization $\overline{\alpha}_k^{[0]}(x_k) = \hat{\alpha}_k(x_k)$, that is related to the one-dimensional local linear hazard estimator, see Nielsen and Tanggaard (2001). However, these pilot estimators can be set to different estimators. The asymptotic theory we present here is illustrated for the choice $\hat{\alpha}_k$. In Section A3 of the appendix, we illustrate how one can obtain the same estimator $\overline{\alpha}_k$ by first minimizing (8) without an additive constraint, yielding the pilot estimator $\hat{\alpha}_k$ and then running an additive minimization of $\hat{\alpha}_k$.

The complete smooth backfitting algorithm for the local constant additive hazard estimator $\overline{\alpha}$ is as follows.

1. Compute $\hat{O}_k$, $\hat{E}_k$, and $\hat{E}_{j,k}$ from Equations (16–18) and set $\hat{\alpha}_k(x_k) = \hat{O}_k(x_k)/\hat{E}_k(x_k)$ for $k, j = 0,\ldots, d$.
2. Set $r = 0$ and $\overline{\alpha}_k^{[r]} = \hat{\alpha}_k$ for $k = 0,\ldots, d$.
3. For $k = 0,\ldots, d$, compute $\overline{\alpha}_k^{[r+1]}(x_k)$ via Equation (20) for all points x_k.
4. If the convergence criterion
$$\frac{\sum_{k=0}^d \int \left(\overline{\alpha}_k^{[r+1]}(x_k) - \overline{\alpha}_k^{[r]}(x_k)\right)^2 \mathrm{d}x_k}{\sum_{k=0}^d \int \left(\overline{\alpha}_k^{[r+1]}(x_k)\right)^2 \mathrm{d}x_k + 0.0001} < 0.0001$$
is fulfilled, stop; otherwise set r to $r + 1$ and go to step 3.
5. After convergence in step r, set $\overline{\alpha}_k = \overline{\alpha}_k^{[r+1]}$ for $k = 0,\ldots, d$, and $\overline{\alpha}^* = \sum_{i=1}^n \int \mathrm{d}N_i(s)/\sum_{i=1}^n \int Y_i(s)\mathrm{d}s$.

Note that the quantities $\hat{E}_{j,k}(x_j, x_k)$, $\hat{E}_k(x_k)$, $\hat{\alpha}(x_k)$, and $\overline{\alpha}^*$ can be calculated once in the beginning and they are not updated during the iteration process. This is a computational advantage. However, we want to emphasize that the downside of the analogous local linear approach described in this section is that the local linear pilot estimator does not necessarily exist when there are

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

observations in high dimensions. The local constant estimator, on the other hand, suffers from bad performance at boundaries.

4.4 | Asymptotic properties of the local constant smooth backfitting additive kernel hazard estimator

We now derive the asymptotic behavior of the local constant estimator under weak assumptions. Indeed, we don't assume existence of $\hat{O}, \hat{E}$ but only existence of some one- and two-dimensional marginal estimators $\hat{O}_k, \hat{O}_{k,j}, \hat{E}_k, \hat{E}_{k,j}$, $j, k = 0, \ldots, d$, which is satisfied under the conditions illustrated below.

The following conditions are sufficient to derive asymptotic normality of the resulting smooth backfitting estimators $\overline{\alpha}_j$, $j = 0, \ldots, d$.

A1. The exposure satisfies $\inf_{x\in\mathcal{X}} E(x) > 0$ and its marginals E_j are differentiable for every j. Moreover, the conditional density f_t of Z given $Y(t) = 1$ is continuous for every $t \in [0, T]$ and it holds $\sup_{x\in\mathcal{X}} f_t(x) < C_f$ for some constant C_f.
A2. There exists a function $\gamma \in C^2([0, \mathcal{T}])$ such that it holds $n^{-1}\sum_{i=1}^n Y_i(t) \to \gamma(t)$ in probability as $n \to \infty$ for every $t \in [0, \mathcal{T}]$.
A3. The function k is a second-order kernel; that is, it satisfies $\int k(u)\mathrm{d}u = 1$, $\int uk(u)\mathrm{d}u = 0$. Furthermore, k is a symmetric and Lipschitz continuous function with support $[-1, 1]$.
A4. It holds $n^{1/5}h \to c_h$ for a constant $0 < c_h < \infty$ as $n \to \infty$.
A5. The hazard α is two times continuously differentiable in every component of $x \in \mathcal{X}$.

Note that in our notation, $\gamma(t)$ from A2 and $E_0(t)$ are almost surely identical. However, the definition of E_0 does not assure $E_0 \in C^2([0, \mathcal{T}])$ without A2.

Theorem 1 (Local constant smooth backfitting estimator). *Let $\hat{\alpha}_j = \hat{O}_j/\hat{E}_j$ be the pilot estimator for $j = 0, \ldots, d$. Under Assumptions A1–A5, with probability tending to 1, there exists a unique solution $\{\overline{\alpha}^*, \overline{\alpha}_j : j = 0, \ldots, d\}$ to (9), and the backfitting algorithm converges to it:*

$$\int \left[\overline{\alpha}_j^{[r]}(x_j) - \overline{\alpha}_j(x_j)\right]^2 E_j(x_j)\mathrm{d}x_j \to 0.$$

For $x_0 \in (0, \mathcal{T})$ and $x_l \in (0, R)$, $l = 1, \ldots, d$, the solution satisfies

$$n^{2/5}\left\{\begin{pmatrix} \overline{\alpha}_0(x_0) - \alpha_0(x_0) \\ \vdots \\ \overline{\alpha}_d(x_d) - \alpha_d(x_d) \end{pmatrix}\right\} \to \mathcal{N}\left(\begin{pmatrix} c_h^2 b_0(x_0) \\ \vdots \\ c_h^2 b_d(x_d) \end{pmatrix}, \begin{pmatrix} v_0(x_0) & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & v_d(x_d) \end{pmatrix}\right),$$

and in particular $\overline{\alpha}(x) = \overline{\alpha}^* + \sum_{j=0}^d \overline{\alpha}_j$ with $\overline{\alpha}^*$ from Equation (14) satisfies

$$n^{2/5}\{\overline{\alpha}(x) - \alpha(x)\} \to \mathcal{N}\left(c_h^2 \sum_{j=0}^d b_j(x_j), \sum_{j=0}^d v_j(x_j)\right),$$

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for $n \to \infty$, *where*

$$v_j(x_j) = c_h^{-1} \int k(u)^2 \mathrm{d}u \; \sigma_j^2(x_j) E_j(x_j)^{-1},$$

$$\sigma_j^2(x_j) = \alpha^* E_j(x_j)^{-1} + \sum_{l \neq j} \int \alpha_l(u) E_{jl}(x_j, u) E_j(x_j)^{-1} \mathrm{d}u + \alpha_j(x_j).$$

and where b_j *is given through*

$$(b_0, b_1, \ldots, b_d) = \arg\min_{\mathcal{B}} \int [\beta(x) - \beta_0 - \beta_1(x_1) - \cdots - \beta_d(x_d)]^2 E(x) \mathrm{d}x,$$

for

$$\beta(x) = \sum_{j=0}^{d} \int u^2 k(u) \mathrm{d}u \left[\alpha_j'(x_j) \frac{\partial \log E(x)}{\partial x_j} + \frac{1}{2} \alpha_j''(x_j) \right],$$

and $\mathcal{B} = \{\tilde{\beta} = (\beta_0, \beta_1, \ldots, \beta_d) : \int \beta_j(x_j) E_j(x_j) \mathrm{d}x_j = 0; j = 0, \ldots, d\}$.

The proof of Theorem 1 is given in Appendix A1.

Remark 1. Define the martingale $M_i = N_i - \Lambda_i$ where Λ_i is the compensator of N_i. The term $\int k(u)^2 \mathrm{d}u \; \sigma_j^2(x_j) E_j(x_j)$ occurs as the asymptotic variance of the martingale $\int k_h(x_j, X_{ij}(s)) \mathrm{d}M_i(s)$. The convergence rate is the same as for a one-dimensional local constant hazard estimator, see, for example, Nielsen and Tanggaard (2001). In the nonparametric regression setting $Y = m(X) + \varepsilon$ of Mammen et al. (1999), and in contrast to our hazard estimator, the asymptotic variance under certain regularity conditions is specified through $\sigma_j^2(x_j) = \mathrm{Var}(Y - m(X) | X_j = x_j)$ without any closed form expression.

Remark 2. By Lemma 1 in the appendix, $\tilde{\alpha}^*$ is an unbiased estimator of α^* if the identification conditions $\int \alpha_j(x_j) E_j(x_j) \mathrm{d}x_j = 0$ hold for $j = 0, \ldots, d$.

4.5 | The local linear smooth backfitting additive kernel hazard estimator

The local linear smooth backfitting estimator $\tilde{\alpha}_j(x_j)$ for $j = 0, \ldots, d$, can be described by the minimization in Equation (10). As described in Section 4.2, this is equivalent to the minimization in Equation (8) for $p = 1$ with respect to $(\hat{\alpha}, \hat{\alpha}^{(1)})$ under the constraints $\theta_0(x) = \hat{\alpha}^* + \sum_{j=0}^{d} \hat{\alpha}_j(x_j)$, $\theta_{1,j}(x_j) = \hat{\alpha}_j^{(1)}(x_j)$ for a certain weighting function w.

Denoting the estimator of derivatives α_j' by $\tilde{\alpha}^j$ in the following, the first-order conditions for the minimization in $\tilde{\alpha}_j(x_j) + \tilde{\alpha}^*$ and $\tilde{\alpha}^j(x_j)$ can be written as

$$[\tilde{\alpha}_j(x_j) + \tilde{\alpha}^*] \hat{V}^j(x_j) + \tilde{\alpha}^j(x_j) \hat{V}_j^j(x_j) = \frac{1}{n} \sum_{i=1}^{n} \int k_h(x_j, X_{ij}(s)) \mathrm{d}N_i(s)$$

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$
\begin{aligned}
&- \sum_{l \neq j} \int \tilde{\alpha}_l(x_l) \hat{V}^{l,j}(x_l, x_j) \mathrm{d}x_l && (21)\\
&- \sum_{l \neq j} \int \tilde{\alpha}^l(x_l) \hat{V}^{l,j}_l(x_l, x_j) \mathrm{d}x_l
\end{aligned}
$$

$$
\begin{aligned}
[\tilde{\alpha}_j(x_j) + \tilde{\alpha}^*]\hat{V}^j_j(x_j) + \tilde{\alpha}^j(x_j)\hat{V}^j_{j,j}(x_j) = &\frac{1}{n}\sum_{i=1}^n \int \left(\frac{x_j - X_{i,j}(s)}{h}\right) k_h(x_j, X_{ij}(s)) \mathrm{d}N_i(s),\\
&- \sum_{l \neq j} \int \tilde{\alpha}_l(x_l) \hat{V}^{l,j}_j(x_l, x_j) \mathrm{d}x_l\\
&- \sum_{l \neq j} \int \tilde{\alpha}^l(x_l) \hat{V}^{l,j}_{l,j}(x_l, x_j) \mathrm{d}x_l, \qquad (22)
\end{aligned}
$$

with the new notation

$$
\hat{V}^j(x_j) = \frac{1}{n}\sum_{i=1}^n \int k_h(x_j, X_{ij}(s)) Y_i(s) \mathrm{d}s, \tag{23}
$$

$$
\begin{aligned}
\hat{V}^{l,j}(x_l, x_j) &= \frac{1}{n}\sum_{i=1}^n \int k_h(x_l, X_{il}(s)) k_h(x_j, X_{ij}(s)) Y_i(s) \mathrm{d}s,\\
\hat{V}^j_j(x_j) &= \frac{1}{n}\sum_{i=1}^n \int \left(\frac{x_j - X_{i,j}(s)}{h}\right) k_h(x_j, X_{ij}(s)) Y_i(s) \mathrm{d}s,\\
\hat{V}^{l,j}_l(x_l, x_j) &= \frac{1}{n}\sum_{i=1}^n \int \left(\frac{x_l - X_{i,l}(s)}{h}\right) k_h(x_l, X_{il}(s)) k_h(x_j, X_{ij}(s)) Y_i(s) \mathrm{d}s, \qquad (24)\\
\hat{V}^{l,j}_j(x_l, x_j) &= \frac{1}{n}\sum_{i=1}^n \int \left(\frac{x_j - X_{i,j}(s)}{h}\right) k_h(x_l, X_{il}(s)) k_h(x_j, X_{ij}(s)) Y_i(s) \mathrm{d}s,\\
\hat{V}^j_{j,j}(x_j) &= \frac{1}{n}\sum_{i=1}^n \int \left(\frac{x_j - X_{i,j}(s)}{h}\right)^2 k_h(x_j, X_{ij}(s)) Y_i(s) \mathrm{d}s,\\
\hat{V}^{l,j}_{l,j}(x_l, x_j) &= \frac{1}{n}\sum_{i=1}^n \int \left(\frac{x_l - X_{i,l}(s)}{h}\right)\left(\frac{x_j - X_{i,j}(s)}{h}\right) k_h(x_l, X_{il}(s)) k_h(x_j, X_{ij}(s)) Y_i(s) \mathrm{d}s.
\end{aligned}
$$

Here, x_{-k} denotes $(x_0, \dots, x_{k-1}, x_{k+1}, \dots, x_d)$ and $\mathcal{X}_{x_k}$ denotes the set $\{(x'_0, \dots, x'_d) \in \mathcal{X} : x'_k = x_k\}$.

Note that $\hat{V}^j(x_j)$ and $\hat{V}^{l,j}(x_l, x_j)$ are identical to the one- and two-dimensional local constant fits $\hat{E}_j(x_j)$ and $\hat{E}_{j,k}(x_j, x_k)$ from the local constant estimator. For simplicity of notation, we relabel them in the sequel. The terms $\hat{V}^j_j(x_j)$, $\hat{V}^{l,j}_l(x_l, x_j)$, $\hat{V}^{l,j}_j(x_l, x_j)$, $\hat{V}^j_{j,j}(x_j)$ and $\hat{V}^{l,j}_{l,j}(x_l, x_j)$ contain linear and quadratic components, which distinguish this approach from the one in the last section.

Furthermore, for $j = 0, \dots, d$ we introduce the same identification condition as Equation (13) in the local constant case and require

$$
\int \tilde{\alpha}_j(x_j) \hat{V}^j(x_j) \mathrm{d}x_j = 0 \tag{25}
$$

to get a unique solution of (21) and (22).

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

We can derive a local constant estimator from the same conditions (21) and (22) for $\hat{\alpha}_k(x_k)$ but with $\hat{\alpha}_j'(x_j)$ set to zero for every j. If we choose $w \equiv 1$, this local constant estimator coincides with the one from Section 4.3.

Conditions (21–25) uniquely define our estimator, and for the derivation of asymptotic theory (21) and (22) can be written in one equation as

$$\hat{M}_j(x_j)\begin{pmatrix} \tilde{\alpha}_j(x_j) - \hat{\alpha}_j(x_j) \\ \tilde{\alpha}^j(x_j) - \hat{\alpha}^j(x_j) \end{pmatrix} = -\tilde{\alpha}^* \begin{pmatrix} \hat{V}^j(x_j) \\ \hat{V}^j_j(x_j) \end{pmatrix} - \sum_{l \neq j} \int \hat{S}_{l,j}(x_l, x_j) \begin{pmatrix} \tilde{\alpha}_l(x_l) \\ \tilde{\alpha}^l(x_l) \end{pmatrix} \mathrm{d}x_l, \tag{26}$$

where we have used the matrices

$$\hat{M}_j(x_j) = \begin{pmatrix} \hat{V}^j(x_j) & \hat{V}^j_j(x_j) \\ \hat{V}^j_j(x_j) & \hat{V}^j_{j,j}(x_j) \end{pmatrix}, \tag{27}$$

$$\hat{S}_{l,j}(x_l, x_j) = \begin{pmatrix} \hat{V}^{l,j}(x_l, x_j) & \hat{V}^{l,j}_l(x_l, x_j) \\ \hat{V}^{l,j}_j(x_l, x_j) & \hat{V}^{l,j}_{l,j}(x_l, x_j) \end{pmatrix}, \tag{28}$$

and the one-dimensional local linear fit of the observations

$$\begin{pmatrix} \hat{\alpha}_j(x_j) \\ \hat{\alpha}^j(x_j) \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} \int \hat{M}_j(x_j)^{-1} \begin{pmatrix} 1 \\ h^{-1}(x_j - X_{ij}(s)) \end{pmatrix} k_h(x_j, X_{ij}(s))\mathrm{d}N_i(s).$$

Note, that we would get the same asymptotic result for any estimator which arises from Equation (26) by replacing $\hat{V}^j_{0,0}$, $\hat{V}^j_{0,0}$ and $(\hat{\alpha}_j, \hat{\alpha}^j)$ with asymptotically equivalent estimators that satisfy the same regularity conditions in Appendix A2.

For the implementation as an iterative algorithm, step $r+1$ of the backfitting algorithm is given by:

$$\begin{pmatrix} \hat{m}_j(x_j) \\ \tilde{\alpha}^{[r+1],j}(x_j) \end{pmatrix} = \begin{pmatrix} \hat{\alpha}_j(x_j) \\ \hat{\alpha}^j(x_j) \end{pmatrix} - \hat{M}_j(x_j)^{-1} \sum_{l \neq j} \int \hat{S}_{l,j}(x_l, x_j) \begin{pmatrix} \tilde{\alpha}^{[r]}_l(x_l) \\ \tilde{\alpha}^{[r],l}(x_l) \end{pmatrix} \mathrm{d}x_l, \tag{29}$$

$$\tilde{\alpha}^{[r+1]}_j(x_j) = \hat{m}_j(x_j) - \left(\int \hat{V}^j(u_j)\mathrm{d}u_j \right)^{-1} \int \hat{m}_j(u_j)\hat{V}^j(u_j)\mathrm{d}u_j, \tag{30}$$

for $r = 0, 1, 2, \ldots$.

Note that $\tilde{\alpha}^*$ from Equation (26) vanishes in the component $\alpha^{[r+1],j}(x_j)$ and it is made redundant in the other component by the norming condition (30). Theorem 2 assures the convergence of this estimator.

We recommend avoiding the inverse of the matrices $\hat{M}_j$ in the implementation for computational stability. Solving Equations (21) and (22) for $\tilde{\alpha}_j(x_j)$ and $\tilde{\alpha}^j(x_j)$, respectively, and first replacing $\tilde{\alpha}^j(x_j)$ in Equation (21) by its latest fit $\tilde{\alpha}^{[r],j}(x_j)$ and then $\tilde{\alpha}_j(x_j)$ in Equation (22) by $\tilde{\alpha}^{[r+1]}_j(x_j)$ in step $r+1$, we get the asymptotically equivalent, more stable backfitting equations

$$\begin{aligned} \tilde{\alpha}^{[r+1]}_j(x_j) = \hat{V}^j(x_j)^{-1}\Big(& \hat{U}^j(x_j) - \tilde{\alpha}^{[r],j}(x_j)\hat{V}^j_j(x_j) - \tilde{\alpha}^*\hat{V}^j(x_j), \\ & - \sum_{l \neq j} \int \tilde{\alpha}^{[r]}_l(x_l)\hat{V}^{l,j}(x_l, x_j)\mathrm{d}x_l - \sum_{l \neq j} \int \tilde{\alpha}^{[r],l}(x_l)\hat{V}^{l,j}_l(x_l, x_j)\mathrm{d}x_l \Big), \end{aligned} \tag{31}$$

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\tilde{\alpha}^{[r+1],j}(x_j) = \hat{V}^j_{j,j}(x_j)^{-1}\Big(\hat{U}^j_j(x_j) - \tilde{\alpha}^{[r]}_j(x_j)\hat{V}^j_j(x_j) - \tilde{\alpha}^*\hat{V}^j_j(x_j)$$
$$-\sum_{l\neq j}\int \tilde{\alpha}^{[r+1]}_l(x_l)\hat{V}^{l,j}_j(x_l,x_j)\mathrm{d}x_l - \sum_{l\neq j}\int \tilde{\alpha}^{[r],l}(x_l)\hat{V}^{l,j}_{l,j}(x_l,x_j)\mathrm{d}x_l\Big), \quad (32)$$

for step $r+1$ with the notation

$$\hat{U}^j(x_j) = \frac{1}{n}\sum_{i=1}^n \int k_h(x_j, X_{ij}(s))\mathrm{d}N_i(s), \quad (33)$$
$$\hat{U}^j_j(x_j) = \frac{1}{n}\sum_{i=1}^n \int \left(\frac{x_j - X_{ij}(s)}{h}\right)k_h(x_j, X_{ij}(s))\mathrm{d}N_i(s). \quad (34)$$

Note that $\hat{U}^j(x_j)$ is identical to $\hat{O}_j(x_j)$, the local constant occurrence estimator described in Section 4.3. We set the initialization in step $r = 0$ to $(\tilde{\alpha}^{[0]}_j(x_j), \tilde{\alpha}^{[0],j}(x_j)) = (0, 0)$.

The complete smooth backfitting algorithm for the local linear additive hazard estimator $\tilde{\alpha}$ is as follows.

1. Compute $\hat{V}^j$, $\hat{V}^{l,j}$, $\hat{V}^j_j$, $\hat{V}^{l,j}_l$, $\hat{V}^{l,j}_j$, $\hat{V}^j_{j,j}$, and $\hat{V}^{l,j}_{l,j}$ from Equations (23) and (24) and set $\hat{\alpha}(x_k) = \hat{O}_k(x_k)/\hat{E}_k(x_k)$ for $k, j = 0, \dots, d$.
2. Set $r = 0$ and $\overline{\alpha}^{[r]}_k = \hat{\alpha}_k$ for $k, j = 0, \dots, d$.
3. For $k = 0, \dots, d$, calculate for all points x_k Set $r = 1$, compute $\tilde{\alpha}^{[r+1]}_k(x_k)$ via Equations (31) and (32). Then replace $\tilde{\alpha}^{[r+1]}_j$ by
$$\tilde{\alpha}^{[r+1]}_j - \left(\int \hat{V}^j(u_j)\mathrm{d}u_j\right)^{-1}\int \tilde{\alpha}^{[r^*]}_j(u_j)\hat{V}^j(u_j)\mathrm{d}u_j.$$
4. If the convergence criterion
$$\frac{\sum_{k=0}^d \int \left(\tilde{\alpha}^{[r+1]}_k(x_k) - \tilde{\alpha}^{[r]}_k(x_k)\right)^2\mathrm{d}x_k}{\sum_{k=0}^d \int \left(\tilde{\alpha}^{[r+1]}_k(x_k)\right)^2\mathrm{d}x_k + 0.0001} < 0.0001$$
is fulfilled, stop; otherwise set r to $r+1$ and go to step 3.
5. After convergence in step r, set $\tilde{\alpha}_k = \tilde{\alpha}^{[r+1]}_k$ for $k = 0, \dots, d$, and $\tilde{\alpha}^* = \sum_{i=1}^n \int \mathrm{d}N_i(s)/\sum_{i=1}^n \int Y_i(s)\mathrm{d}s$.

4.6 | Asymptotic properties of the local linear smooth backfitting additive kernel hazard estimator

For the asymptotic behavior of $\tilde{\alpha}_j$, we assume the same Assumptions A1–A5 as for the local constant estimator.

Theorem 2 (Local linear smooth backfitting estimator). *Under Assumptions A1–A5, with probability tending to 1, there exists a unique solution $\{\tilde{\alpha}_j, \tilde{\alpha}^j : j = 0, \dots, d\}$ to (10) and the backfitting algorithm (29) converges to it:*

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\int \left[\tilde{\alpha}_j^{[r]}(x_j) - \tilde{\alpha}_j(x_j)\right]^2 E_j(x_j)\mathrm{d}x_j \to 0,$$
$$\int \left[\tilde{\alpha}^{j,[r]}(x_j) - \tilde{\alpha}^j(x_j)\right]^2 E_j(x_j)\mathrm{d}x_j \to 0.$$

For $x_0 \in (0, \mathcal{T})$ *and* $x_l \in (0, R)$, $l = 1, \dots, d$, *the solution satisfies*

$$n^{2/5}\left\{\begin{pmatrix} \tilde{\alpha}_0(x_0) - \alpha_0(x_0) + v_{n,0} \\ \vdots \\ \tilde{\alpha}_d(x_d) - \alpha_d(x_d) + v_{n,d} \end{pmatrix}\right\} \to \mathcal{N}\left(\begin{pmatrix} c_h^2 b_0(x_0) \\ \vdots \\ c_h^2 b_d(x_d) \end{pmatrix}, \begin{pmatrix} v_0(x_0) & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & v_d(x_d) \end{pmatrix}\right),$$

for $n \to \infty$, *where*

$$v_{n,j} = \int\int \alpha_j(x_j) k_h(x_j, u) E_j(u)\mathrm{d}u\, \mathrm{d}x_j,$$
$$b_j(x_j) = \frac{1}{2}\int u^2 k(u)\mathrm{d}u \left[\alpha_j''(x_j) - \int \alpha_j''(x_j)E_j(x_j)\mathrm{d}x_j\right],$$
$$v_j(x_j) = c_h^{-1}\int k(u)^2\mathrm{d}u\, \sigma_j^2(x_j)E_j(x_j)^{-1},$$
$$\sigma_j^2(x_j) = \alpha^* E_j(x_j)^{-1} + \sum_{l \neq j}\int \alpha_l(u)E_{jl}(x_j, u)E_j(x_j)^{-1}\mathrm{d}u + \alpha_j(x_j).$$

This result yields in particular

$$n^{2/5}\{\tilde{\alpha}(x) - \alpha(x)\} \to \mathcal{N}\left(c_h^2 \sum_{j=0}^{d} b_j(x_j), \sum_{j=0}^{d} v_j(x_j)\right),$$

for $\tilde{\alpha}(x) = \tilde{\alpha}^* + \sum_{j=0}^{d}\tilde{\alpha}_j(x_j)$ *with* $\tilde{\alpha}^* = \sum_{i=1}^{n}\int \mathrm{d}N_i(s)/\sum_{i=1}^{n}\int Y_i(s)\mathrm{d}s$.

The proof of Theorem 2 is given in Appendix A2.

Remark 3. Note that the convergence rate is the same as for a one-dimensional local linear hazard estimator, see, for example, Nielsen and Tanggaard (2001). Furthermore, $\tilde{\alpha}_j(x_j)$ estimates $\alpha_j(x_j) - \int \alpha_j(x_j)\hat{V}^j(x_j)\mathrm{d}x_j$ instead of $\alpha_j(x_j)$. The terms $v_{n,j}$ correct for this shift in the estimation of each component. The sum $\sum_{j=0}^{d} v_{n,j}$ vanishes as the additive adjustments cancel each other off.

The component $\tilde{\alpha}^*$ of the estimator $\tilde{\alpha}$, which estimates α^*, is identical to $\overline{\alpha}^*$ from the local constant case. Its asymptotic behavior is explained in Remark 2.

5 | SIMULATION STUDY

5.1 | Simulation setting

We assume that the survival times T_i follow a Gompertz–Makeham distribution, with hazard function given by

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\alpha(t, Z_i) = \alpha_0(t) + \sum_{k=1}^{d} \alpha_k(Z_{ik}) = e^{0.01t} + \frac{4}{\sqrt{d}} \sum_{k=1}^{d} (-1)^{k+1} \sin(\pi Z_{ik}),$$

($i = 1, \ldots, n$). We add right censoring with censoring variables C_i that follow the same distribution as T_i, except with the scale parameter divided by 1.75. The factor $4d^{-1/2}$ is chosen so that the distribution of T_i doesn't vary too much in the number of covariates d. Note that, for convenience, the components are identified differently than in Equations (25), (15). We now describe how the covariates $(Z_{i1}, \ldots, Z_{id})$ are generated. We first simulate $(\widetilde{Z}_{i1}, \ldots, \widetilde{Z}_{id})$ from a d-dimensional multi-normal distribution with mean equal 0 and $\mathrm{Corr}(Z_{ij}, Z_{il}) = \rho$ if $j \neq l$, else 1. Afterwards we set

$$Z_{ik} = 2.5\pi^{-1} \arctan(\widetilde{Z}_{ik}).$$

We repeat the procedure and take the first $i = 1, \ldots, n$ observations such that $4d^{-1/2} \sum_{k=1}^{d} (-1)^{k+1} \sin(\pi Z_{ik})$ is positive. Technically, the values of the covariates are conditioned such that the resulting hazard is positive and hence well defined.

As kernel function k, we used the Epanechnikov kernel. Performance is measured via the integrated squared error:

$$MISE_k = n^{-1} \sum_i \left(\eta_k(Z_{ik}) - \widehat{\eta}_k(Z_{i_k})\right)^2.$$

5.2 | Simulation results

We compare the performance of the local linear smooth backfitting estimator to the local constant smooth backfitting estimator. We also compare these proposed estimators to a version of the classical backfitting equivalent where only the updated component is smoothed, see Buja et al. (1989).

Figure 1 shows the estimation results for the first component from 100 simulations in a setting with sample size $n = 5{,}000$, dimension $d = 3$, and correlation $\rho = 0.5$, calculated with a MISE optimal bandwidth. We find that the classical backfitting estimators produce more noise than their smooth backfitting counterparts. The local constant smooth backfitting estimator is less smooth (more "wiggly") than the local linear version. This first impression can be further verified in Table 1: Classical backfitting estimators perform significantly worse than the smooth alternatives. The local linear classical backfitting estimator only gives sensible results in the easiest settings, that is, when $n = 5{,}000$ and or $d = 3$, while breaking down in all other cases. Another observation is that the local linear smooth backfitting estimator is almost always to be preferred over the local constant smooth backfitting estimator. Only in the most challenging setting, that is, $n = 500$, $d = 30$, did the local constant smooth backfitting estimator outperform the local linear version. But even in that case, the advantage is only by a small margin.

6 | DATA APPLICATION: THE TRACE STUDY

The TRACE study group (see, e.g., Jensen et al. (1997)) has collected information on more than 4,000 consecutive patients with acute myocardial infarction (AMI) with the aim of studying the

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

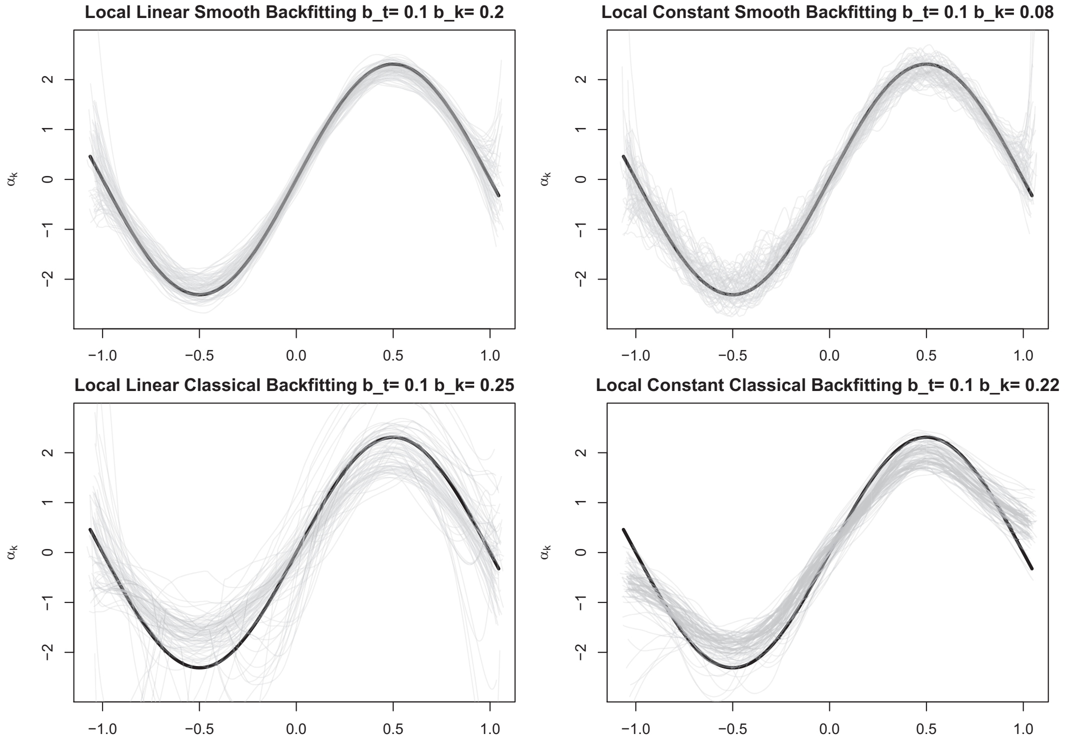

FIGURE 1 Simulation results for $k = 1$ comparing four different estimators: Local constant smooth backfitting, local linear smooth backfitting, local constant backfitting, and local linear backfitting. The gray lines represent 100 Monte Carlo simulations with MISE-optimal bandwidth estimating the true curve (black).

prognostic importance of ventricular fibrillation (vf) on mortality. We here consider a subset of these patients that are available in the `timereg` R package. We furthermore only consider those patients with more than 40 years of age, and only consider the first five years of follow-up time after the diagnosis. This results in $n = 1799$ observations. At entry, that is, time of AMI occurrence, the patients had various risk factors recorded. Here, additionally to duration, that is, time since AMI occurrence, we will consider age at AMI occurrence of the patient, a_i, and wall motion index (heart pumping effect based on ultrasound measurements where 2 is normal and 0 is worst (Scheike, 2009)), wmi_i. We will ignore additional binary covariates that have been recorded, as our framework only covers continuous covariates. With that regard, this section should be seen as a simple illustration of our theoretical work rather than a serious attempt to answer a real-world question. In summary, we consider the model

$$\lambda_i(t) = Y_i(t)\{\alpha_0(t) + \alpha_2(a_i) + \alpha_3(\mathrm{wmi}_i)\},$$

under the identifiability condition $\int \alpha_j(x_j)dx_j = 0$ for $j = 1, 2$. The initially estimated curve for α_0 can be seen in Figure 2. We find that the duration effect has two distinct periods with an increased risk in the beginning that flattens after approximately three months. This suggests that it might be beneficial to apply different amounts of smoothing to those two periods. We therefore generate two different data sets from our original data set: The first data set covers the risk in the first three months (this can be achieved by censoring all patients who survived beyond

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Simulation results comparing four different estimators: Local constant smooth backfitting, local linear smooth backfitting, local constant backfitting, and local linear backfitting.

	$n = 500$			$n = 5000$		
	MISE	**Bias^2**	**Variance**	**MISE**	**Bias^2**	**Variance**
$d = 3$						
LL-SBF	0.25	0.07	0.17	0.031	0.007	0.024
LC-SBF	0.30	0.05	0.25	0.051	0.011	0.041
LL-BF	43.14	0.69	42.46	0.779	0.041	0.737
LC-BF	1.44	0.48	0.96	0.077	0.020	0.058
$d = 10$						
LL-SBF	0.22	0.05	0.17	0.020	0.005	0.015
LC-SBF	0.24	0.08	0.17	0.030	0.006	0.025
LL-BF	1118.80	10.88	1107.91	0.135	0.057	0.078
LC-BF	1.02	0.03	0.99	0.031	0.005	0.026
$d = 30$						
LL-SBF	0.18	0.03	0.15	0.014	0.0007	0.0133
LC-SBF	0.16	0.05	0.10	0.029	0.0172	0.0114
LL-BF	NA	NA	NA	0.171	0.1494	0.0217
LC-BF	NA	NA	NA	0.033	0.0227	0.0105

Note: Values are calculated from 500 Monte Carlo simulations with MISE-optimal bandwidth.

three months) and the second data set covers the risk conditional on surviving the first three months (i.e., omits all patients in the data set with failure or censoring in the first three months). The results with our local linear estimator for the two different cohorts, that is, those with ventricular fibrillation (vf = 1) and those without ventricular fibrillation (vf = 0) can is depicted in Figures 3 and 4.

The smoothing parameter was chosen manually: For the cohort with vf = 0 we have $n = 1655$ patients when considering the first three months and chose the bandwidths for (t, a, wmi) as $(0.1, 15, 0.8)$; for the data set after surviving the first three months, we have $n = 1482$ and chose a bandwidth of $(1, 15, 0.8)$. For the cohort with vf = 1 we have $n = 132$ patients for the first three months and chose a bandwidth of (t, a, wmi) as $(0.1, 20, 0.8)$; for the data set, after surviving the first three months, we have $n = 75$ and chose a bandwidth of (t, a, wmi) as $(1, 20, 0.8)$. The dashed lines show a point-wise asymptotic 95% confidence interval based on Theorem 2. Note that it is hereby in particular assumed that (a) the bias can be neglected and (b) that the true underlying model is indeed additive. Therefore, the confidence intervals should be seen as rather optimistic. They nevertheless give an impression of the uncertainty under optimal conditions. Looking at Figure 3, we find that in the first three months, vf = 1 leads to a significant increase in mortality risk. We also find that the risk increase is more severe for older patients. Figure 4 does not provide evidence that vf = 1 leads to an increased risk after surviving the first three months. In the next section, we want to look at how confident we can be with the model results.

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

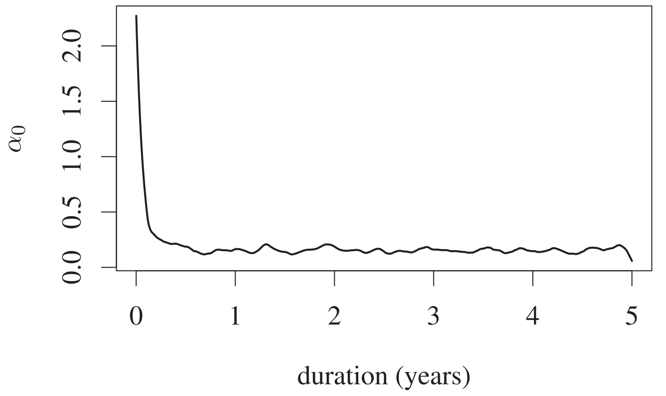

FIGURE 2 Local linear additive smooth backfitting fit of α_0 on the full data.

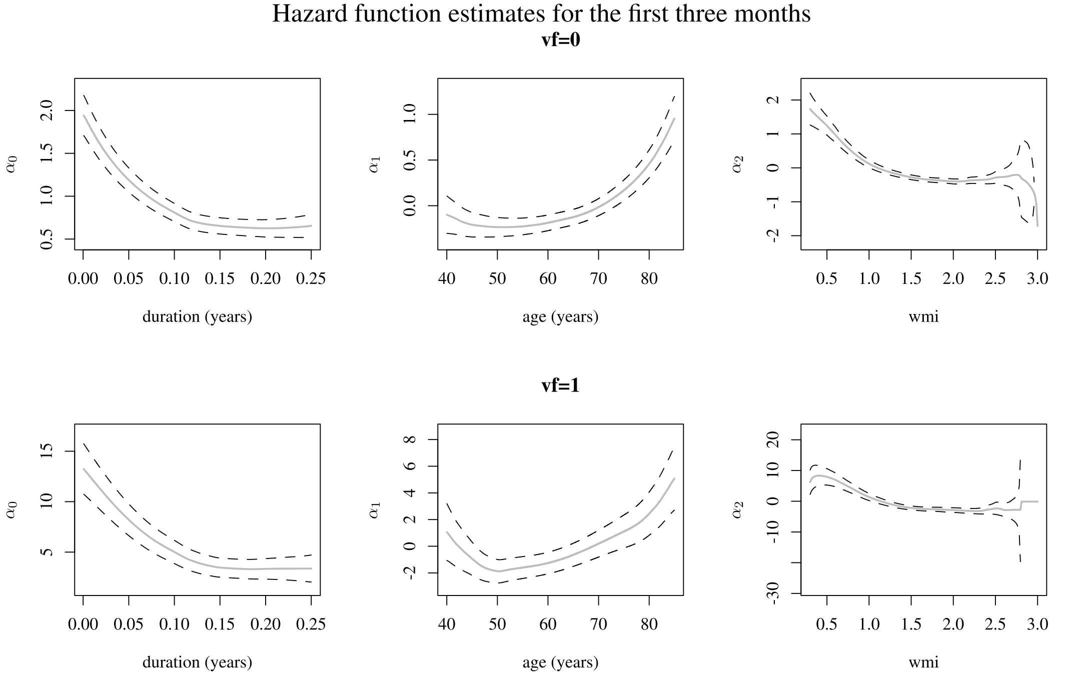

FIGURE 3 Local linear fit of $(\alpha_0, \alpha_1, \alpha_2)$ for the first three months for two different strata, depending on the value of vf. The dashed line indicates the asymptotic 95% point-wise confidence interval.

6.1 | Model robustness

6.1.1 | CRPS score

We transform our estimated hazard function $\alpha = \alpha_0 + \alpha_1 + \alpha_2$ into a plug-in estimator of the survival function via the relationship $S(t|z) = \prod_{s \le t}(1 - \alpha(s, z)\mathrm{d}s)$. We then split our data randomly into an 80% training set and a 20% test set. We train our model on the training set and evaluate the CPRS score (Avati et al., 2020) on the test set (note that a lower score indicates better

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

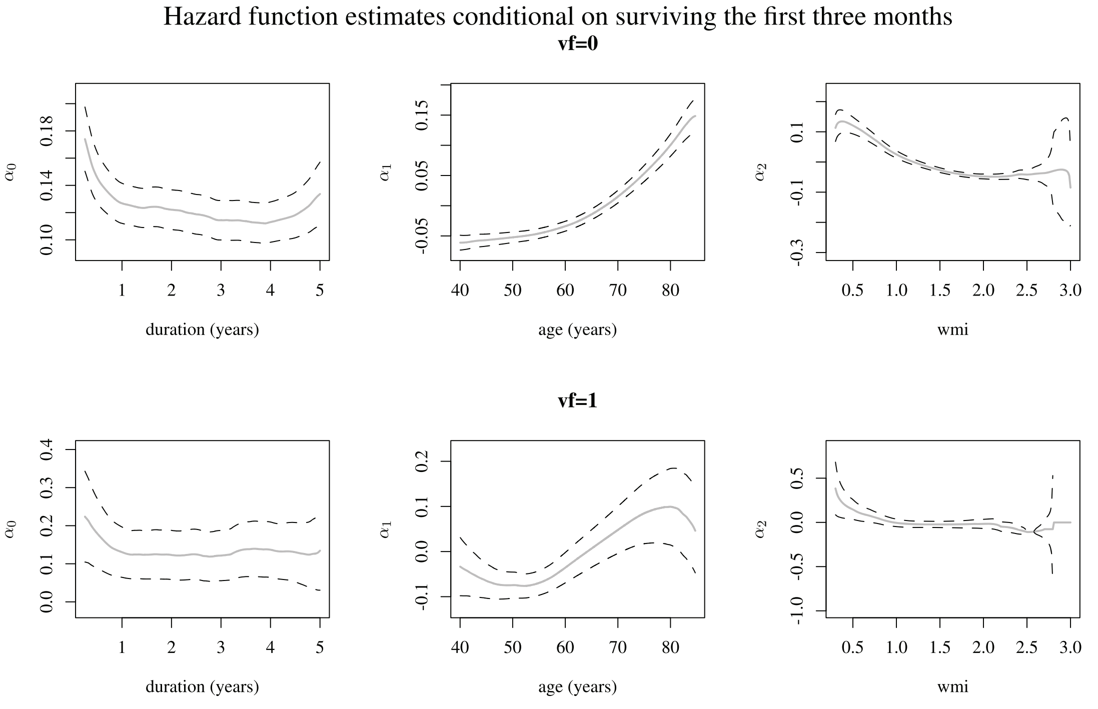

FIGURE 4 Local linear fit of $(\alpha_1, \alpha_2, \alpha_3)$ conditional on surviving the first three months for two different strata, depending on the value of vf. The dashed line indicates the asymptotic 95% point-wise confidence interval.

performance):

$$CRPS = m^{-1} \sum_{i=1}^{m} \int_0^{T_i} (1 - \hat{S}(s|z_i))^2 \mathrm{d}s + \delta_i \int_{T_i}^{\infty} \hat{S}(s|z_i)^2 \mathrm{d}s,$$

where m is the size of the test set. Due to the additive structure, our survival prediction—although consistent—can still be negative. We therefore consider a simple adjustment where we numerically calculate

$$\hat{S}^{\mathrm{adj}}(s|z) = \prod_{s \le t} (1 - \hat{\alpha}^{adj}(s,z)\mathrm{d}s), \quad \hat{\alpha}^{adj}(s,z) = \max(\hat{\alpha}(s,z), 0).$$

Lastly, we compare our local linear additive fit with the local constant multiplicative smooth backfitting estimator from Hiabu, Mammen, et al. (2021). The results from 200 simulation runs can be seen in Figure 5. We have two main observations. Firstly, the model choice does not seem to have a big impact when considering survival conditional on surviving the first three months. Secondly, for survival during the first three months, using the adjusted survival probability estimates improves the performance, but even better performance can be achieved by using a multiplicative model. Nevertheless, we want to emphasize that our smooth backfitting additive estimators have the desirable projection property that, if the additive model assumption is violated, the estimators converge to the closest additive fit, making the results still interpretable. We investigate this property in the next subsection.

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

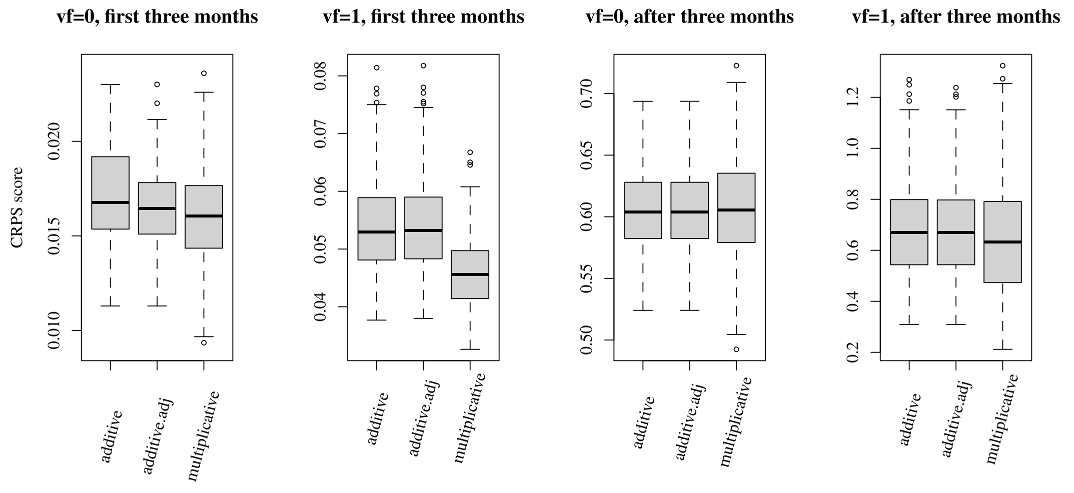

FIGURE 5 CPRS scores from 200 simulations of an 80/20 training-test-split. Boxplots are given for the four different data sets as described on top of the plots and each time for three different models: Smooth backfitting additive model, smooth backfitting additive model using the adjusted survival estimates $\hat{S}^{\mathrm{adj}}(s)$ and the smooth backfitting multiplicative model from Hiabu, Nielsen, and Scheike (2021).

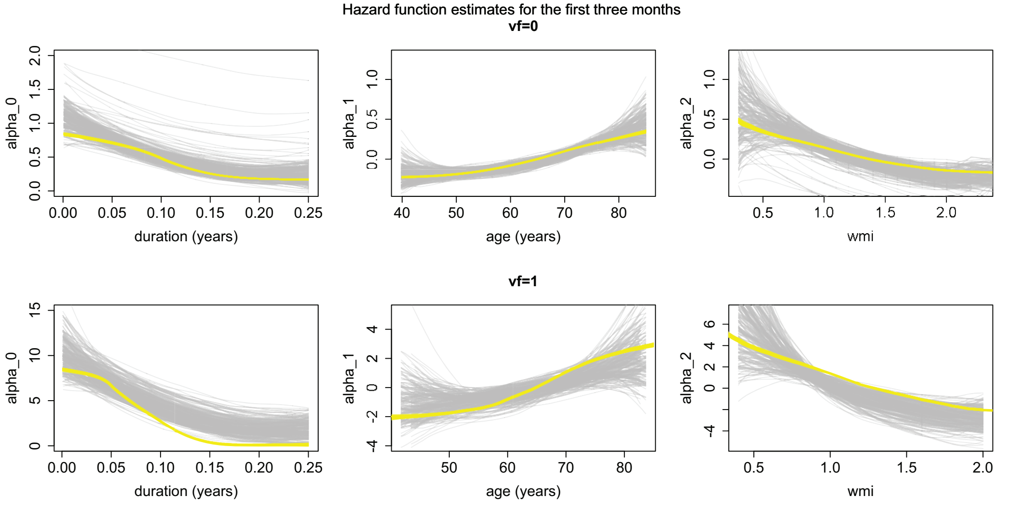

FIGURE 6 200 simulations from a multiplicative hazard model, see Figure B1. Gray curves are fitted local linear smooth backfitting estimators. Yellow curves are approximately optimal additive fits derived from a smooth backfitting additive regression fit with the true hazard as response and an inflated sample size of 10,000.

6.1.2 | Stability under model misspecification

We take the estimated multiplicative smooth backfitting model from the previous subsection, see also Figures B1 and B2 in the Appendix, as a true model and investigate how, in this case, our additive estimator would look. When generating the four data sets (vf $= 0, 1$; risk in the first three months, risk conditional on surviving the first three months), we keep the same number of samples as in the original data sets while sampling (a, wmi) with replacement from the original

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

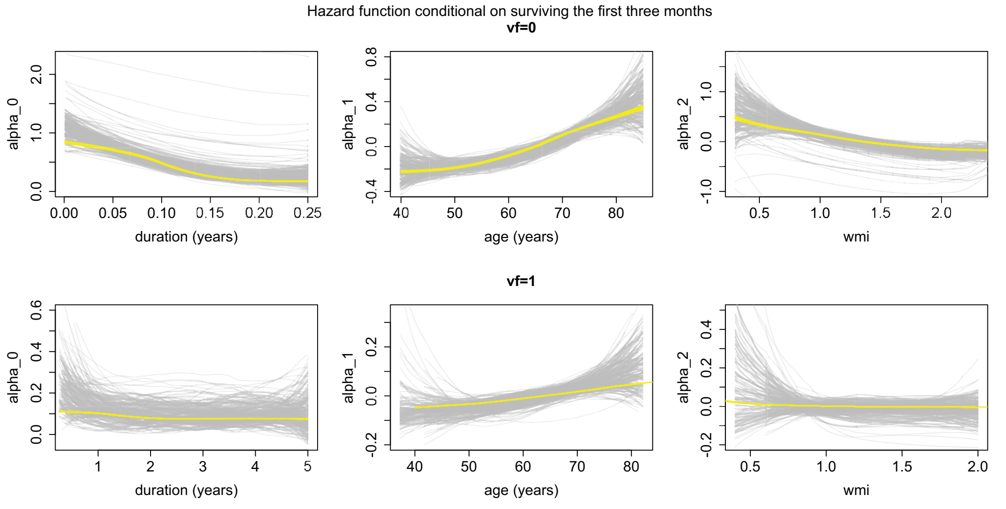

FIGURE 7 200 simulations from a multiplicative hazard model, see Figure A2. Gray curves are fitted local linear additive smooth backfitting estimates. Yellow curves are approximately optimal additive fits derived from a smooth backfitting additive regression estimator with the true hazard as response and an inflated sample size of 10,000.

data sets. Afterwards, for each row, we draw a survival time from the multiplicative smooth backfitting model. The survival time is considered censored if it is greater than 0.25 when considering the first three months, and it is considered censored if it is greater than 5 when considering the period after the first three months.

We compare our additive smooth backfitting estimator to a somewhat optimal fit. Note that it is not clear how to derive an optimal fit analytically or even numerically, as it depends on the joint distribution of duration, age, and wmi; which is not known. Therefore, we approximate the optimal fit by estimating an additive smooth backfitting regression function (Hiabu et al., 2023; Mammen et al., 1999) based on 10,000 observations where the response is the known hazard. We consider 200 simulations and the fact that the regression estimator does not vary much as a good indicator, giving us confidence that it is a good approximation of the optimal additive fit. The results are given in Figures 6 and 7. We find that our proposed estimators (grey lines)—despite the limited sample sizes—are reasonably close to the regression fit, such that we can conclude that our approach is working reasonably well in estimating the optimal additive fit. Lastly, it should be noted that we also tried a classical backfitting approach with kernel smoothers, with the result that the estimators for all components diverged in every simulation run and did not provide any result.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) through the Research Training Group RTG 1953.

CONFLICT OF INTEREST STATEMENT

None of the authors have a conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in timereg at https://cran.r-project.org/web/packages/timereg/index.html.

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ENDNOTE

[1]https://github.com/MHiabu/Replicate-Smooth-Backfitting-for-Additive-Hazard-Rates

ORCID

Stephan M. Bischofberger https://orcid.org/0000-0002-2751-7419
Munir Hiabu https://orcid.org/0000-0001-5846-667X

REFERENCES

Aalen, O. O. (1980). *A model for nonparametric regression analysis of counting processes*. In W. Klonecki, A. Kozek, & J. Rosiński (Eds.), *Mathematical statistics and probability theory* (pp. 1–25). Springer New York.

Aalen, O. O., Borgan, O., & Gjessing, H. (2008). *Survival and event history analysis: A process point of view*. Springer Science & Business Media.

Aalen, O. O., Stensrud, M., Didelez, V., Daniel, R., Røysland, K., & Strohmaier, S. (2019). Time-dependent mediators in survival analysis: Modeling direct and indirect effects with the additive hazards model. *Biometrical Journal*, *62*, 532–549. https://doi.org/10.1002/bimj.201800263

Andersen, P., Borgan, O., Gill, R., & Keiding, N. (1993). *Statistical models based on counting processes*. Springer.

Avati, A., Duan, T., Zhou, S., Jung, K., Shah, N. H., & Ng, A. Y. (2020). *Countdown regression: Sharp and calibrated survival predictions*. In *Uncertainty in artificial intelligence* (pp. 145–155). PMLR.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Deptartment of Statistics, University of California, Berkeley.

Beslow, N. E., & Day, N. E. (1987). *Statistical methods in cancer research* (Vol. *2*). The Design and Analysis of Cohort Data.

Bissantz, N., Dette, H., Hildebrandt, T., & Bissantz, K. (2016). Smooth backfitting in additive inverse regression. *Annals of the Institute of Statistical Mathematics*, *68*(4), 827–853.

Buja, A., Hastie, T., & Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, *17*(2), 453–510.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–220.

Dukes, O., Martinussen, T., Tchetgen Tchetgen, E. J., & Vansteelandt, S. (2019). On doubly robust estimation of the hazard difference. *Biometrics*, *75*(1), 100–109.

Gregory, K., Mammen, E., & Wahl, M. (2021). Statistical inference in sparse high-dimensional additive models. *The Annals of Statistics*, *49*(3), 1514–1536.

Han, K., Müller, H.-G., & Park, B. U. (2020). Additive functional regression for densities as responses. *Journal of the American Statistical Association*, *115*, 997–1010.

Han, K., & Park, B. U. (2018). Smooth backfitting for errors-in-variables additive models. *The Annals of Statistics*, *46*, 2216–2250.

Hiabu, M., Mammen, E., Martínez-Miranda, M. D., & Nielsen, J. P. (2021). Smooth backfitting of proportional hazards with multiplicative components. *Journal of the American Statistical Association*, *116*(536), 1983–1993.

Hiabu, M., Mammen, E., & Meyer, J. T. (2023). *Local linear smoothing in additive models as data projection*. In D. Belomestny, C. Butucea, E. Mammen, E. Moulines, M. Reiß, & V. V. Ulyanov (Eds.), *Foundations of modern statistics* (pp. 197–223). Springer International Publishing. 978-3-031-30114-8.

Hiabu, M., Nielsen, J. P., & Scheike, T. H. (2021). Nonsmooth backfitting for the excess risk additive regression model with two survival time scales. *Biometrika*, *108*(2), 491–506.

Huang, L.-S., & Yu, C.-H. (2019). Classical backfitting for smooth-backfitting additive models. *Journal of Computational and Graphical Statistics*, *28*(2), 1–22.

Huffer, F. W., & McKeague, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association*, *86*(413), 114–129.

Jensen, G. V. H., Torp-Pedersen, C., Hildebrandt, P., Kober, L., Nielsen, F., Melchior, T., Joen, T., & Andersen, P. (1997). Does in-hospital ventricular fibrillation affect prognosis after myocardial infarction? *European Heart Journal*, *18*(6), 919–924.

Jeon, J. M., & Park, B. U. (2020). Additive regression with Hilbertian responses. *The Annals of Statistics*, *48*, 2671–2697.

Kravdal, Ø. (1997). The attractiveness of an additive hazard model: An example from medical demography. *European Journal of Population=Revue Européenne de Démographie*, *13*(1), 33–47.
Lin, D. Y., & Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, *81*(1), 61–71.
Linton, O. B., Nielsen, J. P., & Van de Geer, S. (2003). Estimating multiplicative and additive hazard functions by kernel methods. *The Annals of Statistics*, *31*(1), 464–492.
Mammen, E., Linton, O., & Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics*, *27*, 1443–1490.
Mammen, E., Marron, J. S., Turlach, B. A., & Wand, M. P. (2001). A general framework for constrained smoothing. *Statistical Science*, *16*, 232–248.
Mammen, E., & Nielsen, J. P. (2003). Generalised structured models. *Biometrika*, *90*, 551–566.
Mammen, E., Park, B. U., & Schienle, M. (2014). *Additive models: Extensions and related models*. In J. S. Racine, L. Su, & A. Ullah (Eds.), *The Oxford handbook of applied nonparametric and semiparametric econometrics and statistics*. Oxford University Press.
Mammen, E., & Sperlich, S. (2022). Backfitting tests in generalized structured models. *Biometrika*, *109*(1), 137–152.
Mammen, E., & Yu, K. (2009). Nonparametric estimation of noisy integral equations of the second kind. *Journal of the Korean Statistical Society*, *38*, 99–110.
Martinussen, T., & Scheike, T. H. (2002). A flexible additive multiplicative hazard model. *Biometrika*, *89*(2), 283–298.
Martinussen, T., & Scheike, T. H. (2006). *Dynamic regression models for survival data*. Springer.
McDaniel, L. S., Yu, M., & Chappell, R. (2019). Analysis and design of clinical trials using additive hazards survival endpoints. *Statistics in Biopharmaceutical Research*, *11*(3), 274–282.
McKeague, I. W. (1988). *Asymptotic theory for weighted least squares estimators in aalen's additive risk model*. In *Statistical inference from stochastic processes: Proceedings of the AMS-IMS-SIAM joint summer research conference held august 9–15, 1987, with support from the National Science Foundation and the Army Research Office* (Vol. *80*, pp. 139–152). American Mathematical Society.
McKeague, I. W., & Utikal, K. J. (1990). Inference for a nonlinear counting process regression model. *The Annals of Statistics*, *18*(3), 1172–1187.
Nielsen, J. P. (1998). Marker dependent kernel hazard estimation from local linear estimation. *Scandinavian Actuarial Journal*, *1998*(2), 113–124.
Nielsen, J. P., & Linton, O. B. (1995). Kernel estimation in a non-parametric marker dependent hazard model. *The Annals of Statistics*, *23*, 1735–1748.
Nielsen, J. P., & Sperlich, S. (2005). Smooth backfitting in practice. *Journal of the Royal Statistical Society: Series B*, *67*, 43–61.
Nielsen, J. P., & Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. *Scandinavian Journal of Statistics*, *28*, 675–698.
Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics*, *11*, 453–466.
Scheike, T. (2009). Timereg package. *R package version*, 3(0):1–1.
Spierdijk, L. (2008). Nonparametric conditional hazard rate estimation: A local linear approach. *Computational Statistics & Data Analysis*, *52*(5), 2419–2434.
Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, *8*(6), 1348–1360.
Tchetgen Tchetgen, E. J., Walter, S., Vansteelandt, S., Martinussen, T., & Glymour, M. (2015). Instrumental variable estimation in a survival context. *Epidemiology*, *26*(3), 402–410.
Van Keilegom, I., & Veraverbeke, N. (2001). Hazard rate estimation in nonparametric regression with censored data. *Annals of the Institute of Statistical Mathematics*, *53*(4), 730–745.
Yu, K., Park, B. U., & Mammen, E. (2008). Smooth backfitting in generalized additive models. *The Annals of Statistics*, *36*, 228–260.

How to cite this article: Bischofberger, S. M., Hiabu, M., Mammen, E., & Nielsen, J. P. (2025). Smooth backfitting for additive hazard rates. *Scandinavian Journal of Statistics*, 1–45. https://doi.org/10.1111/sjos.70004

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

APPENDIX A

A1 Asymptotic theory for the local constant estimator

For the proof of Theorem 1, we apply the general theory for smooth backfitting estimators. We split the estimator into a stochastic part and a part consisting of its bias plus a function that vanishes. For counting processes martingales, these two parts are usually referred to as the variable and the stable part, respectively. One has to show three things: The convergence of the backfitting algorithm, the asymptotic normality of the stochastic part, and that the bias part vanishes asymptotically. In Mammen et al. (1999), conditions for these three properties have been stated for a nonparametric regression setup. The main part of our proof is to verify these conditions under Assumptions A1–A5. For completeness, we restate the modified conditions in our notation.

We also state propositions from Mammen et al. (1999), adapted to our notation, which imply the properties we need if the following assumptions hold. The difference to Mammen et al. (1999) is that we make use of martingale properties and counting process theory instead of the usual arguments for kernel density estimators.

We start with assumptions about the marginal exposures and convergence of marginal exposure estimators. Note that we don't assume any particular definition of $\hat{E}_j$ and $\hat{E}_{j,k}$, $j, k = 0, \ldots, d$, for the following propositions.

B1. For all $j \neq k$ it holds

$$\int \frac{E_{j,k}(x_j, x_k)^2}{E_j(x_j)E_k(x_k)} \mathrm{d}x_j \ \mathrm{d}x_k < \infty.$$

B2. It holds

$$\int \left[\frac{\hat{E}_j(x_j) - E_j(x_j)}{E_j(x_j)}\right]^2 E_j(x_j)\mathrm{d}x_j = o_P(1),$$

$$\int \left[\frac{\hat{E}_{j,k}(x_j, x_k)}{E_j(x_j)E_k(x_k)} - \frac{E_{j,k}(x_j, x_k)}{E_j(x_j)E_k(x_k)}\right]^2 E_j(x_j)E_k(x_k)\mathrm{d}x_j \ \mathrm{d}x_k = o_P(1),$$

$$\int \left[\frac{\hat{E}_{j,k}(x_j, x_k)}{\hat{E}_j(x_j)E_k(x_k)} - \frac{E_{j,k}(x_j, x_k)}{E_j(x_j)E_k(x_k)}\right]^2 E_j(x_j)E_k(x_k)\mathrm{d}x_j \ \mathrm{d}x_k = o_P(1).$$

Moreover, $\hat{E}_j$ vanishes outside the support of E_j, $\hat{E}_{j,k}$ vanishes outside the support of $E_{j,k}$ and $\hat{E}$ is symmetric, that is, $\hat{E}_{j,k}(x_j, x_k) = \hat{E}_{k,j}(x_k, x_j)$.

We assume that the marginal pilot estimator and proportions of the marginal exposure estimators are somehow bounded in probability:

B3. There exists a constant C such that with probability tending to 1 for all j,

$$\int \hat{\alpha}_j(x_j)^2 E_j(x_j)\mathrm{d}x_j \leq C.$$

B4. For some finite intervals $S_j \subset \mathbb{R}$ that are contained in the support of E_j, $j = 1, \ldots, d$, we suppose that there exists a finite constant C such that with probability tending to 1 for all $j \neq k$,

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\sup_{x_j \in S_j} \int \frac{\hat{E}_{j,k}(x_j, x_k)}{E_j(x_j)\hat{E}_k(x_k)^2} \mathrm{d}x_k \le C.$$

We now introduce the notation $\hat{\alpha}_j = \hat{\alpha}_j^A + \hat{\alpha}_j^B$ for the one-dimensional smoother with

$$\hat{\alpha}_j^A = \hat{E}_j(x_j)^{-1} \frac{1}{n} \sum_{i=1}^n \int k_h(x_j, X_{ij}(s)) \mathrm{d}M_i(s),$$

the variable part and

$$\hat{\alpha}_j^B = \hat{E}_j(x_j)^{-1} \frac{1}{n} \sum_{i=1}^n \int k_h(x_j, X_{ij}(s)) \mathrm{d}\Lambda_i(s),$$

the stable part of $\hat{\alpha}_j$. Here, the compensator Λ_i of N_i is defined such that M_i is a martingale and $N_i = M_i + \Lambda_i$. The definition of M_i will be given later. Now we define the stochastic and stable components of the local constant smooth backfitting estimator, $\overline{\alpha}_{0,j}^s, \overline{\alpha}_j^s$, for $s \in \{A, B\}$, as the solution of

$$\overline{\alpha}_k^s(x_k) = \hat{\alpha}_k^s(x_k) - \hat{\alpha}_{0,k}^s - \sum_{j \neq k} \int_{\mathcal{X}_j} \overline{\alpha}_j^s(x_j) \left[\frac{\hat{E}_{j,k}(x_j, x_k)}{\hat{E}_k(x_k)} - \hat{E}_{j,[k+]}(x_j) \right] \mathrm{d}x_j, \tag{A1}$$

where $\hat{\alpha}_{0,k}^s = \int \hat{\alpha}_k^s(x_k)\hat{E}_k(x_k)\mathrm{d}x_k / \int \hat{E}_k(x_k)\mathrm{d}x_k$. Existence and uniqueness of $\hat{\alpha}_k^A, \hat{\alpha}_k^B$ is stated in Proposition 1 under the following assumptions. Assumption B6 assures convergence of the variable part, whereas Assumption B7 is used to describe the structure of the bias part.

B5. There exists a constant C such that with probability tending to 1 for all j, it holds

$$\int \hat{\alpha}_j^A(x_j)^2 E_j(x_j)\mathrm{d}x_j \le C,$$
$$\int \hat{\alpha}_j^B(x_j)^2 E_j(x_j)\mathrm{d}x_j \le C.$$

B6. We assume that there is a sequence $\Delta_n \to 0$ such that

$$\sup_{x_k \in S_k} \left| \int \frac{\hat{E}_{j,k}(x_j, x_k)}{\hat{E}_k(x_k)} \hat{\alpha}_j^A(x_j)\mathrm{d}x_j \right| = o_P(\Delta_n),$$
$$\left\| \int \frac{\hat{E}_{j,k}(x_j, x_k)}{\hat{E}_k(x_k)} \hat{\alpha}_j^A(x_j)\mathrm{d}x_j \right\|_{2,k} = o_P(\Delta_n),$$

where $\| \cdot \|_{2,k}$ denotes norm defined via $\|g\|_{2,k} = \int g(u)^2 E_k(u)\mathrm{d}u$. The sets S_k have been introduced in Assumption B4.

B7. There exist deterministic functions $\mu_{n,j}$ such that

$$\sup_{x_j \in S_j} \left| \overline{\alpha}_j^B(x_j) - \mu_{n,j}(x_j) \right| = o_p(\Delta_n),$$

where S_k has been introduced in Assumption B4.

The following two propositions are results from Mammen et al. (1999), adapted to our setting and notation. Under Assumptions B1–B3 and B5, Proposition 1 ensures that the backfitting algorithm converges, and Propositions 2 and 3 give the asymptotic behavior of the backfitting estimator under Assumptions B1–B9.

Proposition 1 (Convergence of backfitting). *Under Assumptions B*1*–B*3, *with probability tending to* 1, *there exists a unique solution* $\{\overline{\alpha}_j : j = 0,\ldots,d\}$ *to* (19). *Moreover, there exist constants* $0 < \gamma < 1$ *and* $c > 0$ *such that, with probability tending to* 1, *it holds*:

$$\int \left[\overline{\alpha}_j^{[r]}(x_j) - \overline{\alpha}_j(x_j)\right]^2 E_j(x_j)\mathrm{d}x_j \leq c\gamma^{2r}\left(1 + \sum_{l=0}^{d}\int \left[\overline{\alpha}_l^{[0]}(x_l)\right]^2 E_l(x_l)\mathrm{d}x_l\right),$$

for $j = 0,\ldots,d$. *The functions* $\overline{\alpha}_l^{[0]}$ *are the starting values of the backfitting algorithm. For* $r > 0$ *the functions* $\overline{\alpha}_0^{[r]},\ldots,\overline{\alpha}_d^{[r]}$ *are defined by Equation* (20).

*Moreover, under the additional Assumption B*5, *with probability tending to* 1, *there exists a solution* $\{\overline{\alpha}_j^s : j = 0,\ldots,d\}$ *of* (A1) *that is unique for* $s = A, B$, *respectively*.

Proposition 2 (Asymptotic behavior of stochastic part). *Suppose that Assumptions B*1*–B*6 *hold for a sequence* Δ_n *and intervals* $S_j, j = 0,\ldots,d$. *Then it holds that*

$$\sup_{x_j\in S_j}\left|\overline{\alpha}_j^A(x_j) - [\hat{\alpha}_j^A(x_j) - \overline{\alpha}_{0,j}^A]\right| = o_P(\Delta_n).$$

*Under the additional Assumption B*7, *it holds*

$$\sup_{x_j\in S_j}\left|\overline{\alpha}_j^A(x_j) - [\hat{\alpha}_j^A(x_j) - \overline{\alpha}_{0,j}^A + \mu_{n,j}(x_j)]\right| = o_P(\Delta_n).$$

For the convergence of the bias term, we need the following.

B8. For all $j \neq k$, it holds

$$\sup_{x_j\in S_j}\int\left|\frac{\hat{E}_{j,k}(x_j,x_k)}{\hat{E}_j(x_j)\hat{E}_k(x_k)} - \frac{E_{j,k}(x_j,x_k)}{E_j(x_j)E_k(x_k)}\right|E_k(x_k)\mathrm{d}x_k = o_p(1).$$

At last, Assumption B9 is about the structure of the bias term of the estimators.

B9. There exist deterministic functions $a_{n,0}(x_0),\ldots,a_{n,d}(x_d)$ and constants $a_n^*, \gamma_{n,0},\ldots,\gamma_{n,d}$ and a function $\beta : \mathbb{R} \to \mathbb{R}$ (not depending on n), such that

$$\int a_{n,j}(x_j)^2E_j(x_j)\mathrm{d}x_j < \infty,$$

$$\int \beta(x)^2E(x)\mathrm{d}x < \infty,$$

$$\sup_{x_1\in S_1,\ldots,x_d\in S_d}|\beta(x)| < \infty,$$

$$\gamma_{n,j} - \int a_{n,j}(x_j)\hat{E}_j(x_j)\mathrm{d}x_j = o_P(\Delta_n),$$

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\sup_{x_j \in S_j} \left| \hat{\alpha}_j^B(x_j) - \hat{\mu}_{n,0} - \hat{\mu}_{n,j}(x_j) \right| = o_P(\Delta_n),$$

$$\int \left| \hat{\alpha}_j^B(x_j) - \hat{\mu}_{n,0} - \hat{\mu}_{n,j}(x_j) \right|^2 E_j(x_j)\mathrm{d}x_j = o_P(\Delta_n^2),$$

for random variables $\hat{\mu}_{n,0}$ and where

$$\hat{\mu}_{n,j}(x_j) = a_n^* + a_{n,j}(x_j) + \sum_{k \neq j} \int a_{n,k}(x_k) \frac{\hat{E}_{j,k}(x_j, x_k)}{\hat{E}_j(x_j)} \mathrm{d}x_k + \Delta_n \int \beta(x) \frac{E(x)}{E_j(x_j)} \mathrm{d}x_{-j}.$$

The following Proposition is taken from Mammen et al. (1999) and we have adapted it to our notation. It implies in particular that the bias term of the smooth backfitting estimators equals the projections of the bias of the full-dimensional estimator of Linton et al. (2003).

Proposition 3 (Asymptotic behavior of bias part). *Under Assumptions* $B1$–$B6$, $B8$, $B9$, *for* $j = 0, \ldots, d$, *it holds*

$$\sup_{x_j \in S_j} \left| \overline{\alpha}_j^B(x_j) - \mu_{n,j}(X_j) \right| = o_P(\Delta_n),$$

for $\mu_{n,j}(x_j) = a_{n,j}(x_j) - \gamma_{n,j} + \Delta_n \beta_j(x_j)$ *with*

$$(\beta_0, \beta_1, \ldots, \beta_d) = \arg\min_{\mathcal{B}} \int [\beta(x) - \beta_0 - \beta_1(x_1) - \cdots - \beta_d(x_d)]^2 E(x)\mathrm{d}x,$$

and $\mathcal{B} = \{\tilde{\beta} = (\beta_0, \beta_1, \ldots, \beta_d) : \int \beta_j(x_j)E_j(x_j)\mathrm{d}x_j = 0; j = 0, \ldots, d\}$. *In particular, does Assumption B7 hold with this choice of* $\mu_{n,j}(x_j)$?

With the next lemma, we ensure that the constant α^* is estimated at a parametric rate in the local constant setting. This standard result will also be needed in the proof of Theorem 1.

Lemma 1. *Let* $\overline{\alpha}^* = \left(\sum_{i=1}^n \int \mathrm{d}N_i(s)\right) / \left(\sum_{i=1}^n \int Y_i(s)\mathrm{d}s\right)$ *as defined in Equation (14). Under the condition* $\int \alpha_j(x_j)E_j(x_j)\mathrm{d}x_j = 0$, *for* $j = 0, \ldots, d$ *together with Assumption* $A2$, *it holds*

$$n^{1/2}\left(\overline{\alpha}^* - \alpha^*\right) \to \mathcal{N}\left(0, \sigma_{\alpha^*}^2\right),$$

as $n \to \infty$ *and for* $\sigma_{\alpha^*}^2 = \alpha^*(1 - \alpha^*)$. *This implies in particular* $\overline{\alpha}^* - \alpha^* = O_p(n^{-1/2})$.

Proof. We first note that it holds $E_0(t) = \int E(x)\mathrm{d}x_{-0} = \gamma(t)$ for $x = (t, z)$ and with γ from Assumption A2. Using $\frac{1}{n}\sum_{i=1}^n Y_i(s) = \gamma(s) + o_P(1)$ in the denominator and the usual martingale decomposition for counting processes in the numerator, we get

$$\mathbb{E}\left[n^{1/2}\overline{\alpha}^*\right] = n^{1/2}\alpha^* + o(1),$$
$$\mathrm{Var}\left(n^{1/2}\overline{\alpha}^*\right) = \alpha^*(1 - \alpha^*) + o(1),$$

because of the identification $\int \alpha_0(s)\gamma(s)\mathrm{d}s = 0$. The terms $\mathbb{E}\left[\int \alpha_j(Z_{i,j}(s))\gamma(s)\mathrm{d}s\right]$ in the stable part of the martingale vanish because of $\gamma(t) = \int E(x)\mathrm{d}x_{-0}$ and the identification criterion. The Central Limit Theorem for i.i.d. observations then yields the result. ∎

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Moreover, we will make use of the following counting process martingale central limit theorem, which is a direct application of Rebolledo's Theorem (Theorem II.5.1 in Andersen et al. (1993)). It is a multivariate extension of the central limit theorem for martingales in Ramlau-Hansen (1983).

Lemma 2 (Multivariate Ramlau-Hansen). *Let $\{M_i : i = 1, \dots, n\}$ be a sequence of i.i.d. martingales and let $g^{(n)}_{i,j}$ be predictable functions for $j = 1, \dots, d$. Furthermore, suppose it holds for $j, k = 1, \dots, d$,*

$$\sum_{i=1}^{n} \int g^{(n)}_{i,j}(s) g^{(n)}_{i,k}(s) \mathrm{d}\langle M_i\rangle(s) \to \sigma^2_{j,k}, \tag{A2}$$

$$\sum_{i=1}^{n} \int \left[g^{(n)}_{i,j}(s)\right]^2 I_{\{|g^{(n)}_{i,j}(s)|>\varepsilon\}} \mathrm{d}\langle M_i\rangle(s) \to 0, \tag{A3}$$

in probability for $n \to \infty$ with $\sigma^2_{j,k} > 0$ and for every $\varepsilon > 0$. Then

$$\sum_{i=1}^{n} \begin{pmatrix} \int g^{(n)}_{i,1}(s) \mathrm{d}M_i(s) \\ \vdots \\ \int g^{(n)}_{i,d}(s) \mathrm{d}M_i(s) \end{pmatrix} \to \mathcal{N}(0, \Sigma),$$

in distribution for $n \to \infty$, where $\sigma^2_{j,k}$, $j, k = 1, \dots, d$ are the entries of the covariance matrix Σ.

To show Theorem 1, we apply Propositions 1–3 and Lemmas 1 and 2. According to the propositions, it is sufficient to verify Assumptions B1–B9. In the proof of Theorem 1 we will show that our Assumptions A1–A5 imply Assumptions B1–B9 for the right choices of $\Delta_n, a_{n,j}, \beta, \gamma_{n,j}$.

Proof of Theorem 1. In the following, we show how Assumptions A1–A5 imply B1–B6, B8–B9 with our choice of marginal pilot estimators. Assumption B7 is established through Proposition 3 once the other assumptions are verified.

Without loss of generality, the proofs are done for $\mathcal{T} = R = 1$, that is, for survival time and covariates with support $[0, 1]$ and we will show that Assumptions B1–B9 are satisfied on closed subsets $S_0 \subset (0, \mathcal{T})$ and $S_j \subset (0, R)$, $j = 1, \dots, d$.

We first note that Assumption B1 follows directly from A1.

For the remaining stochastic statements, we start with the derivation of convergence rates for the marginal exposure estimators. Moreover, we will show all statements for the rate $\Delta_n = h^2$. With $I_h = [2h, 1 - 2h]$, it holds for $j = 0, \dots, d$,

$$\sup_{x_j \in I_h} |\hat{E}_j(x_j) - E_j(x_j)| = O_P\left((\log n)^{1/2} n^{-2/5}\right), \tag{A4}$$

$$\sup_{x_j, x_k \in I_h} |\hat{E}_{j,k}(x_j, x_k) - E_{j,k}(x_j, x_k)| = O_P\left((\log n)^{1/2} n^{-3/10}\right), \tag{A5}$$

$$\sup_{0 \le x_j \le 1} |\hat{E}_j(x_j) - \int_0^1 k_h(x_j, u) \mathrm{d}u\, E_j(x_j)| = O_P\left(n^{-1/5}\right), \tag{A6}$$

$$\sup_{0 \le x_j, x_k \le 1} |\hat{E}_{j,k}(x_j, x_k) - \int_0^1 k_h(x_j, u) \mathrm{d}u \int_0^1 k_h(x_k, v) \mathrm{d}v\, E_{j,k}(x_j, x_k)| = O_P\left(n^{-1/5}\right). \tag{A7}$$

Before proving Equations (A4–A7) we emphasize that they imply in particular

$$\sup_{x_j\in[0,1]} |\hat{E}_j(x_j)| = O_P(1), \tag{A8}$$

$$\sup_{x_j\in[0,1]} |\hat{E}_j(x_j)^{-1}| = O_P(1), \tag{A9}$$

$$\sup_{x_j,x_k\in[0,1]} |\hat{E}_{j,k}(x_j, x_k)| = O_P(1). \tag{A10}$$

Condition (A4) follows with standard arguments (chaining, Bernstein inequality, c.f. Mammen et al. (1999) for the regression case) from

$$\mathbb{E}[\hat{E}_j(x_j)] - E_j(x_j) = O\left(n^{-2/5}\right), \tag{A11}$$

$$|\hat{E}_j(x_j)| \leq C_1 \qquad \text{a.s.}, \tag{A12}$$

$$|\hat{E}_j(u_1) - \hat{E}_j(u_2)| \leq C_2|u_1 - u_2|n^m O_P(1), \tag{A13}$$

$$\mathrm{Var}(\hat{E}_j(x_j)) = O(n^{-4/5}), \tag{A14}$$

for constants $0 < C_1, C_2 < \infty$, $m > 0$ and all $u_1 \neq u_2, x_j \in [0, 1]$. This can be seen with Taylor expansions and using the Lipschitz continuity of K. Condition (A5–A7) can be shown in the same way. For Equations (A6) and (A7) note that $\int_0^1 k_h(x_j, u)\mathrm{d}u$ corrects the kernel at the boundaries where it does not integrate to unity.

We now show (A11–A14). Condition (A12) follows directly from A3 with K being bounded and the covariates having compact support. With usual kernel estimator arguments and a Taylor expansion of f_s around x_j we get

$$\mathbb{E}[\hat{E}_j(x_j)] - E_j(x_j) = o(h^2), \tag{A15}$$

which implies condition (A11) immediately. Condition (A14) can be derived analogously. Eventually, the Lipschitz continuity of K in A3 yields (A13).

Since the kernel k is cut off outside $[0, 1]$, Assumption B2 follows directly from Equations (A8–A10).

For the remaining assumptions we split the marginal estimator $\hat{\alpha}_j(x_j)$ as described for B5 into the variable part

$$\hat{\alpha}_j^A(x_j) = \frac{\frac{1}{n}\sum_{i=1}^n \int k_h(x_j, X_{ij}(s))\mathrm{d}M_i(s)}{\hat{E}_j(x_j)},$$

and the stable part

$$\hat{\alpha}_j^B(x_j) = \frac{\frac{1}{n}\sum_{i=1}^n \int k_h(x_j, X_{ij}(s))\mathrm{d}\Lambda_i(s)}{\hat{E}_j(x_j)},$$

via $\hat{\alpha}_j(x_j) = \hat{\alpha}_j^A(x_j) + \hat{\alpha}_j^B(x_j)$. With the choice $\Lambda_i(t) = \int_0^t \lambda_i(s)\mathrm{d}s$ for the intensity λ_i that was introduced in Equation (4), we get that $M_i = N_i - \Lambda_i$ defines a unique square integrable martingale arising from the counting process N_i.

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Next we derive the asymptotic behavior of $\hat{\alpha}_j^A(x_j)$ and $\hat{\alpha}_j^B(x_j)$ separately. With M_i being a martingale and $k_h(x_j, X_{ij}(s))$ being predictable, the integral $\int k_h(x_j, X_{ij}(s))\mathrm{d}M_i(s)$ is a martingale as well. Using the multivariate Ramlau-Hansen martingale central limit theorem in Lemma 2, we will show that $\hat{\alpha}_j^A(x_j)$ is asymptotically normally distributed whereas the difference between the stable part $\hat{\alpha}_j^B(x_j)$ and $\alpha_j(x_j)$ asymptotically behaves like the bias term $b_j(x_j)$.

For $x_j \in I_h$, we now show conditions (A2) and (A3) of Lemma 2 for $g_{ij}^{(n)}(s) = n^{-3/5}k_h(x_j - X_{ij}(s))$. Note that with Λ_i being the compensator of M_i, we get in particular $d\langle M_i\rangle(s) = d\Lambda_i(s) = \left[\alpha^* + \sum_{k=0}^{d}\alpha_k(X_{ik}(s))\right]Y_i(s)\mathrm{d}s$.

For cross-terms with $j \neq l$ in Equation (A2), it holds with this choice of $g_{ij}^{(n)}$ that

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{n}\int g_{i,j}^{(n)}(s)g_{i,k}^{(n)}(s)\mathrm{d}\langle M_i\rangle(s)\right] \\
&= \mathbb{E}\left[\left(\frac{1}{n}n^{2/5}\right)^2\sum_{i=1}^{n}\int k_h(x_j - X_{ij}(s))k_h(x_l - X_{il}(s))d\Lambda_i(s)\right] \\
&= n^{-1/5}\int\int k_h(x_j - u_j)k_h(x_l - u_l)\left[\alpha^* + \alpha_0(s) + \sum_{k=1}^{d}\alpha_k(u_k)\right] \\
&\quad \times \gamma(s)f_s(u_1,\ldots,u_d)\mathrm{d}(u_1,\ldots,u_d)\mathrm{d}s \\
&= O(h), \qquad\qquad (A16)
\end{aligned}
$$

because of the bounded support of the covariates and with the hazard rates being continuous. We write $f_s(u_1,\ldots,u_d)$ for the conditional density of $(X_{i1}(s),\ldots,X_{id}(s))$ at $(u_1,\ldots,u_d)$ given $Y_i(s) = 1$. Moreover, it can be shown easily with similar arguments that the variance of these terms satisfies

$$
\mathrm{Var}\left(\sum_{i=1}^{n}\int g_{i,j}^{(n)}(s)g_{i,k}^{(n)}(s)\mathrm{d}\langle M_i\rangle(s)\right) = O(h^6), \qquad (A17)
$$

and hence $\sigma_{k,l}^2 = 0$ for $j \neq l$ is assured for (A2). For the diagonal of the asymptotic covariance matrix $\tilde{\Sigma}$, we start with the following preliminary results. For $x_j \in I_h$ it holds

$$
\begin{aligned}
&n^{4/5}\mathbb{E}\left[n^{-2}\sum_{i=1}^{n}\int k_h(x_j - X_{ij}(s))^2\alpha_j(X_{ij}(s))Y_i(s)\mathrm{d}s\right] \\
&= n^{4/5}n^{-1}\int\int k_h(x_j - u)^2\alpha_j(u)f_s(u)\gamma(s)\mathrm{d}u\ \mathrm{d}s \\
&= n^{-1/5}h^{-1}\int\int k(v)^2\alpha_j(x_j + vh)f_s(x_j + vh)\gamma(s)\mathrm{d}v\ \mathrm{d}s \\
&= (nh^5)^{-1/5}\int k(v)^2\alpha_j(x_j)\mathrm{d}vE_j(x_j) + o(1) \\
&= c_h^{-1}\int k(v)^2\mathrm{d}v\ \alpha_j(x_j)E_j(x_j) + o(1), \qquad\qquad (A18)
\end{aligned}
$$

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

with usual kernel estimator arguments. Analogously, we get for $l \neq j$ that

$$n^{4/5}\mathbb{E}\left[n^{-2}\sum_{i=1}^{n}\int k_h(x_j - X_{ij}(s))^2\alpha_l(X_{il}(s))Y_i(s)\mathrm{d}s\right]$$
$$= c_h^{-1}\int k(v)^2\mathrm{d}v\int\int \alpha_k(u_l)f_s(x_j,u_l)\gamma(s)\mathrm{d}u_l\ \mathrm{d}s + o(1). \tag{A19}$$

For the variance of the diagonal terms, one can derive

$$\mathrm{Var}\left(\sum_{i=1}^{n}\int\left(g_{i,j}^{(n)}(s)\right)^2\mathrm{d}\langle M_i\rangle(s)\right) = O(h^5), \tag{A20}$$

which yields the stochastic convergence of diagonal variance terms together with (A18) and (A19).

Summarizing, Equations (A16–A20) imply condition (A2) of Lemma 2 with $\sigma_{j,j}^2 = \tilde{\sigma}_j^2(x_j)$ for

$$\tilde{\sigma}_j^2(x_j) = c_h^{-1}\int k^2(v)\mathrm{d}v\left(\alpha^* + \sum_{l\neq j}\int\int \alpha_k(u_l)f_s(x_j,u_l)\gamma(s)\mathrm{d}u_l\ \mathrm{d}s + \alpha_j(x_j)E_j(x_j)\right),$$

and $\sigma_{j,k}^2 = 0, j \neq k$.

The Lindeberg condition (A3) is satisfied under Assumption A3 since we assume bounded support for all covariates.

Hence, Lemma 2 implies

$$n^{2/5}\begin{pmatrix}\hat{\alpha}_0^A(x_0)\hat{E}_0(x_0)\\ \vdots \\ \hat{\alpha}_d^A(x_d)\hat{E}_d(x_d)\end{pmatrix} \to \mathcal{N}(0,\tilde{\Sigma}), \tag{A21}$$

where Σ is a diagonal matrix with the entries $\tilde{\sigma}_j^2(x_j), j = 0,\dots,d$.

Equations (A14) and (A15) imply convergence in probability of $\hat{E}_j(x_j)$ to $E_j(x_j)$ at a fast enough rate and hence, we get

$$n^{2/5}\begin{pmatrix}\hat{\alpha}_0^A(x_0)\\ \vdots \\ \hat{\alpha}_d^A(x_d)\end{pmatrix} \to \mathcal{N}(0,\Sigma), \tag{A22}$$

from Equation (A21) with Σ being a diagonal matrix with the entries $\sigma_j^2(x_j) = \tilde{\sigma}_j^2(x_j)E_j(x_j)^{-2}, j = 0,\dots,d$.

Note that condition (A22), implies in particular $\mathrm{Var}\left(\hat{\alpha}_j^A(x_j)\right) = O(n^{-4/5})$. Following the line of argumentation we used to prove (A4) for $\hat{E}_j(x_j)$, this leads to

$$\sup_{x_j\in I_h}|\hat{\alpha}_j^A(x_j)| = O_P\left((\log n)^{1/2}n^{-2/5}\right). \tag{A23}$$

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Analogously, one can get a similar result at the boundary and thus

$$\sup_{x_j \in [0,1]} |\hat{\alpha}_j^A(x_j)| = O_P(1) \tag{A24}$$

on the whole, this provides support.

For the stable part, we refer to Nielsen and Linton (1995) who have shown for

$$B_j(x_j) = \frac{1}{n}\sum_{i=1}^{n} \int k_h(x_j, X_{ij}(s))\mathrm{d}\Lambda_i(s)$$

that

$$\sup_{x_j \in [0,1]} |B_j(x_j) - \mathbb{E}[B_j(x_j)]| = o_P(1), \tag{A25}$$

$$\sup_{x_j \in [0,1]} |\mathbb{E}[B_j(x_j)]| = o(1), \tag{A26}$$

making use of the Lipschitz continuity of K from Assumption A3 and of Assumption A1. Together with (A9), Equations (A25) and (A26) imply

$$\sup_{x_j \in [0,1]} |\hat{\alpha}_j^B(x_j)| = O_P(1). \tag{A27}$$

One can get Assumptions B3 and B5 immediately from Equations (A24) and (A27). Assumptions B2, B4 and B8 follow from Equations (A4–A7).

We illustrate the derivation of Assumption B6 for $x_j \in I_h$. First note that $\int E_{j,k}(x_j, x_k)(E_j(x_j))^{-1}k_h(x_j - X_{i,j}(s))\mathrm{d}x_j$ is a bounded function $g(h, x_k, X_{i,j}(s))$ of arguments h, x_k, and $X_{i,j}(s)$ and hence predictable. This leads to

$$\mathrm{Var}\left(\int g(h, x_k, X_{i,j}(s))\mathrm{d}M_i(s)\right) = O(1),$$

due to M_i being a square integral martingale and a similar derivation to (A16–A20). Thus, it holds that

$$n^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\int\int \frac{E_{j,k}(x_j, x_k)}{E_j(x_j)} k_h(x_j, X_{i,j}(s))\mathrm{d}x_j \ \mathrm{d}M_i(s)\right)$$

is asymptotically normally distributed and in particular

$$\frac{1}{n}\sum_{i=1}^{n}\int\int \frac{E_{j,k}(x_j, x_k)}{E_j(x_j)} k_h(x_j, X_{i,j}(s))\mathrm{d}x_j \ \mathrm{d}M_i(s) = O_P\left(n^{-1/2}\right).$$

Note that by integrating over x_k, we achieve the parametric rate $n^{1/2}$ making the usual rate h^{-1} vanish. Together with (A4) and (A5), the last equation yields

$$\int \frac{\hat{E}_{j,k}(x_j, x_k)}{\hat{E}_k(x_k)}\hat{\alpha}_j^A(x_j)\mathrm{d}x_j = \int \frac{E_{j,k}(x_j, x_k)}{E_k(x_k)}\hat{\alpha}_j^A(x_j)\mathrm{d}x_j + O_P(n^{-3/10}n^{-2/5}\log n)$$

$$= E_k(x_k)^{-1}\frac{1}{n}\sum_{i=1}^{n}\int\int \frac{E_{j,k}(x_j,x_k)}{E_j(x_j)}k_h(x_j,X_{i,j}(s))\mathrm{d}x_j\ \mathrm{d}M_i(s) + O_P(n^{-3/10}n^{-2/5}\log n)$$
$$= O_P\left(n^{-1/2}\right),$$

since (A4) further implies $\hat{\alpha}_j^A(x_j) = E_j(x_j)_h^{-1}(x_j - X_{i,j}(s))\mathrm{d}M_i(s) + O_P\left(n^{-2/5}(\log n)^{1/2}\right)$.

The last equation proves Assumption B6.

We prove Assumption B9 for the following choices for $j = 0,\dots,d$.

$$a_n^* = \alpha^*,$$
$$a_{n,j}(x_j) = \alpha_j(x_j) + \alpha_j'(x_j)\int k_h(x_j,u)(u-x_j)\left[\int k_h(x_j,v)\mathrm{d}v\right]^{-1}\mathrm{d}u,$$
$$\beta(x) = \sum_{j=0}^{d}\left[\alpha_j'(x_j)\frac{\partial \log E(x)}{\partial x_j} + \frac{1}{2}\alpha_j''(x_j)\right]\int u^2k(u)\mathrm{d}u,$$
$$\gamma_{n,j} = 0.$$

The first three statements of B9 hold immediately with this choice of $a_{n,j}$ and Assumptions A1 and A3.

For the fourth statement, it holds

$$\int a_{n,j}(x_j)\hat{E}_j(x_j)\mathrm{d}x_j = \int \alpha_j(x_j)\hat{E}_j(x_j)\mathrm{d}x_j + \int \alpha_j'(x_j)\hat{E}_j(x_j)\frac{\int k_h(x_j,u)(u-x_j)}{\int k_h(x_j,v)\mathrm{d}v}\mathrm{d}x_j, \tag{A28}$$

and we investigate the two summands separately. For the first one, it holds

$$\begin{aligned}\int \alpha_j(x_j)\hat{E}_j(x_j)\mathrm{d}x_j &= \frac{1}{n}\sum_{i=1}^{n}\int\int \alpha_j(x_j)k_h(x_j,X_{ij}(s))\mathrm{d}x_jY_i(s)\mathrm{d}s\\ &= \frac{1}{n}\sum_{i=1}^{n}\int g_h(X_{i,j}(s))Y_i(s)\mathrm{d}s\\ &= \mathbb{E}\left[\int \alpha_j(x_j)\hat{E}_j(x_j)\mathrm{d}x_j\right] + o_P\left(n^{-1/2}\right)\\ &= \int\int\int \alpha_j(x_j)k_h(x_j-u)\gamma(s)f_s(u)\mathrm{d}u\ \mathrm{d}s\ \mathrm{d}x_j + o_P\left(n^{-1/2}\right)\\ &= \int\int \alpha_j(x_j)k_h(x_j-u)E_j(u)\mathrm{d}u\ \mathrm{d}x_j + o_P\left(n^{-1/2}\right)\\ &= \int \alpha_j(x_j)E_j(x_u)\mathrm{d}x_j + o_P\left(n^{-1/2}\right),\end{aligned}$$

since $\int g_h(X_{i,j}(s))Y_i(s)\mathrm{d}s$ are i.i.d. random variables with the definition $g_h(X_{i,j}(s)) = \int \alpha_j(x_j)k_h(x_j - X_{ij}(s))\mathrm{d}x_j$ and the Central Limit Theorem applies as for B6. The last equation follows from a substitution, a Taylor expansion of E_j and the fact that k is a kernel of order one.

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The second summand can be treated analogously, yielding

$$\begin{aligned}
&\int \alpha_j'(x_j)\hat{E}_j(x_j)\frac{\int k_h(x_j,u)(u-x_j)}{\int k_h(x_j,v)\mathrm{d}v}\mathrm{d}x_j \\
&= \int\int \alpha_j'(x_j)k_h(x_j-u)(u-x_j)E_j(u)\mathrm{d}u\ \mathrm{d}x_j + o_P\left(n^{-1/2}\right), \\
&= o_P\left(n^{-1/2}\right),
\end{aligned}$$

and hence in total

$$\int a_j(x_j)\hat{E}_j(x_j)\mathrm{d}x_j = o_P\left(n^{-1/2}\right). \tag{A29}$$

because of the identification $\int \alpha_j(x_j)E_j(x_u)\mathrm{d}x_j = 0$. This verifies the fourth statement of B9 with $\gamma_{n,j} = 0$.

To prove B9, we start with two preliminary results:

$$\sup_{x_j\in I_h}|\hat{\alpha}_j^B(x_j) - \hat{\mu}_{n,j}(x_j)| = o_P\left(h^2\right), \tag{A30}$$

$$\sup_{x_j\in I_h^c}|\hat{\alpha}_j^B(x_j) - \hat{\mu}_{n,j}(x_j)| = o_P(h). \tag{A31}$$

Recall that by definition, it holds

$$\begin{aligned}
\hat{\alpha}_j^B(x_j) &= \frac{1}{n}\sum_{i=1}^n \int k_h(x_j - X_{ij}(s))\mathrm{d}\Lambda_i(s)\left(\hat{E}_j(x_j)\right)^{-1} \\
&= \frac{1}{n}\sum_{i=1}^n \int k_h(x_j - X_{ij}(s))\left[\alpha^* + \sum_{l=0}^d \alpha_l(X_{il}(s))\right]Y_i(s)\mathrm{d}s\left(\hat{E}_j(x_j)\right)^{-1},
\end{aligned}$$

and

$$\begin{aligned}
\hat{\mu}_{n,j}(x_j) &= a_{n,0} + a_{n,j}(x_j) + \sum_{k\neq j}\int a_{n,k}(x_k)\frac{\hat{E}_{j,k}(x_j,x_k)}{\hat{E}_j(x_j)}\mathrm{d}x_k - \Delta_n\int \beta(x)\frac{E(x)}{E_j(x_j)}\mathrm{d}x_{-j} \\
&= \alpha^* + \alpha_j(x_j) + \alpha_j'(x_j)\int k_h(x_j,u)(u-x_j)\left[\int k_h(x_j,v)\mathrm{d}v\right]^{-1}\mathrm{d}u \\
&\quad + \sum_{k\neq j}\int\left(\alpha_k(x_k) + \alpha_k'(x_k)\int k_h(x_k,u)(u-x_k)\left[\int k_h(x_k,v)\mathrm{d}v\right]^{-1}\mathrm{d}u\right) \\
&\quad \times \frac{\hat{E}_{j,k}(x_j,x_k)}{\hat{E}_j(x_j)}\mathrm{d}x_k \\
&\quad + \Delta_n\int u^2k(u)\mathrm{d}u\int\sum_{j=0}^d\left[\alpha_j'(x_j)\frac{\partial \log E(x)}{\partial x_j} + \frac{1}{2}\alpha_j''(x_j)\right]\frac{E(x)}{E_j(x_j)}\mathrm{d}x_{-j}.
\end{aligned}$$

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Next, it holds for $j = 0, \dots, d$,

$$\frac{1}{n}\sum_{i=1}^{n}\int k_h(x_j, X_{ij}(s))\alpha_j(X_{ij}(s))Y_i(s)\mathrm{d}s\left(\hat{E}_j(x_j)\right)^{-1}$$

$$= \alpha_j(x_j) + \alpha_j'(x_j)\int k_h(x_j, u)(u - x_j)\mathrm{d}u\left(\int k_h(x_j, u)\mathrm{d}u\right)^{-1} \tag{A32}$$

$$+ h^2\int u^2k(u)\mathrm{d}u\left[E_j'(x_j)\alpha_j'(x_j) + \frac{1}{2}E_j(x_j)\alpha_j''(x_j)\right]E_j(x_j)^{-1} + R_{n,j}(x_j), \tag{A33}$$

with $\sup_{x_j\in I_h}|R_{n,j}(x_j)| = o_p(h^2)$ and $\sup_{x_j\in[0,1]\setminus I_h}|R_{n,j}(x_j)| = O_p(h^2)$. Similarly, for $k \neq j$, we get

$$\frac{1}{n}\sum_{i=1}^{n}\int k_h(x_j, X_{ij}(s))\alpha_k(X_{ik}(s))Y_i(s)\mathrm{d}s\left(\hat{E}_j(x_j)\right)^{-1}$$

$$= \int \alpha_k(x_k)\frac{\hat{E}_{j,k}(x_j, x_k)}{\hat{E}_j(x_j)}\mathrm{d}x_k$$

$$+ \int \alpha_k'(x_k)\frac{\hat{E}_{j,k}(x_j, x_k)}{\hat{E}_j(x_j)}k_h(x_k, u)(u - x_k)\mathrm{d}u\left(\int k_h(x_j, u)\mathrm{d}u\right)^{-1}$$

$$+ h^2\int u^2k(u)\mathrm{d}u\int\left[\frac{\partial E_{j,k}(x_j, x_k)}{\partial x_k}\alpha_k'(x_k) + \frac{1}{2}E_{j,k}(x_j, x_k)\alpha_j''(x_j)\right]E_j(x_j)^{-1}$$

$$+ R_{n,j,j}(x_j), \tag{A34}$$

with $\sup_{x_j\in I_h}|R_{n,j,k}(x_j)| = o_p(h^2)$ and $\sup_{x_j\in[0,1]\setminus I_h}|R_{n,j,k}(x_j)| = O_p(h^2)$. Equation (A33) follows straightforward with a Taylor expansion of each α_j and E_j and for the derivation of (A34) we refer to the proof of Theorem 4 in Mammen et al. (1999), where the analogue is shown for the nonparametric regression case. Equations (A33) and (A34) imply (A30) and (A31) with above choices of $a_{n,j}$, β and $\gamma_{n,j}$. Eventually, together with (A29), conditions (A30) and (A31) imply A9.

For the last statement of the theorem, we note that the constant component α^* in the conditional hazard can be estimated at a parametric rate $n^{-1/2}$ by $\overline{\alpha}^*$ due to Lemma 1. ∎

A2 Asymptotic theory for the local linear estimator

For the local linear estimator, we follow the same procedure as in Section A. We first introduce general assumptions as well as a set of results from Mammen et al. (1999) which we will apply to prove Theorem 2. Then we verify the new assumptions under Assumptions A1–A5.

Let $E : \mathcal{X} \to [0, 1]$ be the exposure as defined earlier and let W be a (deterministic) positive definite $(d + 1) \times (d + 1)$-matrix with elements $W_{r,s}$ such that $W_{0,0} = 1$. We set

$$M_j(x_j) = \begin{pmatrix} W_{0,0} & W_{j,0} \\ W_{j,0} & W_{j,j} \end{pmatrix} E_j(x_j), \tag{A35}$$

$$S_{l,j}(x_l, x_j) = \begin{pmatrix} W_{0,0} & W_{l,0} \\ W_{j,0} & W_{l,j} \end{pmatrix} E_{l,j}(x_l, x_j). \tag{A36}$$

These will later be the fixed but unknown matrices to which $\hat{M}_j$ and $\hat{S}_j$, respectively, converge.

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Now we make the following assumptions, which are all of similar nature to B1–B9. Note that these are assumptions on $\hat{V}^j(x_j)$, $\hat{V}^j_j(x_j)$, $\hat{V}^j_j(x_j)$, $\hat{V}^j_{j,j}(x_j)$, $\hat{V}^{l,j}(x_l,x_j)$, $\hat{V}^{l,j}_l(x_l,x_j)$, $\hat{V}^{l,j}_j(x_l,x_j)$, $\hat{V}^{l,j}_{l,j}(x_l,x_j)$ and $\hat{\alpha}_j(x_j)$, $\hat{\alpha}^j(x_j)$, and all $x_j, x_l, j, l = 0,\dots,d$ and we don't assume any particular definition of these terms for the following propositions.

B1'. For all $j \neq k$ it holds

$$\int \frac{E_{j,k}(x_j,x_k)^2}{E_j(x_j)E_k(x_k)} \mathrm{d}x_j \mathrm{d}x_k < \infty.$$

B2'. For $\hat{M}_j$ and $\hat{S}_{l,j}$ as in Equations (27) and (28) it holds

$$\int \left[\frac{\hat{V}^j(x_j) - E_j(x_j)}{E_j(x_j)}\right]^2 E_j(x_j)\mathrm{d}x_j = o_P(1),$$

$$\int \left[\frac{\hat{V}^{j,k}(x_j,x_k)}{E_j(x_j)E_k(x_k)} - \frac{E_{j,k}(x_j,x_k)}{E_j(x_j)E_k(x_k)}\right]^2 E_j(x_j)E_k(x_k)\mathrm{d}x_j\ \mathrm{d}x_k = o_P(1),$$

$$\int \left[\hat{M}_j(x_j)^{-1}\hat{S}_{k,j}(x_k,x_j) - M_j(x_j)^{-1}S_{k,j}(x_k,x_j)\right]^2_{r,s} E_j(x_j)E_k^{-1}(x_k)\mathrm{d}x_j\ \mathrm{d}x_k = o_P(1),$$

for $r,s = 1,2$. Here $[A]_{r,s}$ denotes the element (r,s) of a matrix A. Moreover, $\hat{M}_j$ vanishes outside the support of E_j, $\hat{S}_{j,k}$ vanishes outside the support of $E_{j,k}$ and $\hat{S}$ is symmetric, that is, $\hat{S}_{j,k}(x_j,x_k)^T = \hat{S}_{k,j}(x_k,x_j)$.

B3'. There exists a constant C such that with probability tending to 1 for all j,

$$\int \hat{\alpha}_j(x_j)^2 E_j(x_j)\mathrm{d}x_j \leq C,$$

and

$$\int \hat{\alpha}^j(x_j)^2 E_j(x_j)\mathrm{d}x_j \leq C.$$

B4'. For some finite intervals $S_j \subset \mathbb{R}$ that are contained in the support of E_j, $j = 0,\dots,d$, we suppose that there exists a finite constant C such that with probability tending to 1 for all $j \neq k$,

$$\sup_{x_j \in S_j} \int \operatorname{trace}\left[\hat{S}_{k,j}(x_k,x_j)\hat{M}_j(x_j)^{-2}\hat{S}_{k,j}(x_k,x_j)\right] E_k(x_k)^{-1}\mathrm{d}x_k \leq C.$$

We now introduce the notation $\hat{\alpha}_j = \hat{\alpha}_j^A + \hat{\alpha}_j^B$ and $\hat{\alpha}^j = \hat{\alpha}^{j,A} + \hat{\alpha}^{j,B}$. Where $(\hat{\alpha}_j^A, \hat{\alpha}^{j,A})$ is the variable part and $(\hat{\alpha}_j^B, \hat{\alpha}^{j,B})$ is the stable part of the initialization $(\hat{\alpha}_j, \hat{\alpha}^j)$. The terms are given by

$$\hat{\alpha}_j^A(x_j) = \left\{(\hat{V}^j_j(x_j))^2 - \hat{V}^j_{j,j}(x_j)\hat{V}^j(x_j)\right\}^{-1} \frac{1}{n}\sum_{i=1}^n \int g_{i,j}(x_j)\mathrm{d}M_i(s),$$

$$\hat{\alpha}^{j,A}(x_j) = \left\{ (\hat{V}^j_j(x_j))^2 - \hat{V}^j_{j,j}(x_j)\hat{V}^j(x_j) \right\}^{-1} \frac{1}{n}\sum_{i=1}^n \int g^j_i(x_j)\mathrm{d}M_i(s),$$

$$\hat{\alpha}^B_j(x_j) = \left\{ (\hat{V}^j_j(x_j))^2 - \hat{V}^j_{j,j}(x_j)\hat{V}^j_{0,0}(x_j) \right\}^{-1} \frac{1}{n}\sum_{i=1}^n \int g_{i,j}(x_j)\mathrm{d}\Lambda_i(s),$$

$$\hat{\alpha}^{j,B}(x_j) = \left\{ (\hat{V}^j_j(x_j))^2 - \hat{V}^j_{j,j}(x_j)\hat{V}^j_{0,0}(x_j) \right\}^{-1} \frac{1}{n}\sum_{i=1}^n \int g^j_i(x_j)\mathrm{d}\Lambda_i(s),$$

with

$$g_{i,j}(x_j) = \left[\hat{V}^j_j(x_j)\left(\frac{x_j - X_{ij}(s)}{h}\right) - \hat{V}^j_{j,j}(x_j) \right] k_h(x_j - X_{ij}(s)),$$

$$g^j_i(x_j) = \left[\hat{V}^j_j(x_j) - \hat{V}^j(x_j)\left(\frac{x_j - X_{ij}(s)}{h}\right) \right] k_h(x_j - X_{ij}(s)).$$

Equivalently, we can write

$$\begin{pmatrix} \hat{\alpha}^A_j(x_j) \\ \hat{\alpha}^{j,A}(x_j) \end{pmatrix} = \frac{1}{n}\sum_{i=1}^n \int \begin{pmatrix} 1 \\ h^{-1}(x_j - X_{ij}(s)) \end{pmatrix} k_h(x_j, X_{ij}(s))\mathrm{d}M_i(s),$$
$$\begin{pmatrix} \hat{\alpha}^B_j(x_j) \\ \hat{\alpha}^{j,B}(x_j) \end{pmatrix} = \frac{1}{n}\sum_{i=1}^n \int \hat{M}_j(x_j)^{-1} \begin{pmatrix} 1 \\ h^{-1}(x_j - X_{ij}(s)) \end{pmatrix} k_h(x_j, X_{ij}(s))\mathrm{d}\Lambda_i(s),$$

As in Assumption B4, M_i is the martingale arising from N_i, and Λ_i is its compensator. Later on, we will verify the following assumptions on $(\hat{\alpha}^A_j, \hat{\alpha}^{j,A})$ and $(\hat{\alpha}^B_j, \hat{\alpha}^{j,B})$. Moreover, for the whole estimator we define, for $s \in \{A, B\}$, $\tilde{\alpha}^s_{0,j}$, $\tilde{\alpha}^s_j$ and $\tilde{\alpha}^{j,s}$ as the solution of the equations

$$\hat{M}_j(x_j)\begin{pmatrix} \tilde{\alpha}^s_j(x_j) - \hat{\alpha}^s_j(x_j) \\ \tilde{\alpha}^{j,s}(x_j) - \hat{\alpha}^{j,s}(x_j) \end{pmatrix} = \tilde{\alpha}^s_{0,j}\begin{pmatrix} \hat{V}^j(x_j) \\ \hat{V}^j_j(x_j) \end{pmatrix} - \sum_{l \neq j} \int \hat{S}_{l,j}(x_l, x_j)\begin{pmatrix} \tilde{\alpha}^s_l(x_l) \\ \tilde{\alpha}^{l,s}(x_l) \end{pmatrix}\mathrm{d}x_l, \tag{A37}$$

$$\int \tilde{\alpha}^s_j(x_j)\hat{V}^j(x_j)\mathrm{d}x_j = 0. \tag{A38}$$

Existence and uniqueness of $\tilde{\alpha}^A_j, \tilde{\alpha}^B_j, \tilde{\alpha}^{j,A}, \tilde{\alpha}^{j,B}$ is stated in Proposition 4. We make further assumptions

B5' There exists a constant C such that with probability tending to 1 for all j, it holds

$$\int \hat{\alpha}^s_j(x_j)^2 E_j(x_j)\mathrm{d}x_j \leq C,$$

and

$$\int \hat{\alpha}^{j,s}(x_j)^2 E_j(x_j)\mathrm{d}x_j \leq C,$$

for $s = A, B$.

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

B6'. We assume that there is a sequence Δ_n such that

$$\sup_{x_k \in S_k} \left\| \int \hat{M}_k(x_k)^{-1} \hat{S}_{k,j}(x_k, x_j) \begin{pmatrix} \hat{\alpha}_j^A(x_j) \\ \hat{\alpha}^{j,A}(x_j) \end{pmatrix} \mathrm{d}x_j \right\|_2 = o_P(\Delta_n),$$

$$\left\| \int \hat{M}_k(x_k)^{-1} \hat{S}_{k,j}(x_k, x_j) \begin{pmatrix} \hat{\alpha}_j^A(x_j) \\ \hat{\alpha}^{j,A}(x_j) \end{pmatrix} \mathrm{d}x_j \right\|_{M_k,2} = o_P(\Delta_n),$$

where $\|\cdot\|_2$ denotes the L_2 norm in $\mathbb{R}^2$ and where for functions $g : \mathbb{R} \to \mathbb{R}^2$ we define $\|g\|_{M_k,2}^2 = \int g(u) M_k(u) g(u) \mathrm{d}u$. The sets S_k have been introduced in Assumption B4'.

B7'. There exist deterministic functions $\mu_{n,j}$ such that

$$\sup_{x_j \in S_j} \left| \tilde{\alpha}_j^B(x_j) - \mu_{n,j}(x_j) \right| = o_p(\Delta_n),$$

where S_k has been introduced in Assumption B4'.

The local linear equivalents to Propositions 1 and 2 are the following results from Mammen et al. (1999), adapted to our setting. The following two propositions assure convergence of the backfitting algorithm and asymptotic normality of the stochastic part of the estimator under Assumptions B1'–B7'.

Proposition 4 (Convergence of backfitting). *Under Assumptions B1'–B3', with probability tending to 1, there exists a unique solution $\{\tilde{m}_{0,l}, \tilde{m}_l, \tilde{m}^l : l = 0, \ldots, d\}$ to (26–28). Moreover, there exist constants $0 < \gamma < 1$ and $c > 0$ such that, with probability tending to 1, it holds:*

$$\int \left[\tilde{\alpha}_j^{[r]}(x_j) - \tilde{\alpha}_j(x_j) \right]^2 E_j(x_j) \mathrm{d}x_j \leq c\gamma^{2r} \Gamma,$$

$$\int \left[\tilde{\alpha}^{j,[r]}(x_j) - \tilde{\alpha}^j(x_j) \right]^2 E_j(x_j) \mathrm{d}x_j \leq c\gamma^{2r} \Gamma,$$

where

$$\Gamma = 1 + \sum_{l=0}^{d} \int \left[\tilde{\alpha}_l^{[0]}(x_l) \right]^2 E_l(x_l) \mathrm{d}x_l + \int \left[\tilde{\alpha}^{l,[0]}(x_l) \right]^2 E_l(x_l) \mathrm{d}x_l.$$

The functions $\tilde{\alpha}_{0,l}^{[0]}, \tilde{\alpha}_l^{[0]}$ and $\tilde{\alpha}^{l,[0]}$ are the starting values of the backfitting algorithm. For $r > 0$ the functions $\tilde{\alpha}_l^{[r]}$ and $\tilde{\alpha}^{l,[r]}$ are defined by Equations (29) and (30).

Moreover, under the additional Assumption B5', with probability tending to 1, there exists a solution $\{\tilde{\alpha}_0^s, \tilde{\alpha}_j^s, \tilde{\alpha}^{j,s} : j = 0, \ldots, d\}$ of (A37), (A38) that is unique for $s = A, B$, respectively.

Proposition 5 (Asymptotic behavior of stochastic part). *Suppose that Assumptions B1'–B6' hold for a sequence Δ_n and intervals $S_j, j = 0, \ldots, n$. Then it holds that*

$$\sup_{x_j \in S_j} \left| \tilde{\alpha}_j^A(x_j) - [\hat{\alpha}_j^A(x_j) - \tilde{\alpha}_{0,j}^A] \right| = o_P(\Delta_n).$$

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Under the additional Assumption B7', it holds

$$\sup_{x_j \in S_j}\left|\tilde{\alpha}_j(x_j) - [\hat{\alpha}_j^A(x_j) - \tilde{\alpha}_{0,j}^A + \mu_{n,j}(x_j)]\right| = o_P(\Delta_n).$$

Before stating a result for the bias part, we assume the following.

B8'. For all $j \neq k$, it holds

$$\sup_{x_j \in S_j} \int \left|\left[\hat{M}_j(x_j)^{-1}\hat{S}_{k,j}(s_k, x_j) - M_j^{-1}(x_j)S_{k,j}(x_k, x_j)\right]_{r,s}\right| E_k(x_k)\mathrm{d}x_k = o_p(1),$$

for $r, s = 1, 2$.

B9'. There exist deterministic functions $a_{n,0}(x_0), \ldots, a_{n,d}(x_d), a_n^0(x_0), \ldots, a_n^d(x_d)$ and constants a_n^*, $\gamma_{n,0}, \ldots, \gamma_{n,d}$ such that

$$\int a_{n,j}(x_j)^2 E_j(x_j)\mathrm{d}x_j < \infty,$$

$$\int a_n^j(x_j)^2 E_j(x_j)\mathrm{d}x_j < \infty,$$

$$\gamma_{n,j} - \int a_{n,j}(x_j)\hat{V}^j(x_j)\mathrm{d}x_j = o_P(\Delta_n),$$

$$\sup_{x_j \in S_j}\left|\tilde{\alpha}_j^B(x_j) - \hat{\mu}_{n,0} - \hat{\mu}_{n,j}(x_j)\right| = o_P(\Delta_n),$$

$$\int \left|\hat{\alpha}_j^B(x_j) - \hat{\mu}_{n,0} - \hat{\mu}_{n,j}(x_j)\right|^2 E_j(x_j)\mathrm{d}x_j = o_P(\Delta_n^2)$$

$$\sup_{x_j \in S_j}\left|\hat{\alpha}^{j,B}(x_j) - \hat{\mu}_{n,0} - \hat{\mu}_n^j(x_j)\right| = o_P(\Delta_n),$$

$$\int \left|\hat{\alpha}^{j,B}(x_j) - \hat{\mu}_n^j(x_j)\right|^2 E_j(x_j)\mathrm{d}x_j = o_P(\Delta_n^2),$$

for random variables $\hat{\mu}_{n,0}$ and where

$$\begin{pmatrix} \hat{\mu}_{n,j}(x_j) \\ \hat{\mu}_n^j(x_j) \end{pmatrix} = \begin{pmatrix} a_{n,0} + a_{n,j}(x_j) \\ a_n^j(x_j) \end{pmatrix} + \sum_{k \neq j} \int \hat{M}_j(x_j)^{-1}\hat{S}_{k,j}(x_k, x_j)\begin{pmatrix} a_{n,k}(x_k) \\ a_n^k(x_k) \end{pmatrix}\mathrm{d}x_k.$$

The next proposition appears in Mammen et al. (1999) with different notation for the nonparametric regression case. It assures convergence of the deterministic part of the estimator.

Proposition 6 (Asymptotic behavior of bias part). *Under Assumptions B1'–B6', B8', B9', it holds*

$$\sup_{x_j \in S_j}\left|\tilde{\alpha}_j^B(x_j) - \mu_{n,j}(X_j)\right| = o_P(\Delta_n),$$

$$\sup_{x_j \in S_j}\left|\tilde{\alpha}^{j,B}(x_j) - \mu_n^j(X_j)\right| = o_P(\Delta_n),$$

for $\mu_{n,j}(x_j) = a_{n,j}(x_j) - \gamma_{n,j}$ and $\mu_n^j(x_j) = a_n^j(x_j)$. Assumption B7' holds with this choice of $\mu_{n,j}(x_j)$.

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Proof of Theorem 2. To apply Propositions 4–6, we have to prove that Assumptions A1–A5 imply B1–B6, B8, B9. The proof is analogous to the proof of Theorem 1, and the assumptions can be shown in a similar way.

We now focus on the variance and bias part

$$\begin{pmatrix} \hat{\alpha}_j^A(x_j) \\ \hat{\alpha}^{j,A}(x_j) \end{pmatrix} = \hat{M}_j(x_j)^{-1}\frac{1}{n}\sum_{i=1}^{n}\int \begin{pmatrix} 1 \\ h^{-1}(x_j - X_{ij}(s)) \end{pmatrix} k_h(x_j, X_{ij}(s))\mathrm{d}M_i(s),$$
$$\begin{pmatrix} \hat{\alpha}_j^B(x_j) \\ \hat{\alpha}^{j,B}(x_j) \end{pmatrix} = \hat{M}_j(x_j)^{-1}\frac{1}{n}\sum_{i=1}^{n}\int \begin{pmatrix} 1 \\ h^{-1}(x_j - X_{ij}(s)) \end{pmatrix} k_h(x_j, X_{ij}(s))\mathrm{d}\Lambda_i(s).$$

Analogously to (A4–A7), we show uniform convergence of $\hat{M}_j(x_j)$ and $\hat{S}_{l,j}(x_l, x_j)$ to $M_j(x_j)$ and $S_{l,j}(x_l, x_j)$, respectively, and then focus on

$$\frac{1}{n}\sum_{i=1}^{n}\int \begin{pmatrix} 1 \\ h^{-1}(x_j - X_{ij}(s)) \end{pmatrix} k_h(x_j, X_{ij}(s))\mathrm{d}M_i(s)$$

for asymptotic normality and on

$$\frac{1}{n}\sum_{i=1}^{n}\int \begin{pmatrix} 1 \\ h^{-1}(x_j - X_{ij}(s)) \end{pmatrix} k_h(x_j, X_{ij}(s))\mathrm{d}\Lambda_i(s)$$

for a bias term.

With M_i being the same martingale as in the proof of Theorem 1 occurring in the stochastic part, we get the same asymptotic variance σ_j^2. Moreover, Assumptions A6–A9 can be verified with the choices

$$\begin{aligned}
\Delta_n &= h^2, \\
a_n^* &= \alpha^*, \\
a_{n,j}(x_j) &= \alpha_j(x_j) + \frac{1}{2}h^2\alpha_j''(x_j)\int u^2k(u)\mathrm{d}u, \\
a_n^j(x_j) &= h\alpha_j'(x_j), \\
\beta(x) &= \sum_{j=1}^{d}\frac{1}{2}\int u^2k(u)\mathrm{d}u\left[\alpha_j''(x_j) - \int \alpha_j''(x_j)E_j(x_j)\mathrm{d}x_j\right], \\
\gamma_{n,j} &= \nu_{n,j} + \frac{h^2}{2}\int u^2k(u)\mathrm{d}u\int \alpha_j''(x_j)E_j(x_j)\mathrm{d}x_j, \\
\nu_{n,j} &= \int\int \alpha_j(x_j)k_h(x_j, u)E_j(u)\mathrm{d}u\ \mathrm{d}x_j.
\end{aligned}$$

■

A3 Two-step smooth backfitting estimator

The interpretation as a projection motivates two different ways to compute the smooth backfitting hazard estimator. For the minimization over all additive hazard functions, we can either minimize

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

directly or we first minimize over the subspace of all (unstructured) local polynomial functions of degree p obtaining a solution $\hat{\alpha}_{pilot}$ from Equation (8) which is a non-additive estimator and then minimize the integrated squared errors between $\hat{\alpha}_{pilot}$ and all additive local polynomial functions of degree p:

$$\begin{aligned}\underset{\substack{\alpha^* \in \mathbb{R}, \\ \alpha_j^{(l)} : \mathbb{R} \to \mathbb{R}, \\ j=0,\dots,d \\ l=0,\dots,p}}{\arg\min} \sum_{i=1}^{n} \int \int \Big\{ & \hat{\alpha}_{pilot}(x) - \Big[\alpha^* + \alpha_0(t) + \alpha_1(z_1) + \cdots \alpha_d(z_d) \\ & + \alpha_0^{(p)}(x_0)\left(\frac{x_0 - X_{i0}(s)}{h}\right)^p + \cdots + \alpha_d^{(p)}(x_d)\left(\frac{x_d - X_{id}(s)}{h}\right)^p\Big]\Big\}^2 \\ & \times K_h(x - X_i(s))Y_i(s)\mathrm{d}s \, \mathrm{d}\nu(x). \end{aligned} \tag{A39}$$

We want to emphasize that the estimator we obtain via direct minimization (9) or (10), respectively, and the one obtained through the two-step minimization (A39) are identical.

In the following, we want to illustrate how the estimator can be obtained from an unstructured hazard estimator. Although we don't make use of it, this representation enables us to derive the asymptotic theory for the final estimator, making use of the known asymptotic behavior of the established unstructured local constant which is defined below. Moreover, the derivation is less technical and easier to follow, and the implementation is more straightforward.

Let $\hat{\alpha}$ be the unstructured local constant pilot estimator, $\hat{\alpha}^{LC}$ defined in Section 4.3. Then, for a weighting w, the local constant smooth backfitting estimator $\overline{\alpha}$ can be equivalently defined as

$$\min_{\overline{\alpha}} \int_{\mathcal{X}} \left(\hat{\alpha}(x) - [\overline{\alpha}^* + \sum_{j=0}^{d} \overline{\alpha}_j(x_j)] \right)^2 w(x)\mathrm{d}x.$$

Analogously, for $p = 1$ we get the local linear estimator $\hat{\alpha}^{LL}(x) = \hat{O}^{LL}(x)/\hat{E}^{LL}(x)$ for $x \in \mathcal{X}$ from Equation (8), which is defined through

$$\hat{O}^{LL}(x) = \frac{1}{n}\sum_{i=1}^{n} \int \{1 - (x - X_i(s))D(x)^{-1}c_1(x)\}K_h(x, X_i(s))\mathrm{d}N_i(s),$$

$$\hat{E}^{LL}(x) = \frac{1}{n}\sum_{i=1}^{n} \int \{1 - (x - X_i(s))D(x)^{-1}c_1(x)\}K_h(x, X_i(s))Y_i(s)\mathrm{d}s,$$

where $c_j(x) = n^{-1}\sum_{i=1}^{n} \int K_h(x, X_i(s))(x_j - X_{ij}(s))Y_i(s)\mathrm{d}s$ and for the $(d+1) \times (d+1)$-matrix $D(x) = [d_{jk}(x)]_{jk}$ with $d_{jk}(x) = \frac{1}{n}\sum_{i=1}^{n} \int K_h(x, X_i(s))(x_j - X_{ij}(s))(x_k - X_{ik}(s))Y_i(s)\mathrm{d}s$.

Note that the matrix D is not necessarily regular for $d > 2$ and hence the existence of D^{-1} and the existence of $\hat{\alpha}^{LL}$ are not guaranteed for $d > 2$.

In contrast to the local linear estimator, the local constant estimator $\hat{\alpha}^{LC}$ is always well defined, independent of the dimension d.

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

APPENDIX B

B1 Fitted values from the multiplicative model

In this section, we show the fitted values from the local constant multiplicative smooth backfitting model Hiabu, Mammen, et al. (2021) applied to the TRACE study data application from Section 6. The fit for the risk in the first three months is given in Figure B1, and the fit for the risk conditional on surviving the first three months is given in Figure B2.

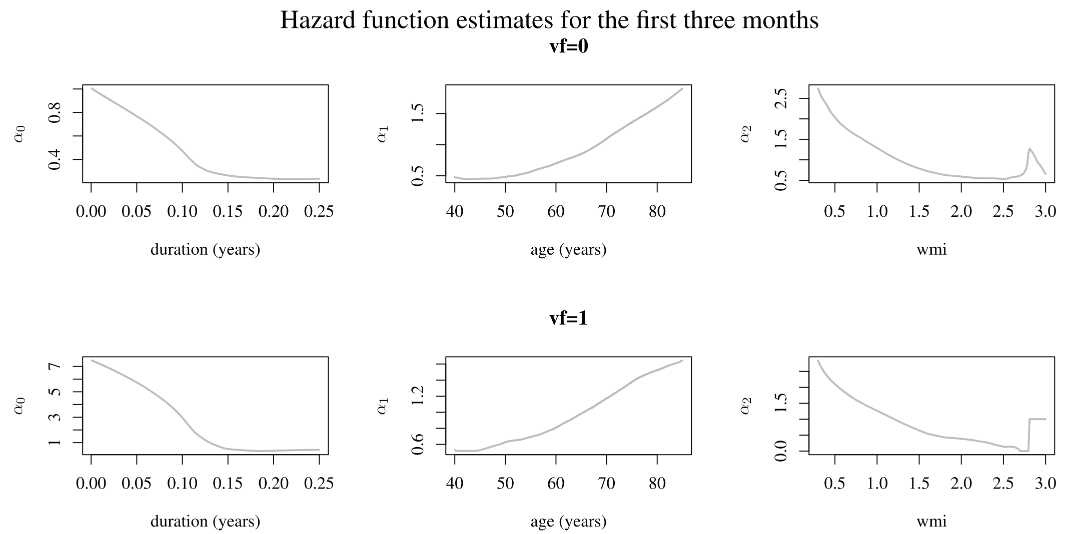

FIGURE B1 Local constant multiplicative smooth backfitting fit of $(\alpha_0, \alpha_1, \alpha_2)$ conditional on surviving the first three months for two different strata depending on the value of vf.

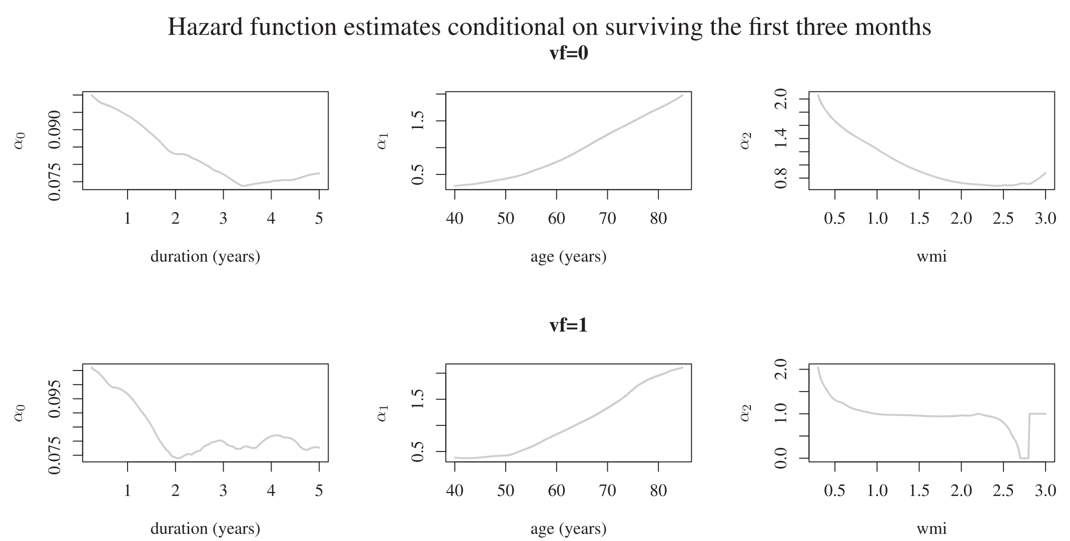

FIGURE B2 Local constant multiplicative smooth backfitting fit of $(\alpha_0, \alpha_1, \alpha_2)$ conditional on surviving the first three months for two different strata depending on the value of vf.

14679469, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.70004 by City University Of London, Wiley Online Library on [29/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License