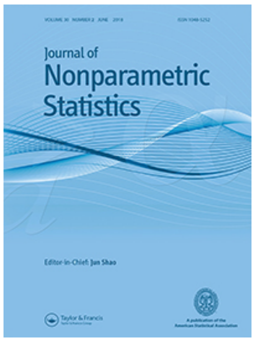

Journal of Nonparametric Statistics

ISSN: 1048-5252 (Print) 1029-0311 (Online) Journal homepage: http://www.tandfonline.com/loi/gnst20

Partially linear transformation model for length-biased and right-censored data

Wenhua Wei, Alan T. K. Wan & Yong Zhou

To cite this article: Wenhua Wei, Alan T. K. Wan & Yong Zhou (2018) Partially linear transformation model for length-biased and right-censored data, Journal of Nonparametric Statistics, 30:2, 332-367, DOI: 10.1080/10485252.2018.1424335

To link to this article: https://doi.org/10.1080/10485252.2018.1424335

JOURNAL OF NONPARAMETRIC STATISTICS, 2018
VOL. 30, NO. 2, 332–367
https://doi.org/10.1080/10485252.2018.1424335

Partially linear transformation model for length-biased and right-censored data

Wenhua Wei[a], Alan T. K. Wan[b] and Yong Zhou[a,c]

[a]School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, People's Republic of China; [b]Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong; [c]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, People's Republic of China

ABSTRACT

In this paper, we consider a partially linear transformation model for data subject to length-biasedness and right-censoring which frequently arise simultaneously in biometrics and other fields. The partially linear transformation model can account for nonlinear covariate effects in addition to linear effects on survival time, and thus reconciles a major disadvantage of the popular semiparamnetric linear transformation model. We adopt local linear fitting technique and develop an unbiased global and local estimating equations approach for the estimation of unknown covariate effects. We provide an asymptotic justification for the proposed procedure, and develop an iterative computational algorithm for its practical implementation, and a bootstrap resampling procedure for estimating the standard errors of the estimator. A simulation study shows that the proposed method performs well in finite samples, and the proposed estimator is applied to analyse the Oscar data.

ARTICLE HISTORY
Received 19 March 2017
Accepted 28 December 2017

KEYWORDS
Estimating equations; length-biased sampling; local linear fitting technique; partially linear transformation model; right-censoring

1. Introduction

Incident and prevalent cohort designs are two primary types of epidemiological study designs. An incident cohort study follows subjects that are disease-free at the time of sampling to a failure event or censoring due to loss of follow-up. While incident sampling represents an ideal form of analysis, it is often an expensive undertaking because it typically requires a large cohort with lengthy follow-up. In contrast, prevalent designs, which recruit only the living subjects diagnosed with the disease before the time of sampling, are more economical and efficient. However, by excluding subjects who died before sampling took place, prevalent cohort data are intrinsically biased towards cases of longer survival as well as being left-truncated, where the truncation time is the observed time interval between the disease onset and recruitment into the prevalent cohort. These are serious statistical problems that render standard methods of survival analysis inapplicable. In the case of a stable disease, the occurrence of disease incidence is a stationary Poisson process over time, and the truncation time has a Uniform distribution. Under this set-up, the

CONTACT Wenhua Wei wei.wenhua@mail.shufe.edu.cn

survival time in the prevalent cohort is said to have a length-biased distribution, where the probability of a survival time being sampled is proportional to its length.

Broadly speaking, there are two main methodological strategies for estimating the unbiased survival distribution from length-biased data. The majority of this work also accounts for right-censoring, which is another common feature of survival data due to loss of follow-up. The conditional approach, popularised by the work of Turnbull (1976), Lagakos, Barraj, and De Gruttola (1988), Wang (1991), and others, is conditional on the observed truncation times. There are pros and cons of this approach. On the one hand, it yields a simple and an easily implementable estimator; on the other hand, there is the drawback of efficiency loss when the Uniform distributional property of the truncation times is ignored. This shortcoming leads to the development of the alternative unconditional approach that fully utilises the aforementioned distributional information of the truncation times by maximising the full likelihood, see Vardi (1982, 1985, 1989), Gill, Vardi, and Wellner (1988), Asgharian, M'Lan, and Wolfson (2002) and Asgharian and Wolfson (2005). However, as noted by Luo and Tsai (2009), the estimator obtained by the unconditional approach has neither a closed-form expression nor an explicit limiting variance, and the method is difficult to implement. These limitations of the unconditional approach motivated Luo and Tsai (2009) to develop a pseudo-partial likelihood approach that has the advantage of simplicity of the conditional approach and yields an estimator that is only marginally inferior to that obtained under the unconditional approach.

There has also been a growth of interest in the modelling of risk factors on the unbiased failure times when the observed failure times are length-biased. Studies based on Cox's proportional hazards (PH) model and its variants include Wang (1996), Shen (2009), Tsai (2009), Qin and Shen (2010), Qin, Ning, Liu, and Shen (2011), Huang and Qin (2012), Hu, Chen, and Sun (2015), among others. When the PH model is inappropriate, the accelerated failure time (AFT) model is often a useful alternative, and several authors have considered the AFT model under length-biased sampling, see Shen, Ning, and Qin (2009), Chen (2010) and Ning, Qin, and Shen (2014a,b). One approach that has garnered considerable interests in survival analysis in recent years is the semiparametric linear transformation (SLT) model (Cheng, Wei, and Ying 1995), which is a flexible formulation that includes the PH, proportional odds (PO) and several other well-known models as special cases. The SLT model affords greater flexibility than the traditional survival models, and is evidently gaining prominence and replacing the PH model as the workhorse of survival analysis. Inferential procedures for the SLT model under various types of biased sampling schemes including length-biased sampling have been developed by Shen et al. (2009), Liu, Qin, and Shen (2012), Kim, Lu, Sit, and Ying (2013), Cheng and Huang (2014) and Wang and Wang (2015). Despite the SLT model's flexibility and many advantages, one important weakness of this model is that it constrains the effects of covariates to be linear. This is an unduly restrictive assumption adopted primarily for mathematical convenience and inappropriate in many situations. Indeed, nonlinear covariate effects are commonplace in survival analysis. Lu and Zhang (2010) cited examples from a lung cancer study (Kalbfleisch and Prentice 2002), where the survival time has a nonlinear dependence on age, and a study on women's health by New York University, where the time of developing breast carcinoma is thought to depend nonlinearly on sex hormone levels. Clearly, to dissect the potential nonlinear penetrance of the covariates, there is a need to develop a more powerful tool than the SLT model.

The partially linear transformation (PLT) model developed by Lu and Zhang (2010) (see also Ma and Kosorok 2005) is an attempt to address the aforementioned deficiency of the SLT model. The PLT model extends the SLT model by incorporating nonlinear covariate effects in the model through the inclusion of an unknown smooth function of covariates. As such, the PLT model is a generalisation of the SLT model. Several models that have been used to study the nonlinear and linear covariate effects for survival data, including the partially linear PH (PLPH) (Cai, Fan, Jiang, and Zhou 2007) and partially linear proportional odds (PLPO) models (Lu and Zhang 2010), are nested as special cases in the PLT framework. Lu and Zhang (2010) developed a martingle-based estimating equations approach to estimate the linear and nonlinear covariate effects in the PLT model and an asymptotic theory for the properties of the estimators. They also proposed an efficient iterative algorithm for implementing the procedure. There have been several interesting attempts to extend the basic PLT set-up by incorporating, for example, an additive nonparametric specification (Liu, Li, and Zhang 2014), a varying-coefficients function (Qiu and Zhou 2015), and a single index function (Liu et al. 2014) in the model. However, to the best of our knowledge, no study has considered the PLT model under length-biased sampling, and the purpose of this paper is to take steps in this direction.

In this paper, we consider the PLT model when the observed failure times are length-biased and subject to right-censoring. We adopt the same martingale-based estimation procedure of Lu and Zhang (2010) to deal with the difficulties associated with the simultaneous estimation of the transformation and covariate functions in the model. However, refinements to the procedure of Lu and Zhang (2010) are made in the following aspects. Firstly, we modify Lu and Zhang (2010)'s estimating equation to account for the two challenges encountered in the length-biased sampling data, i.e. the biasedness and informative censoring. Also, our approach fully utilises the exchangeability of the left-truncation time and the residual survival time of the length-biased data. The utilisation of this information is expected to lead to an efficiency gain in the estimator. Furthermore, we establish the asymptotic properties of the estimators by overcoming the difficulties caused by the biasedness of the data and employ a simple bootstrap scheme to obtain the estimator's variance.

We organise the rest of this paper as follows. In Section 2, we describe our model set-up and introduce the notations. Section 3 develops the estimation method and an algorithm for computing the estimates. In Section 4, we develop an asymptotic theory for the proposed estimator and a resampling method for estimating the estimator's standard deviation. Simulations results on the finite sample performance of the proposed method are reported in Section 5. A real data example illustrating the method is contained in Section 6. Section 7 concludes the paper and proofs of technical results are contained in an appendix.

2. Data and model specification

Let $\tilde{T}$ be the failure time of interest measured from the initial event to the failure event, A be the truncation time (or backward recurrence time) measured from the initial event to the time of enrolment, and V be the residual survival time (or forward recurrence time) measured from the time of enrolment to the failure event. Under length-biased sampling (Shen et al. 2009; Huang and Qin 2012; Liu et al. 2012), we only observe $T = A + V$, the length-biased version of $\tilde{T}$ within the subset of $\tilde{T} > A$. Allowing for loss of follow-up, which often occurs with clinical trial studies, V is often right-censored, and we let C be the associated censoring variable measured from the time of enrolment to censoring and assume that C is independent of A and V. Thus, the total censoring time is represented by $A + C$.

Our method for analysing $\tilde{T}$ is based on the following PLT model:

$$H(\tilde{T}) = -Z^\top \beta - f(W) + \epsilon, \tag{1}$$

where $H(\cdot)$ is an unknown monotonic increasing function of transformation, Z is a $p \times 1$ dimensional time-independent covariate, β is an unknown $p \times 1$ vector of regression coefficients, W is a scalar covariate, $f(\cdot)$ is an unspecified smooth function with $f(0) = 0$ for identifiability purpose, and ϵ is an error term with a completely specified distribution and independent of Z and W. We use $\lambda_\epsilon(t)$ and $\Lambda_\epsilon(t)$ to denote the hazard and the cumulative hazard functions of ϵ, respectively. Thus, model (1) can exploit both linear and nonlinear predictability patterns in the covariates on the unbiased lifetime $\tilde{T}$. When ϵ follows the extreme value and standard logistic distributions, model (1) reduces to the PLPH and PLPO models, respectively. When $f(\cdot) \equiv 0$, model (1) degenerates to the conventional SLT model.

The observed data set $\{(A_i, X_i, \delta_i, Z_i, W_i),\ i = 1, \ldots, n\}$ consists of n independently and identically distributed (i.i.d.) realisations from the population (A, X, δ, Z, W), where $X = \min(T, A + C)$, $T = A + V$, $\delta = I(V \leq C)$, and $I(\cdot)$ is an indicator function. Throughout our analysis, we assume that C and (A, V) are independent given the covariates Z and W. As well, we let the survival function of C be $S_C(\cdot)$, and allow it to be covariate-dependent. It is worth noting that T and $A + C$ may be dependent as they share a common component A. Actually, Asgharian and Wolfson (2005) showed that except for trivial cases, $\mathrm{Cov}(A + V, A + C) = \mathrm{Cov}(A, V) + \mathrm{Var}(A) > 0$. The data are thus informatively censored under length-biased sampling. This informative censoring feature is the major challenge in analysing length-biased and right-censored data as the methods for the conventional right-censored data may be failed to account for this. In the next section, we describe the approach for estimating β, $H(\cdot)$ and $f(\cdot)$.

3. Estimation methodology and computational algorithm

3.1. Estimating methodology

Denote the conditional density and survival functions of $\tilde{T}$ given Z and W as $f_U(t \mid Z, W)$ and $S_U(t \mid Z, W)$, respectively. Under the stationarity assumption for length-biased data, the conditional density function of T (Shen et al. 2009) is

$$f_{LB}(t \mid Z, W) = \frac{t f_U(t \mid Z, W)}{u(Z, W)},$$

where $u(Z, W) = \int_0^\infty t f_U(t \mid Z, W)\, \mathrm{d}t < \infty$ is a normalising constant. In the absence of right-censoring, Asgharian and Wolfson (2005) showed that (A, V) has an exchangeable joint density conditional on Z and W, i.e.

$$f_{A,V}(a, v \mid Z, W) = \frac{f_U(a + v \mid Z, W)}{u(Z, W)}.$$

Thus, A and V share the same marginal conditional density function

$$f_A(A = t \mid Z, W) = f_V(V = t \mid Z, W) = \frac{S_U(t \mid Z, W)}{u(Z, W)},$$

and the conditional density functions of A given V,Z and W and that of V given A,Z and W have the same form, with the former density function given by

$$f_{A|V}(A = a \mid V = v, Z, W) = \frac{f_U(a + v \mid Z, W)}{S_U(v \mid Z, W)}, \tag{2}$$

and the latter function being defined analogously.

However, in the presence of right-censoring, A and V no longer have an exchangeable joint density function because A is always observed, whereas V is right-censored. Consider the bivariate variable $(A, \tilde{V})$ of the uncensored observations, where $\tilde{V} = \min(V, C)$ is the observed residual lifetime. From Huang and Qin (2012), the conditional density function of $A = a$ given $\delta = 1, \tilde{V} = v, Z$ and W is given by

$$P(A = a \mid \delta = 1, \tilde{V} = v, Z, W) = \frac{f_U(a + v \mid Z, W)}{S_U(v \mid Z, W)}. \tag{3}$$

Comparing the density functions (2) and (3), we can infer that given $\delta = 1$, the conditional density of A given $\tilde{V} = v$ is the same as the conditional density function of V given A in the prevalent cohort that is not subject to right-censoring. Cheng and Huang (2014) showed that in the case of the SLT model under length-biased sampling and right-censoring, the utilisation of this exchangeability information can improve estimation efficiency. Cheng and Huang (2014)'s method is based on a combined unbiased estimating equations approach taking into account the aforementioned exchangeability between the conditional density functions of A and V. Our approach, to be described below, generalises Cheng and Huang (2014)'s method from the SLT model to the PLT model.

Let us define $N_i(t) = I(X_i \le t)\delta_i$, $Y_i^1(t) = I(A_i \le t \le X_i)$, $Y_i^2(t) = \delta_i I(\tilde{V}_i \le t \le X_i)$, $Y_i(t) = \frac{1}{2}\{Y_i^1(t) + Y_i^2(t)\}$, and $M_i(t) = N_i(t) - \int_0^t Y_i(u)\, \mathrm{d}\Lambda_\epsilon(H_0(u) + Z_i^\top \beta_0 + f_0(W_i))$, where $\tilde{V}_i = X_i - A_i$, $i = 1, \ldots, n$, and $H_0(\cdot)$, β_0 and $f_0(\cdot)$ are the true values of $H(\cdot)$, β and $f(\cdot)$, respectively. Using results from Cheng and Huang (2014), it can be readily shown that $M_i(t)$ is a mean zero process, and if $f(\cdot)$ is known, it degenerates to the case of SLT model considered by Cheng and Huang (2014). The mean zero property of $M_i(t)$ allows us to construct the following global estimating equations for β and $H(\cdot)$ with fixed $f(\cdot)$:

$$\sum_{i=1}^{n} \left\{\mathrm{d}N_i(t) - Y_i(t)\, \mathrm{d}\Lambda_\epsilon(H(t) + Z_i^\top \beta + f(W_i))\right\} = 0 \tag{4}$$

and

$$\sum_{i=1}^{n} \int_0^\tau Z_i \left\{\mathrm{d}N_i(t) - Y_i(t)\, \mathrm{d}\Lambda_\epsilon(H(t) + Z_i^\top \beta + f(W_i))\right\} = 0, \tag{5}$$

where $\tau = \inf\{t : \Pr(X > t) = 0\}$, $H(\cdot)$ is a nondecreasing function that satisfies $H(0) = -\infty$, and has positive jumps only at the points corresponding to K uncensored observations $0 < t_1 < \cdots < t_K < \infty$. In practice, we can substitute τ by t_K. Moreover, it is instructive to note that Equation (4) is a difference equation for the estimation of the transformation function $H(\cdot)$ when β and $f(\cdot)$ are fixed, and Equation (5) is for the purpose of identifying β with fixed $H(\cdot)$ and $f(\cdot)$.

We use the local linear fitting technique to estimate the smooth nonparametric function $f(\cdot)$. The smoothness assumption of $f(\cdot)$ enables us to apply the Taylor series expansion, and write, for any u in the neighbourhood of w,

$$f(u) \approx \alpha_0(w) + \alpha_1(w)(u - w), \tag{6}$$

where $\alpha_0(w) = f(w), \alpha_1(w) = \dot{f}(w)$ and $\dot{f}(w)$ is the first-order derivative of $f(w)$. We call Equation (6) the local model. Let $K(\cdot)$ be a kernel function and $K_h(t) = K(t/h)/h$, where $h > 0$ is the bandwidth parameter. Then for any fixed β and $H(\cdot)$, by substituting $f(\cdot)$ by Equation (6), the kernel-weighted local estimating equation for $\alpha_0(\cdot)$ and $\alpha_1(\cdot)$ can be constructed as follows:

$$\sum_{i=1}^{n} \int_0^{\tau} \begin{pmatrix} 1 \\ W_i - w \end{pmatrix} K_h(W_i - w)\{\mathrm{d}N_i(t) - Y_i(t)\,\mathrm{d}\Lambda_\epsilon(H(t) + Z_i^\top \beta + \alpha_0(w) + \alpha_1(w)(W_i - w))\} = 0. \tag{7}$$

The introduction of the kernel function $K(\cdot)$ in Equation (7) reflects the fact that the local model (6) is only valid for the data near w. The estimators of $\beta, H(\cdot)$ and $f(\cdot)$ are solutions to the estimating equations (4), (5) and (7).

Remark 3.1: We use the kernel-based local linear fitting technique to estimate $f(\cdot)$ only for the purpose of simplicity. Other fitting techniques and smoothers, such as local polynomial regression (Fan and Gijbels 1996, Chapter 2) or spline smoothers (Schumaker 2007), may also be used.

3.2. Computational algorithm

It is clear from the preceding discussion that solutions to the estimating equations (4), (5) and (7) can only be obtained iteratively. To this end, we propose the following iterative algorithm along the lines of Carroll, Fan, Gijbels, and Wand (1997), Cai et al. (2007), Cai, Fan, Jiang, and Zhou (2008) and Lu and Zhang (2010):

Step 0: Choose an initial value for $f(\cdot)$ and denote it as $\tilde{f}^{(0)}(\cdot)$. Following Carroll et al. (1997), Cai et al. (2007, 2008) and Lu and Zhang (2010), we use the naive one-step estimator as the initial value. We prove that the naive one-step estimator is locally consistent in the appendix. Fix $f(\cdot)$ at this initial value, we then solve Equations (4) and (5) for $H(\cdot)$ and β using Chen, Jin, and Ying (2002)'s algorithm for the SLT model. We denote the estimators as $\tilde{H}(\cdot)$ and $\tilde{\beta}$.

Step 1: Based on $\tilde{H}(\cdot)$ and $\tilde{\beta}$, we solve Equation (7) to obtain the estimators $\tilde{\alpha}_0(W_i)$ and $\tilde{\alpha}_1(W_i)$ of $\alpha_0(w)$ and $\alpha_1(w)$, respectively, at the observed points $w = W_i$, $i = 1, \ldots, n$. This leads to the estimators $\tilde{f}(W_i) = \tilde{\alpha}_0(W_i)$, $i = 1, \ldots, n$.

Step 2: Update the estimators of β and $H(\cdot)$ by solving the estimating equations (4) and (5) again, with $f(W_i)$ replaced by $\tilde{f}(W_i)$, $i = 1, \ldots, n$.

Step 3: Repeat Steps 1 and 2 alternately until the estimators of β and $H(\cdot)$ converge. We denote the final estimators of β and $H(\cdot)$ as $\hat{\beta}$ and $\hat{H}(\cdot)$, respectively.

Step 4: Substituting $\hat{\beta}$ and $\hat{H}(\cdot)$ for β and $H(\cdot)$, we solve Equation (7) to obtain the estimators $\hat{\alpha}_0(w, h, \hat{\beta}, \hat{H})$ and $\hat{\alpha}_1(w, h, \hat{\beta}, \hat{H})$ of $\alpha_0(w)$ and $\alpha_1(w)$, respectively, at the selected grid points $w = w_i,\ i = 1, \ldots, s$. The algorithm ends after completing this step and the final estimators of β, $H(\cdot)$ and $f(\cdot)$ are $\hat{\beta}$, $\hat{H}(\cdot)$ and $\hat{f}(w) = \hat{\alpha}_0(w, h, \hat{\beta}, \hat{H})$, respectively.

Remark 3.2: We use the following convergence criterion based on l_2-norm for Step 3 of the algorithm:

$$\Delta^{(m)} = \left\{ \sum_{j=1}^{p} (\tilde{\beta}_j^{(m)} - \tilde{\beta}_j^{(m-1)})^2 + \sum_{j=1}^{K} (\tilde{H}^{(m)}(t_j) - \tilde{H}^{(m-1)}(t_j))^2 \right\}^{1/2},$$

where $\tilde{\beta}^{(m)}$ and $\tilde{H}^{(m)}(\cdot)$ are the estimates of β and $H(\cdot)$ at the mth iteration. The algorithm exits Step 3 when $\Delta^{(m)}$ is less than a prescribed threshold value.

Remark 3.3: The choice of an appropriate bandwidth parameter h is required for the successful implementation of the algorithm. It is worthwhile to note that h plays different roles for different steps of the algorithm. For Steps 1–3, h should be chosen to be optimal for the estimation of β and $H(\cdot)$. For Step 4, an appropriate h should be selected in order for $\hat{f}(\cdot)$ to attain the desired optimal property. Due to the nonuniform purpose of h, we select two values of h, one for Steps 1–3, and the other for Step 4. Our choice of h for Step 4 is the optimal bandwidth $\hat{h}_{opt} = C_0 n^{-1/5}$ (see Theorem 4.3 and Section 4.1), where C_0 can be estimated by a range of methods such as the rule-of-thumb, cross-validation, or the approaches of Carroll et al. (1997) and Cai et al. (2007, 2008). In our simulation and real data analysis, we will use a simple data-adaptive criterion. See Section 5 for details. The bandwidth used for Steps 1–3 is the *ad-hoc* bandwidth (Carroll et al. 1997; Cai et al. 2007, 2008): $\hat{h}_{ad-hoc} = \hat{h}_{opt} \times n^{1/5} \times n^{-1/3} = \hat{h}_{opt} \times n^{-2/15}$.

4. Asymptotic properties and estimation of asymptotic variance

4.1. Asymptotic properties of the proposed estimator

In this section, we establish the asymptotic properties of the proposed estimators $\hat{\beta}, \hat{H}(\cdot), \hat{\alpha}_0(w)$ and $\hat{\alpha}_1(w)$. First, we define the following quantities for any s and $t \in (0, \tau]$:

$$B_1(t) = E[Y(t)\dot{\lambda}_\epsilon(H_0(t) + Z^\top \beta_0 + f_0(W))],$$

$$B_2(t) = E[Y(t)\lambda_\epsilon(H_0(t) + Z^\top \beta_0 + f_0(W))],$$

$$B(t, s) = \exp\left\{ \int_s^t \frac{B_1(u)}{B_2(u)} \, \mathrm{d}H_0(u) \right\},$$

$$B_1^Z(t) = E[ZY(t)\dot{\lambda}_\epsilon(H_0(t) + Z^\top \beta_0 + f_0(W))],$$

$$B_2^Z(t) = E[ZY(t)\lambda_\epsilon(H_0(t) + Z^\top \beta_0 + f_0(W))],$$

$$z(t) = \frac{1}{B_2(t)} \left\{ B_2^Z(t) + \int_t^\tau \left[B_1^Z(s) - \frac{B_2^Z(s)B_1(s)}{B_2(s)} \right] B(t,s)\,\mathrm{d}H_0(s) \right\} \quad \text{and}$$

$$\lambda^*\{H_0(t)\} = B(t,0), \quad \Lambda^*(x) = \int_{-\infty}^x \lambda^*(u)\,\mathrm{d}u \quad \text{for } x \in (-\infty, \infty),$$

where $\dot{\lambda}_\epsilon(t)$ is the first-order derivative of $\lambda_\epsilon(t)$.

As well, define

$$A_1 = \int_0^\tau E[\{Z - z(t)\}Z^\top Y(t)\dot{\lambda}_\epsilon(H_0(t) + Z^\top\beta_0 + f_0(W))]\,\mathrm{d}H_0(t),$$

$$A_2 = \int_0^\tau E\left\{ [Z - m_Z(t)]Y(t)\frac{e_1^\top(W)}{e_{31}(W)}\dot{\lambda}_\epsilon(H_0(t) + Z^\top\beta_0 + f_0(W)) \right\} dH_0(t),$$

$$\Sigma^* = E\left\{ \int_0^\tau \{[Z - m_Z(t)] - [Z^* - m_{Z^*}]\}\,\mathrm{d}M(t) \right\}^{\otimes 2} \quad \text{and} \quad A = A_1 - A_2,$$

where $M(t) = N(t) - \int_0^t Y(u)\,\mathrm{d}\Lambda_\epsilon(H_0(u) + Z^\top\beta_0 + f_0(W))$ and $a^{\otimes 2} = aa^\top$ for any vector a. Moreover, for $i = 1, \ldots, n$, we have

$$Z_i^* = \frac{\int_0^\tau E[ZY(t)\dot{\lambda}_\epsilon\{H_0(t) + Z^\top\beta_0 + f_0(W)\} \mid W = W_i]\,\mathrm{d}H_0(t)}{\int_0^\tau E[Y(t)\dot{\lambda}_\epsilon\{H_0(t) + Z^\top\beta_0 + f_0(W)\} \mid W = W_i]\,\mathrm{d}H_0(t)} \quad \text{and}$$

$$m_{Z_i^*} = \frac{\int_0^\tau m_Z(t)E[Y(t)\dot{\lambda}_\epsilon\{H_0(t) + Z^\top\beta_0 + f_0(W)\} \mid W = W_i]\,\mathrm{d}H_0(t)}{\int_0^\tau E[Y(t)\dot{\lambda}_\epsilon\{H_0(t) + Z^\top\beta_0 + f_0(W)\} \mid W = W_i]\,\mathrm{d}H_0(t)},$$

where $m_Z(t) = q(t)(\lambda^*\{H_0(t)\}/B_2(t))$ and $q(t)$ is the solution to the following integral equation:

$$q(t) - \int_0^\tau q(s)D_1(s,t)\,\mathrm{d}H_0(s) = \frac{B_2(t)z(t)}{\lambda^*\{H_0(t)\}} - c_3(t), \quad t \in [0,\tau], \tag{8}$$

where the definitions of $e_1(\cdot), e_{31}(\cdot), D_1(\cdot,\cdot)$ and $c_3(\cdot)$ are given in the proof of Theorem 4.1 in the appendix.

We now summarise the asymptotic properties, including the consistency and asymptotic normality, of the proposed estimators $\hat{\beta}, \hat{H}(\cdot), \hat{\alpha}_0(w)$ and $\hat{\alpha}_1(w)$ in the following theorems, the proof of which are given in the appendix.

Theorem 4.1 (Asymptotic Properties of $\hat{\beta}$): *Assume that conditions* (C1)–(C7) *in the appendix are satisfied. If $nh^2/\log(1/h) \to \infty$ and $nh^4 \to 0$, and given $\hat{\beta}$ in a small neighbourhood of β_0, then as $n \to \infty$, we have*

$$\hat{\beta} \xrightarrow{\mathcal{P}} \beta_0,$$

where $\xrightarrow{\mathcal{P}}$ denotes convergence in probability. In addition, as $n \to \infty$, we have

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} N(0, \Sigma),$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and $\Sigma = A^{-1}\Sigma^*(A^{-1})^\top$.

Theorem 4.2 (Asymptotic Properties of $\hat{H}(\cdot)$): *Assume that conditions* (C1)–(C7) *in the appendix are satisfied. If* $nh^2/\log(1/h) \to \infty$ *and* $nh^4 \to 0$, *then as* $n \to \infty$, *we have*

$$\sqrt{n}(\hat{H}(t) - H_0(t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\kappa_i(t)}{\lambda^*\{H_0(t)\}} + o_p(1)$$

for $t \in [0, \tau]$, *where* $\kappa_i(t)$, $i = 1, \ldots, n$, *are independent mean zero functions* (*see Equation* (A23) *in the appendix for definitions of these functions*). *Thus,* $\sqrt{n}(\hat{H}(t) - H_0(t))$ *converges weakly to a mean zero Gaussian process.*

Theorem 4.3 (Asymptotic Properties of $\hat{\alpha}_0(w)$ and $\hat{\alpha}_1(w)$): *Assume that conditions* (C1)–(C7) *in the appendix are satisfied. If* nh^5 *is bounded, and* β *and* $H(\cdot)$ *are estimated at the order* $O_p(n^{-\frac{1}{2}})$, *then as* $n \to \infty$, *we have*

$$\sqrt{nh}\left(\begin{pmatrix} \hat{\alpha}_0(w) - f_0(w) \\ h(\hat{\alpha}_1(w) - \dot{f}_0(w)) \end{pmatrix} - b_n(w)\right) \xrightarrow{\mathcal{D}} N(0, D(w)),$$

where $D(w) = \Sigma_1^{-1}(w)\Sigma_2(w)\Sigma_1^{-1}(w)$, *and* $\Sigma_1(w)$ *and* $\Sigma_2(w)$ *are defined in the appendix.*

4.2. Estimation for the asymptotic variance of $\hat{\beta}$

We have established the asymptotic normality of the estimator $\hat{\beta}$ and derived an expression for its asymptotic variance $\Sigma = A^{-1}\Sigma^*(A^{-1})^\top$ in Theorem 4.1. Unfortunately, the derived expression of Σ involves solving integral equations and cannot be easily computed. For the computation of the variance of $\hat{\beta}$, instead of using the derived formula of Σ, we propose to implement the resampling procedure developed by Gross and Lai (1996). The procedure is described as follows.

Let Φ_n be the empirical distribution that assigns probability $1/n$ on each of the original observations $(A_i, X_i, \delta_i, Z_i, W_i)$, $i = 1, \ldots, n$. A simple bootstrap sample can be obtained by generating n i.i.d. observations $(A_i^*, X_i^*, \delta_i^*, Z_i^*, W_i^*)$, $i = 1, \ldots, n$, from the distribution Φ_n. Based on this bootstrapped sample, estimating equations analogous to Equations (4), (5) and (7) with $(A_i, X_i, \delta_i, Z_i, W_i)$ replaced by $(A_i^*, X_i^*, \delta_i^*, Z_i^*, W_i^*)$, $i = 1, \ldots, n$, can be constructed. The iterative algorithm of Section 3.2 is then applied to solve these estimating equations and obtain the estimators β^*, $H^*(\cdot)$ and $f^*(\cdot)$ of β, $H(\cdot)$ and $f(\cdot)$, respectively. By repeating this procedure B times, a sequence of β_i^*'s, $i = 1, \ldots, B$, is obtained. Gross and Lai (1996) established an asymptotic theory of this simple bootstrap method and showed that the simple bootstrap approximations to the sampling distributions of various non-parametric statistics from left-truncated and right-censored data are accurate to the order of $O_p(n^{-1})$. Thus, the asymptotic variance of the estimator $\hat{\beta}$ can be approximated by the empirical sample variances of β_i^*, $i = 1, \ldots, B$.

5. Simulation results

The purpose of this section is to conduct a simulation exercise to assess the finite sample performance of the proposed method. We generate the unbiased data $\tilde{T}$ from the PLT model (1), where the hazard function of ϵ is $\lambda_\epsilon(t) = \exp(t)/(1 + r * \exp(t))$ with $r = 0,1$

(Dabrowska and Doksum 1988). Note that the model in Equation (1) degenerates to the PLPH and PLPO models when $r = 0$ and $r = 1$, respectively. We consider two independent covariates Z_1 and Z_2, generated from the $N(0, 1)$ and Bernoulli(0.5) distributions, respectively, to enter the linear component of the model, and let the true value β_0 be $(1, -1)^\top$. Additionally, we let the nonparametric function be $f(w) = 2w - w^2$, where $W \sim U(0, 2)$ and is independent of Z_1 and Z_2, and the transformation functions be $H(t) = 2\log(t)$ and $H(t) = \log(\exp(t) - 1)$ when $r = 0$ and $r = 1$, respectively.

For the generation of T, the length-biased data, we use the sampling procedure described in Shen et al. (2009). This process involves generating the truncation variable A from the Uniform distribution $U(0, \tau_A)$ independently of $\tilde{T}$, the unbiased data, where τ_A is a constant that exceeds the upper bound of $\tilde{T}$. The latter constraint on τ_A is imposed to guarantee the stationary of the length-biased data. In our experiment, we set $\tau_A = 100$. We then select

Table 1. Simulation results for β.

		$\beta_1 = 1$				$\beta_2 = -1$			
Censoring mechanism	CR(%)	Bias	SE	SD	CP(95%)	Bias	SE	SD	CP(95%)
		PLPH case							
Covariate-independent	20%	0.0771	0.1637	0.1745	95.7	−0.0659	0.2451	0.2621	98.0
Censoring	40%	0.0881	0.2002	0.2094	97.0	−0.0807	0.2847	0.3105	97.0
Covariate-dependent	20%	0.0721	0.1672	0.1808	96.3	−0.0707	0.2592	0.2614	95.7
Censoring	40%	0.0818	0.2300	0.2397	93.7	−0.0928	0.2959	0.3202	95.7
		PLPO case							
Covariate-independent	20%	−0.0192	0.2927	0.3327	97.3	−0.0216	0.6123	0.6239	96.7
Censoring	40%	−0.0048	0.3293	0.3575	97.7	0.0340	0.5964	0.6440	97.7
Covariate-dependent	20%	−0.0059	0.3444	0.3513	95.0	0.0010	0.6051	0.6205	97.0
Censoring	40%	−0.0185	0.3990	0.4429	97.3	0.0638	0.6344	0.6971	96.7

Table 2. Simulation results for $f(\cdot)$.

				PLPH case				PLPO case			
Censoring mechanism	CR(%)	w_0	$f(w_0)$	$\hat{f}(w_0)$	Bias	SE	SD	$\hat{f}(w_0)$	Bias	SE	SD
Covariate-independent Censoring	20%	0.3	0.5100	0.4827	−0.0273	0.1589	0.1712	0.5084	−0.0016	0.3435	0.4352
		0.6	0.8400	0.8034	−0.0366	0.1244	0.1360	0.8098	−0.0302	0.3402	0.4159
		1.2	0.9600	0.9448	−0.0152	0.1300	0.1387	0.9105	−0.0495	0.3894	0.4343
		1.5	0.7500	0.7221	−0.0279	0.1207	0.1372	0.7190	−0.0310	0.3552	0.3938
		1.8	0.3600	0.3509	−0.0091	0.2213	0.2215	0.3962	0.0362	0.4726	0.5467
	40%	0.3	0.5100	0.4781	−0.0319	0.1759	0.2134	0.5257	0.0157	0.3940	0.4679
		0.6	0.8400	0.8144	−0.0256	0.1449	0.1617	0.8338	−0.0062	0.3747	0.4225
		1.2	0.9600	0.9215	−0.0385	0.1543	0.1672	0.9166	−0.0434	0.3874	0.4471
		1.5	0.7500	0.7015	−0.0485	0.1474	0.1692	0.7053	−0.0447	0.3621	0.4258
		1.8	0.3600	0.3152	−0.0448	0.2424	0.2689	0.2879	−0.0721	0.5130	0.5941
Covariate-dependent Censoring	20%	0.3	0.5100	0.4887	−0.0213	0.1543	0.1743	0.4994	−0.0106	0.3938	0.4494
		0.6	0.8400	0.8082	−0.0318	0.1275	0.1415	0.8205	−0.0195	0.3788	0.4126
		1.2	0.9600	0.9268	−0.0332	0.1288	0.1440	0.9530	−0.0070	0.3910	0.4348
		1.5	0.7500	0.7244	−0.0256	0.1207	0.1398	0.7521	0.0021	0.3648	0.3989
		1.8	0.3600	0.3651	0.0051	0.1909	0.2249	0.3850	0.0250	0.4881	0.5439
	40%	0.3	0.5100	0.4798	−0.0302	0.1833	0.2214	0.4668	−0.0432	0.4354	0.5159
		0.6	0.8400	0.8005	−0.0395	0.1416	0.1699	0.7970	−0.0430	0.3845	0.4688
		1.2	0.9600	0.9217	−0.0383	0.1435	0.1742	0.9431	−0.0169	0.4260	0.4860
		1.5	0.7500	0.7036	−0.0464	0.1453	0.1754	0.7032	−0.0468	0.3768	0.4546
		1.8	0.3600	0.3459	−0.0141	0.2665	0.2752	0.3252	−0.0348	0.5703	0.6316

$n = 100$ pairs of $(A, \tilde{T})$ that satisfy $A < \tilde{T}$ from the sample, and this subset of $\tilde{T}$ constitutes our length-biased data T.

In addition, we consider covariate-independent as well as covariate-dependent cases of right-censoring. For the independent censoring case, we generate the residual censoring variable C from the $U(0, c_1)$ distribution, while for the dependent case, we generate C from $-2Z_1 - Z_2 + EXP(c_2)$, where c_1 and c_2 are chosen such that the censoring percentages (CR) are 20% or 40%. Notably, the total censoring time equals $A + C$. The number of replications is set to 300, and the number of bootstraps associated with the aforementioned described resampling procedure is set to 50. We use the Gaussian kernel in all cases, and

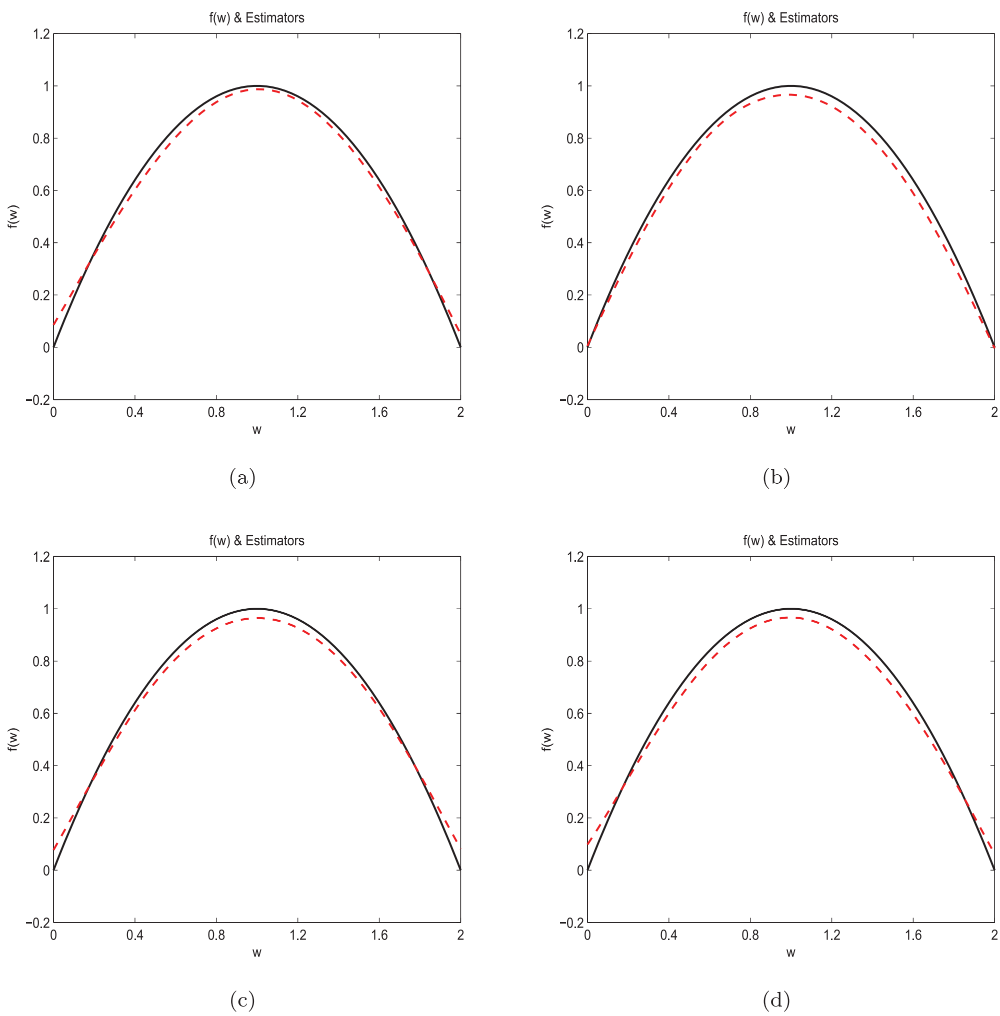

Figure 1. The true and estimated curves of $f(w)$ under the PLPH case. Sub-figure (a) is for the covariate-independent censoring case with a censoring rate of 20%, (b) is for the covariate-independent censoring case with a censoring rate of 40%, (c) is for the covariate-dependent censoring case with a censoring rate of 20% and (d) is for the covariate-dependent censoring case with a censoring rate of 40%. In each of the four sub-figures, the black solid curve is the true curve of $f(w)$, and the red dashed curve is the estimated curve of $f(w)$ based on the proposed method.

choose $h_1 = 0.6n^{-1/3}$ for estimating β and $H(\cdot)$, and $h_2 = 0.6n^{-1/5}$ for estimating $f(w)$ under all scenarios.

The simulation results are presented in Tables 1 and 2. Table 1 summarises the performance of the estimator of β based on bias magnitude (BIAS), standard errors of estimates (SE), estimated standard deviations (SD) and coverage probabilities (CP) corresponding to the nominal 95% confidence interval. The SE is calculated as the standard deviation of the estimates from the replicated samples, and the SD is obtained from the bootstrap resampling procedure. Table 2 reports the performance of the estimator of the nonlinear function $f(w)$ at the fixed points $w = 0.3, 0.6, 1.2, 1.5, 1.8$. At each of these points, we present the average estimate of $f(w)$ across the replications as well as the true value of $f(w)$. The results are presented for CR = 20% and 40%, and for $r = 0$ and $r = 1$ corresponding to the cases of the PLPH and PLPO models respectively. Table 1 shows that when estimating β, the proposed

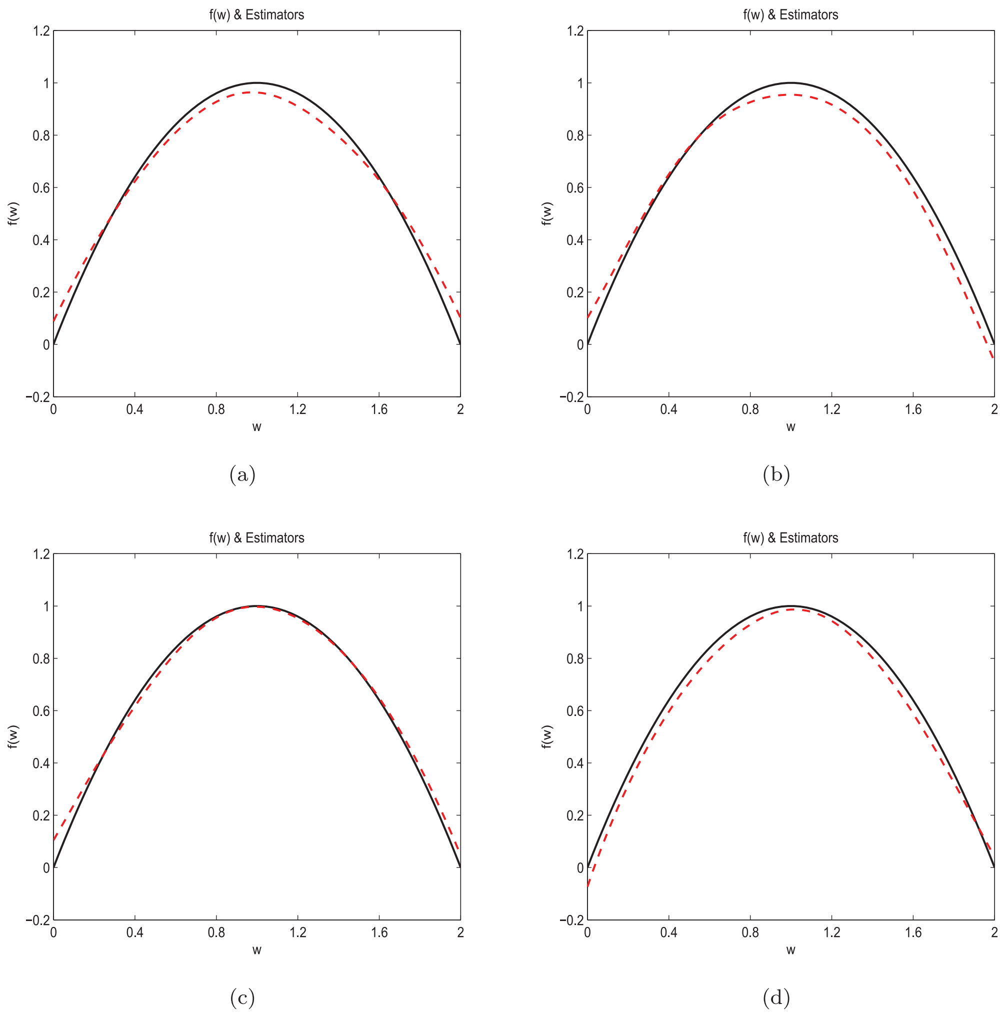

Figure 2. The true and estimated curves of $f(w)$ under the PLPO case. The notations are the same as in Figure 1.

estimator leads to biases of negligible magnitude, and SEs and SDs that are very close to each other, indicating that the resampling procedure works well. In addition, the CPs are all very close to the nominal 95% level. A change from covariate-independent to covariate-dependent censoring has no significant impact on the results, but an increase in CR from 20% to 40% generally has the effect of worsening the performance of the estimator. As far as the estimation of $f(w)$ is concerned, the above comments concerning the bias, SE and SD of the estimator also apply in broad terms. Specifically, irrespective of the values of r and *CR*, the biases are never very large; there are evidently larger deviations between SE and SD compared to the case when one is estimating β, but the differences between the two measures are still not very substantial. Indeed, we plot the the true and estimated curves of $f(w)$ side by side for the cases of $r=0$ and $r=1$ in Figures 1 and 2, respectively. It can be seen that $f(w)$ and $\hat{f}(w)$ nearly coincide everywhere in w. All of the above comments apply to the case of the PLPH model ($r=0$) as much as to the PLPO model ($r=1$). In all cases considered, the iterative computational algorithm converges within a few iterations.

6. A real data example

Social status is often considered to be a determinant of life expectancy. Higher social status is commonly believed to be correlated with lower mortality. To study the relationship between social status and life expectancy, Redelmeier and Singh (2001) and Sylvestre, Huszti, and Hanley (2006) considered the Oscar data set. They found that actors and actresses who had won Oscar awards tended to live longer than those who had not. In this section, we apply the proposed method to the same data set.

The Oscar data set contains information on a number of professional and personal characteristics of 1670 actors and actresses from the first Academy Award to March 2001. Thus, observations corresponding to those who were alive in March 2001 are subject to right-censoring. Among the 1670 actors and actresses included in the data set, 902 were never nominated for an Oscar, 529 received at least one nomination but never won any award, and the remaining 239 were nominated and won at least one Oscar.

In our analysis, the central question is the influence of winning an Oscar on the nominee's life expectancy. Thus, we exclude the 902 actors and actress who did not receive any Oscar nomination and focus only on the 768 Oscar nominees. We further exclude 5 observations with the wrong record (ID 908, 1075, 1192, 1430 and 1521) from the observations. These result in a data set containing 763 observations with a censoring rate of 57.14%. The same data set has been used in the studies of Wolkewitz, Allignol, Schumacher, and Beyersmann (2010), Chen, Wan, and Zhou (2014), and Lin and Zhou (2014).

These 763 Oscar nominees were included in the data set after their first Oscar nomination, and the nominees were alive at the time. Thus, the data are left-truncated and right-censored with the age of the nominee at the first nomination as the left-truncation variable. Chen et al. (2014) applied the test of Addona and Wolfson (2006) to the same data and confirmed that they satisfy the stationarity assumption, therefore, it is reasonable to regard this data set as length-biased and right-censored. Let $\tilde{T}$ be the nominee's lifetime, A be the nominee's age at the first nomination, and C be the time from the nominee's first nomination to death or the end of the study, whichever occurred first. The following characteristics of the nominee are used as covariates in the linear part of the model: gender ($1=$male, $0=$female) (Z_1), country of birth ($1=$ born in the U.S., $0=$ born elsewhere)

(Z_2), ethnicity ($1 =$ white, $0 =$ others) (Z_3), name change ($1 =$ has changed name, $0 =$ has never changed name) (Z_4), the number of four star films acted (Z_5), and whether a winner of Oscar ($1 =$ has won at least one Oscar, $0 =$ has never won any Oscar award) (Z_6). We scale the number of films in which the nominee has starred before the end of study to lie between 0 and 1 and use the resultant scaled variable as the nonlinear factor W. Therefore, our model is

$$H(\tilde{T}) = \sum_{i=1}^{6} \beta_i Z_i + f(W) + \epsilon,$$

where ϵ follows the extreme value distribution under the PLPH model and the standard logistic distribution under the PLPO model, and the transformation function $H(\cdot)$ takes the corresponding form as in simulations. We also use the Gaussian kernel under both models and set the bandwidth to $h = 0.6n^{-1/3}$ for estimating of $H(\cdot)$ and β, and $h = 0.6n^{-1/5}$ for estimating $f(\cdot)$. We report the estimated coefficients (EST), estimated standard deviations (SD) and estimated 95% confidence intervals (CI) for the regression coefficients $\beta_i, i = 1, \ldots, 6$ in Table 3, where SD is calculated based on 50 iterations of the

Table 3. Estimation Results for the Oscar nomination data.

	β_1	β_2	β_3	β_4	β_5	β_6
			PLPH case			
EST	−0.5864	−0.2293	0.0606	−0.0560	0.0351	0.1348
SD	0.1288	0.1141	0.1511	0.1210	0.0123	0.1075
CI(95%)	(−0.839, −0.334)	(−0.453, −0.006)	(−0.236, 0.357)	(−0.293, 0.181)	(0.011, 0.059)	(−0.076, 0.346)
			PLPO case			
EST	−1.0571	−0.4168	0.2234	−0.1445	0.0567	0.2197
SD	0.1316	0.1071	0.2147	0.1333	0.0118	0.1240
CI(95%)	(−1.315, −0.799)	(−0.627, −0.207)	(−0.197, 0.644)	(−0.406, 0.117)	(0.034, 0.080)	(−0.023, 0.463)

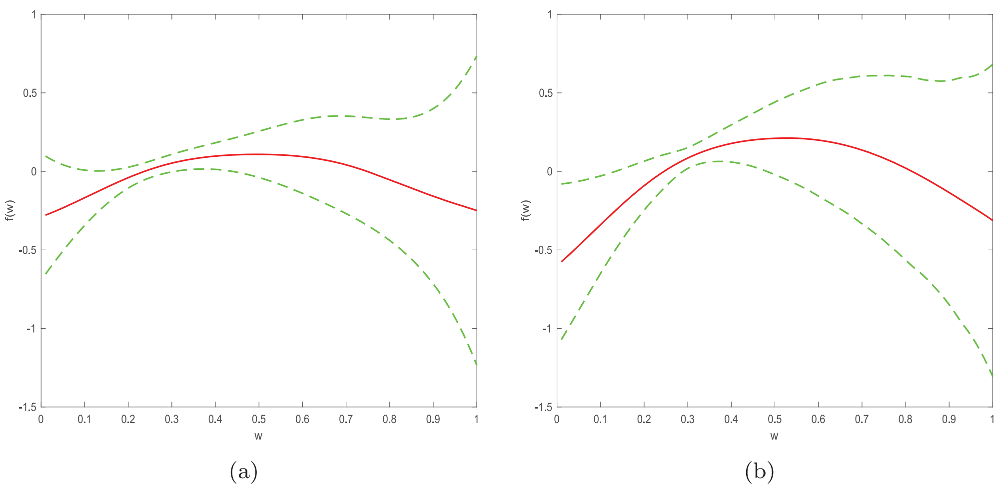

Figure 3. The estimated curve of $f(w)$ and its corresponding 95% pointwise confidence intervals based on the Oscar data set. Sub-figure (a) is for the PLPH case and (b) is for the PLPO case. In each sub-figure, the red solid curve is the estimated nonparametric function, and the green dashed curves represent the 95% pointwise confidence intervals.

bootstrap resampling procedure used in the simulation study. Furthermore, we plot the estimated curve $f(W)$ and its corresponding 95% pointwise confidence intervals based on the two models in Figure 3.

The results in Table 3 show that based on the PLPH model, gender, country of birth and name change are negatively related to a nominee's life expectancy; on the other hand, ethnicity, the number of four star films acted and having been an Oscar winner have positive impacts on life expectancy. However, only the coefficients of gender, country of birth and the number of four star films acted are significantly different from zero, as their corresponding 95% confidence intervals do not contain 0. We therefore conclude that there is no obvious difference in life expectancy between Oscar winners and nominees who never won the award. This conclusion concurs with those of Sylvestre et al. (2006) and Chen et al. (2014). Table 3 shows that the PLPO and PLPH models yield very similar results. Figure 3 shows that the covariate W indeed has a nonlinear effect on life expectancy, and the estimated nonparametric functions of $f(w)$ based on the PLPO and PLPH models exhibit substantial similarities.

7. Concluding remarks

One important advantage of the partially linear transformation model lies in its ability to capture both linear and nonlinear effects of the covariates on the dependent variable. We have considered this model and proposed estimators for the unknown covariate effects when the data are subject to length-biasedness and right-censoring. We have shown that the proposed estimators possess optimal asymptotic properties and fare well in finite samples. The partially linear transformation model may be extended to the following partially linear transformation varying-coefficients model (Qiu and Zhou 2015):

$$H(T) = -Z^T\beta - f^T(W)V + \epsilon, \tag{9}$$

where V is a $q \times 1$ dimensional covariate, $f(\cdot)$ is an unspecified $q \times 1$ dimensional smooth vector function, and other quantities are defined as in Section 2. When $V \equiv 1$, model (9) degenerates to model (1) directly. Model (9) can accommodate interaction effects between covariates and and the dynamic effects of the covariates on the dependent variable through the varying coefficients. Work in progress by the authors considers this model in the context of length-biased and right-censored data.

Moreover, as one of the referees commented, we may also develop a methodology based on a full likelihood approach for estimating β and $H(\cdot)$, in order to pursuit more efficient estimators, even though the actual computation of the estimates will likely be very cumbersome. Ma and Kosorok (2005) considered the current status data under the same model framework as ours and proposed a penalised log-likelihood estimation method. They showed that their proposed estimator of β is asymptotically efficient, and thus a similar method may be developed for the length-biased and right-censored data case. However, this is beyond the scope of this paper and will be an interesting point of departure for future research.

Acknowledgments

We thank the editor, the associate editor and two referees for helpful comments on a earlier draft of this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Wei's work was supported by a research fund from the Shanghai University of Finance and Economics (No. 2017110070). Wan's work was supported by a General Research Fund (No. 9042086) from the Hong Kong Research Grants Council and a strategic grant from the City University of Hong Kong (No. 7004786). Zhou's work was supported by the State Key Program in the Major Research Plan of National Natural Science Foundation of China (No. 91546202), the State Key Program of National Natural Science Foundation of China (No. 71331006), and Innovative Research Team of Shanghai University of Finance and Economics (No. IRTSHUFE13122402).

References

Addona, V., and Wolfson, D.B. (2006), 'A Formal Test for Stationarity of the Incidence Rate Using Data from a Prevalent Cohort Study with Follow-Up', *Lifetime Data Analysis*, 12, 267–284.

Asgharian, M., M'Lan, C.E., and Wolfson, D.B. (2002), 'Length-Biased Sampling with Right Censoring: An Unconditional Approach', *Journal of the American Statistical Association*, 97, 201–209.

Asgharian, M., and Wolfson, D.B. (2005), 'Asymptotic Behavior of the Unconditional NPMLE of the Length-Biased Survivor Function from Right Censored Prevalent Cohort Data', *The Annals of Statistics*, 33, 2109–2131.

Cai, J.W., Fan, J.Q., Jiang, J.C., and Zhou, H.B. (2007), 'Partially Linear Hazard Regression for Multivariate Survival Data', *Journal of the American Statistical Association*, 102, 538–551.

Cai, J.W., Fan, J.Q., Jiang, J.C., and Zhou, H.B. (2008), 'Partially Linear Hazard Regression with Varying Coefficients for Multivariate Survival Data', *Journal of the Royal Statistical Society, Series B*, 70, 141–158.

Carroll, R.J., Fan, J.Q., Gijbels, I., and Wand, M.P. (1997), 'Generalized Partially Linear Single-Index Models', *Journal of the American Statistical Association*, 92, 477–489.

Chen, Y.Q. (2010), 'Semiparametric Regression in Size-Biased Sampling', *Biometrics*, 66, 149–158.

Chen, K.N., Jin, Z.Z., and Ying, Z.L. (2002), 'Semiparametric Analysis of Transformation Models with Censored Data', *Biometrika*, 89, 659–668.

Chen, X.R., Wan, A.T.K., and Zhou, Y. (2014), 'A Quantile Varying-Coefficient Regression Approach to Length-Biased Data Modelling', *Electronic Journal of Statistics*, 8, 2514–2540.

Cheng, Y.J., and Huang, C.Y. (2014), 'Combined Estimating Equation Approaches for Semiparametric Transformation Models with Length-Biased Survival Data', *Biometrics*, 70, 608–618.

Cheng, S.C., Wei, L.J., and Ying, Z. (1995), 'Analysis of Transformation Models with Censored Data', *Biometrika*, 82, 835–845.

Dabrowska, D.M., and Doksum, K.A. (1988), 'Estimation and Testing in A Two-Sample Generalized Odds-Rate Model', *Journal of the American Statistical Association*, 83, 744–749.

Fan, J.Q., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman & Hall.

Gill, R.D., Vardi, Y., and Wellner, J.A. (1988), 'Large-Sample Theory of Empirical Distributions in Biased Sampling Models', *The Annals of Statistics*, 16, 1069–1112.

Gross, S.T., and Lai, T.L. (1996), 'Bootstrap Methods for Truncated and Censored Data', *Statistica Sinica*, 6, 509–530.

Hu, N., Chen, X.R., and Sun, J.G. (2015), 'Regression Analysis for Length-Biased and Right-Censored Failure Time Data with Missing Covariates', *Scandinavian Journal of Statistics*, 42, 438–452.

Huang, C.Y., and Qin, J. (2012), 'Composite Partial Likelihood Estimation under Length-Biased Sampling, with Application to a Prevalent Cohort Study of Dementia', *Journal of the American Statistical Association*, 107, 946–957.

Kalbfleisch, J.D., and Prentice, R.L. (2002), *The Statistical Analysis of Failure Time Data* (2nd ed.), Hoboken, NJ: John Wiley & Sons.
Kim, J.P., Lu, W.B., Sit, T., and Ying, Z.L. (2013), 'A Unified Approach to Semiparametric Transformation Models under General Biased Sampling Schemes', *Journal of the American Statistical Association*, 108, 217–227.
Lagakos, S.W., Barraj, L.M., and De Gruttola, V. (1988), 'Nonparametric Analysis of Truncated Survival Data, with Application to AIDS', *Biometrika*, 75, 515–523.
Lin, C.J., and Zhou, Y. (2014), 'Analyzing Right-Censored and Length-Biased Data with Varying-Coefficient Transformation Model', *Journal of Multivariate Analysis*, 130, 45–63.
Liu, L., Li, J.B., and Zhang, R.Q. (2014), 'General Partially Linear Additive Transformation Model with Right-Censored Data', *Journal of Applied Statistics*, 41, 2257–2269.
Liu, H., Qin, J., and Shen, Y. (2012), 'Imputation for Semiparametric Transformation Models with Biased-Sampling Data', *Lifetime Data Analysis*, 18, 470–503.
Lu, W.B., and Zhang, H.H. (2010), 'On Estimation of Partially Linear Transformation Models', *Journal of the American Statistical Association*, 105, 683–691.
Luo, X.D., and Tsai, W.Y. (2009), 'Nonparametric Estimation for Right-Censored Length-Biased Data: A Pseudo-Partial Likelihood Approach', *Biometrika*, 96, 873–886.
Ma, S.G., and Kosorok, M.R. (2005), 'Penalized Log-Likelihood Estimation for Partly Linear Transformation Models with Current Status Data', *The Annals of Statistics*, 33, 2256–2290.
Ning, J., Qin, J., and Shen, Y. (2014a), 'Semiparametric Accelerated Failure Time Model for Length-Biased Data with Application to Dementia Study', *Statistica Sinica*, 24, 313–333.
Ning, J., Qin, J., and Shen, Y (2014b), 'Score Estimating Equations from Embedded Likelihood Functions under Accelerated Failure Time Model', *Journal of the American Statistical Association*, 109, 1625–1635.
Pollard, D. (1990), *Empirical Processes: Theory and Applications. NSF-CBMS Regional Conference Series in Probability and Statistics (Vol. 2)*, Hayward, CA: IMS.
Qin, J., Ning, J., Liu, H., and Shen, Y. (2011), 'Maximum Likelihood Estimations and EM Algorithms with Length-Biased Data', *Journal of the American Statistical Association*, 106, 1434–1449.
Qin, J., and Shen, Y. (2010), 'Statistical Methods for Analyzing Right-Censored Length-Biased Data under Cox Model', *Biometrics*, 66, 382–392.
Qiu, Z.P., and Zhou, Y. (2015), 'Partially Linear Transformation Models with Varying Coefficients for Multivariate Failure Time Data', *Journal of Multivariate Analysis*, 142, 144–166.
Redelmeier, D.A., and Singh, S.M. (2001), 'Survival in Academy Award–Winning Actors and Actresses', *Annals of Internal Medicine*, 134, 955–962.
Reinhard, H. (1986), *Differential Equations: Foundations and Applications*, London: North Oxford.
Schumaker, L.L. (2007), *Spline Functions: Basic Theory* (3rd ed.), New York: Cambridge University Press.
Shen, P.S. (2009), 'Hazards Regression for Length-Biased and Right-Censored Data', *Statistics and Probability Letters*, 79, 457–465.
Shen, Y., Ning, J., and Qin, J. (2009), 'Analyzing Length-Biased Data with Semiparametric Transformation and Accelerated Failure Time Models', *Journal of the American Statistical Association*, 104, 1192–1202.
Sylvestre, M.P., Huszti, E., and Hanley, J.A. (2006), 'Do Oscar Winners Live Longer Than Less Successful Peers? A Reanalysis of the Evidence', *Annals of Internal Medicine*, 145, 361–363.
Tsai, W.Y. (2009), 'Pseudo-Partial Likelihood for Proportional Hazards Models with Biased-Sampling Data', *Biometrika*, 96, 601–615.
Turnbull, B.U. (1976), 'The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data', *Journal of the Royal Statistical Society, Series B*, 38, 290–295.
van der Vaart, A.W., and Wellner, J.A. (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York: Springer.
Vardi, Y. (1982), 'Nonparametric Estimation in the Presence of Length Bias', *The Annals of Statistics*, 10, 616–620.
Vardi, Y. (1985), 'Empirical Distributions in Selection Bias Models', *The Annals of Statistics*, 13, 178–203.

Vardi, Y. (1989), 'Multiplicative Censoring, Renewal Processes, Deconvolution and Decreasing Density: Nonparametric Estimation', *Biometrika*, 76, 751–761.
Wang, M.C. (1991), 'Nonparametric Estimation from Cross-Sectional Survival Data', *Journal of the American Statistical Association*, 86, 343–354.
Wang, M.C. (1996), 'Hazards Regression Analysis for Length-Biased Data', *Biometrika*, 83, 343–354.
Wang, X., and Wang, Q.H. (2015), 'Estimation for Semiparametric Transformation Models with Length-Biased Sampling', *Journal of Statistical Planning and Inference*, 156, 80–89.
Wolkewitz, M., Allignol, A., Schumacher, M., and Beyersmann, J. (2010), 'Two Pitfalls in Survival Analyses of Time-Dependent Exposure: A Case Study in A Cohort of Oscar Nominees', *The American Statistician*, 64, 205–211.

Appendix 1. Assumptions for the asymptotic theories

This appendix provides the proofs of the main results given in Section 4. Let $\|a\|$ denote the Euclidean norm for a vector a and $\|f\|$ the supremum norm for a function f, i.e. $\|f\| = \sup_{t\in[0,\tau]} |f(t)|$. Our proofs of the theorem require the following conditions:

(C1) The unique true parameter β_0 belongs to the interior of the compact parameter space $\mathcal{B} \in \mathbb{R}^p$.
(C2) The covariate Z is a $p \times 1$ dimensional bounded vector not contained in a $(p-1)$ dimensional hyperplane. The covariate W has a compact support $\mathcal{W}$ and the density function $g(\cdot)$ of W has a bounded second derivative.
(C3) τ is finite with $P(T > \tau) > 0$ and $P(A + C > \tau) > 0$.
(C4) $\lambda_\epsilon(t)$ is positive, bounded and continuously differentiable on $(-\infty, m)$ for any finite constant m, and $\lim_{t\to-\infty} \lambda_\epsilon(t) = 0$.
(C5) $H_0(t)$ has a continuous and positive derivative $\dot{H}_0(t)$ on $[0, \tau]$, and f_0 has a continuous second derivative.
(C6) $D_1(\cdot,\cdot)$ in the integral equation (8) satisfies $\sup_{t\in[0,\tau]} \int_0^\tau \mid D_1(s,t) \mid dH_0(s) < \infty$.
(C7) A and Σ^* are finite and nonsingular matrices.

Appendix 2. Proofs of the theorems

We first propose a naive one-step estimator of the unknown parameters similar to Carroll et al. (1997), Cai et al. (2007, 2008) and Lu and Zhang (2010), which can be used as the initial value of the iterative algorithm described in Section 3.2. In addition, we prove that the naive one-step estimator is locally consistent. Specifically, for any fixed $w \in \mathcal{W}$, the naive estimators of $H(\cdot)$, β and $\alpha_1(w)$ are obtained by solving the following estimating equations:

$$\sum_{i=1}^n K_h(W_i - w)\{\mathrm{d}N_i(t) - Y_i(t)\,\mathrm{d}\Lambda_\epsilon(H(t) + Z_i^\top\beta + \alpha_1(w)(W_i - w))\} = 0, t \geq 0 \tag{A1}$$

and

$$\sum_{i=1}^n \int_0^\tau \begin{pmatrix} Z_i \\ W_i - w \end{pmatrix} K_h(W_i - w)\{\mathrm{d}N_i(t) - Y_i(t)\,\mathrm{d}\Lambda_\epsilon(H(t) + Z_i^\top\beta + \alpha_1(w)(W_i - w))\} = 0. \tag{A2}$$

The estimating equations (A1) and (A2) can be solved by applying the algorithm of Chen et al. (2002). It is worth noting that the intercept term $\alpha_0(w)$ that appears in Equation (7) is included in the function $H(t)$. Denote the resultant estimators from above estimating equations as $\breve{H}(t)$, $\breve{\beta}$ and $\breve{\alpha}_1(w)$ respectively, and $f(w)$ can be estimated by $\breve{f}(w) = \int_0^w \breve{\alpha}_1(u)\,\mathrm{d}u$. Under some regularity conditions, we can show that $\breve{\beta}$, $\breve{\alpha}_1(w)$ and $\breve{f}(w)$ are locally consistent. This is summarised in Lemma A.1 as follows. For the implementation of the algorithm in Section 3.2, the initial values of β and $f(\cdot)$ are set to $\tilde{\beta}^{(0)} = \breve{\beta}$ and $\tilde{f}^{(0)}(w) = \breve{f}(w)$, respectively.

Lemma A.1 (Local consistency of the naive one-step estimator): *Under the regularity conditions* (C1)–(C5), *if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, A_w (given in Equation* (A3)) *is finite and nonsingular for any $w \in \mathcal{W}$, then $\breve{\beta}, \breve{\alpha}_1(w), \breve{f}(w)$ are locally consistent.*

Proof: This proof uses a similar approach to that of Chen et al. (2002). For any given β and $\alpha_1(w)$, the l.h.s. of Equation (A1) is monotone with respect to $H(\cdot)$. Let $\breve{H}(t;\beta,\alpha_1(w))$ be the nondecreasing function uniquely determined by Equation (A1), and $\breve{H}(t;\beta,\alpha_1(w))$ exists when β is in a small neighbourhood of β_0 and $\alpha_1(w)$ is bounded for $w \in \mathcal{W}$. Mimicking Step 1 of the proof of Kim et al. (2013), we can show that $\breve{H}(t;\beta_0,\dot{f}_0(w))$ converges almost surely to $H_0(t)+f_0(w)$ on $[0,\tau]$.

For any $w \in \mathcal{W}$, we define the conditional version of the items given in Section 4 for any $s,t \in [0,\tau]$ as follows:

$$B_{1w}(t) = E[Y(t)\dot{\lambda}_\epsilon(H_0(t)+Z^T\beta_0+f_0(W)) \mid W=w],$$

$$B_{2w}(t) = E[Y(t)\lambda_\epsilon(H_0(t)+Z^T\beta_0+f_0(W)) \mid W=w],$$

$$B_w(t,s) = \exp\left\{\int_s^t \frac{B_{1w}(u)}{B_{2w}(u)}\,\mathrm{d}H_0(u)\right\},$$

$$B^Z_{1w}(t) = E[ZY(t)\dot{\lambda}_\epsilon(H_0(t)+Z^T\beta_0+f_0(W)) \mid W=w],$$

$$B^Z_{2w}(t) = E[ZY(t)\lambda_\epsilon(H_0(t)+Z^T\beta_0+f_0(W)) \mid W=w],$$

$$z_w(t) = \frac{1}{B_{2w}(t)}\left\{B^Z_{2w}(t) + \int_t^\tau \left[B^Z_{1w}(s) - \frac{B^Z_{2w}(s)B_{1w}(s)}{B_{2w}(s)}\right] B_w(t,s)\,\mathrm{d}H_0(s)\right\},$$

$$\lambda^*_w\{H_0(t)\} = B_w(t,0), \quad \Lambda^*_w(x) = \int_{-\infty}^x \lambda^*_w(u)\,\mathrm{d}u \quad \text{for } x \in (-\infty,\infty).$$

Replacing $H(t)$ by $\breve{H}(t;\beta,\alpha_1(w))$ in (A1) and taking the derivative with respect to β on both sides of the resultant equation, for any $t \in [0,\tau]$, we can obtain

$$\left.\frac{\partial \breve{H}(t;\beta,\alpha_1)}{\partial\beta}\right|_{\beta=\beta_0,\alpha_1=\dot{f}_0} = -\int_0^t \frac{B_w(s,t)}{B_{2w}(s)} B^Z_{1w}(s)\,\mathrm{d}H_0(s) + o_p(1)$$

and

$$\mathrm{d}\left.\frac{\partial \breve{H}(t;\beta,\alpha_1)}{\partial\beta}\right|_{\beta=\beta_0,\alpha_1=\dot{f}_0} = -\frac{1}{B_{2w}(t)}\left\{B^Z_{1w}(t) + B_{1w}(t)\left.\frac{\partial \breve{H}(t;\beta,\alpha_1)}{\partial\beta}\right|_{\beta=\beta_0,\alpha_1=\dot{f}_0} + o_p(1)\right\}\mathrm{d}H_0(t).$$

Similarly, taking the derivative with respect to $\alpha_1(w)$ on both sides of the resultant equation, we have

$$\left.\frac{\partial \breve{H}(t;\beta,\alpha_1(w))}{\partial\alpha_1(w)}\right|_{\beta=\beta_0,\alpha_1=\dot{f}_0} = 0.$$

The above calculations imply that for t in a compact subset of the interior of the support of X, the derivative of $\breve{H}(t;\beta,\alpha_1(w))$ with respect to β is bounded in the neighbourhood of β_0, and the derivative of $\breve{H}(t;\beta,\alpha_1(w))$ with respect to $\alpha_1(w)$ is 0 in the neighbourhood of $\dot{f}_0$. Because $\breve{H}(t;\beta_0,\dot{f}_0(w))$ converges uniformly to $H_0(t)+f_0(w)$ on $[0,\tau]$, we obtain that $\breve{H}(t;\breve{\beta},\breve{\alpha}_1(w))$ converges uniformly to $H_0(t)+f_0(w)$ on $[0,\tau]$, provided that $\breve{\beta} \to \beta_0$ and $\breve{\alpha}_1(w)$ is bounded.

We replace $H(t)$ by $\breve{H}(t;\beta,\alpha_1(w))$ in Equation (A2) and denote

$$\breve{U}_w(\beta,\alpha_1(w)) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \begin{pmatrix} Z_i \\ W_i - w \end{pmatrix} K_h(W_i-w)\{\mathrm{d}N_i(t) - Y_i(t)\,\mathrm{d}\Lambda_\epsilon(\breve{H}(t;\beta,\alpha_1(w)) + Z_i^\top\beta + \alpha_1(w)(W_i-w))\}.$$

Similar to Step 4 of Chen et al. (2002), using the law of large numbers and standard nonparametric techniques, we can show that $\breve{U}_w(\beta, \alpha_1(w))$ converges almost surely to a deterministic vector $\breve{u}_w(\beta, \alpha_1(w))$ for β that lies in a small neighbourhood of β_0 and $\alpha_1(w)$ that lies in a small neighbourhood of $\dot{f}_0(w)$. Thus, we have $\breve{u}_w(\beta_0, \dot{f}_0(w)) = 0$. Denote $\breve{U}_w(\beta, \alpha_1(w)) = (\breve{U}_{w1}(\beta, \alpha_1(w)), \breve{U}_{w2}(\beta, \alpha_1(w)))^\top$. Then we have

$$
\begin{aligned}
&\left.\frac{\partial \breve{U}_{w1}(\beta, \alpha_1(w))}{\partial \beta}\right|_{\beta=\beta_0, \alpha_1=\dot{f}_0} \\
&= -\frac{1}{n}\sum_{i=1}^n \int_0^\tau (Z_i - \frac{1}{B_{2w}(t)}[B_{2w}^Z(t) + \int_t^\tau [B_{1w}^Z(s) - \frac{B_{1w}(s)B_{2w}^Z(s)}{B_{2w}(s)}]B_w(t,s)dH_0(s)]) \\
&\quad \times Z_i^T K_h(W_i - w) Y_i(t) \dot{\lambda}_\epsilon (H_0(t) + Z_i^\top \beta_0 + f_0(W_i))\, \mathrm{d}H_0(t) + o_p(1) \\
&= -\frac{1}{n}\sum_{i=1}^n \int_0^\tau (Z_i - z_w(t)) Z_i^\top K_h(W_i - w) Y_i(t) \dot{\lambda}_\epsilon (H_0(t) + Z_i^\top \beta_0 + f_0(W_i))\, \mathrm{d}H_0(t) + o_p(1) \\
&= -\int_0^\tau g(w) E[(Z - z_w(t)) Z^\top Y(t) \dot{\lambda}_\epsilon (H_0(t) + Z_i^\top \beta_0 + f_0(W)) \mid W = w]\, \mathrm{d}H_0(t) + o_p(1) \\
&:= R_1 + o_p(1), \\
&\left.\frac{\partial \breve{U}_{w1}(\beta, \alpha_1(w))}{\partial \alpha_1(w)}\right|_{\beta=\beta_0, \alpha_1=\dot{f}_0} \\
&= -\frac{1}{n}\sum_{i=1}^n \int_0^\tau Z_i K_h(W_i - w) Y_i(t) \dot{\lambda}_\epsilon (H_0(t) + Z_i^\top \beta_0 + f_0(W_i))(W_i - w)\, \mathrm{d}H_0(t) + o_p(1) \\
&= o_p(1), \\
&\left.\frac{\partial \breve{U}_{w2}(\beta, \alpha_1(w))}{\partial \beta}\right|_{\beta=\beta_0, \alpha_1=\dot{f}_0} \\
&= -\frac{1}{n}\sum_{i=1}^n \int_0^\tau (W_i - w) K_h(W_i - w) Y_i(t) \lambda_\epsilon (H_0(t) + Z_i^\top \beta_0 + \dot{f}_0(W_i))\mathrm{d} \left.\frac{\partial \breve{H}(t; \beta, \alpha_1(w))}{\partial \beta}\right|_{\beta=\beta_0, \alpha_1=\dot{f}_0} \\
&\quad - \frac{1}{n}\sum_{i=1}^n \int_0^\tau (W_i - w) K_h(W_i - w) Y_i(t) \dot{\lambda}_\epsilon (H_0(t) + Z_i^\top \beta_0 + \dot{f}_0(W_i)) \\
&\quad \times \left(\left.\frac{\partial \breve{H}(t; \beta, \alpha_1(w))}{\partial \beta}\right|_{\beta=\beta_0, \alpha_1=\dot{f}_0} + Z_i \right) \mathrm{d}H_0(t) + o_p(1) \\
&= o_p(1), \\
&\left.\frac{\partial \breve{U}_{w2}(\beta, \alpha_1(w))}{\partial \alpha_1(w)}\right|_{\beta=\beta_0, \alpha_1=\dot{f}_0} \\
&= -\frac{1}{n}\sum_{i=1}^n \int_0^\tau (W_i - w)^2 K_h(W_i - w) Y_i(t) \dot{\lambda}_\epsilon (H_0(t) + Z_i^\top \beta_0 + f_0(W_i)\, \mathrm{d}H_0(t) + o_p(1) \\
&= -\int_0^\tau h^2 g(w) k_2 E[Y(t) \dot{\lambda}_\epsilon (H_0(t) + Z^\top \beta_0 + f_0(W)) \mid W = w]\, \mathrm{d}H_0(t) + o_p(1) \\
&:= R_2 + o_p(1),
\end{aligned}
$$

where $k_2 = \int w^2 K(w) dw < \infty$. Hence, we obtain

$$\lim_{n\to\infty} \frac{\partial \breve{U}_w(\beta,\alpha_1(w))}{\partial(\beta,\alpha_1(w))}\Bigg|_{\beta=\beta_0,\alpha_1=\dot{f}_0} = \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix} := A_w. \tag{A3}$$

The above calculations also yield

$$\sup_{\beta\in D_{\epsilon_1},\alpha_1\in\mathcal{F}^1_{\epsilon_w}} \left\| \frac{\partial \breve{U}_w(\beta,\alpha_1)}{\partial(\beta,\alpha_1(w))} - A_w \right\| \to 0 \quad \text{as } n\to\infty, \epsilon_1, \epsilon_w \to 0$$

in probability, where $D_{\epsilon_1} = \{\beta \in \mathcal{B} : \|\beta - \beta_0\| \le \epsilon_1\}$ and $\mathcal{F}^1_{\epsilon_w} = \{\alpha_1 : \|\alpha_1(w) - \dot{f}_0(w)\| \le \epsilon_w\}$.

Following arguments used in Step A5 of Chen et al. (2002), consider $\breve{U}_w(\beta,\alpha_1(w))$ as a random mapping from an arbitrarily small but fixed ball $Q_\epsilon = \{(\beta,\alpha_1) : \|(\beta,\alpha_1(w)) - (\beta_0,\dot{f}_0(w)\| \le \epsilon\}$ to another open connected set in $\mathbb{R}^{p+1}$. By the assumption of Lemma A.1, A_w is finite and nonsingular. Then with probability 1, $\breve{U}_w(\beta,\alpha_1(w))$ is homeomorphic from Q_ϵ to E_n, its image. The convergence of $\breve{U}_w(\beta,\dot{f}_0)$ to 0 indicates that E_n contains $0 \in \mathbb{R}^{p+1}$ with probability tending to 1. Because $\breve{U}_w(\breve{\beta},\breve{\alpha}_1(w)) = 0$ and Q_ϵ is an arbitrarily small neighbourhood centred at $(\beta_0,\dot{f}_0(w))$, $\breve{\beta}$ and $\breve{\alpha}_1(w)$ are locally consistent, resulting in the local consistency of $\breve{f}(w)$. This completes the proof. ■

The above analysis has established the local consistency of the naive one-step estimator used as the initial value in the iterative algorithm. Next, we establish the asymptotic properties of the fully iterated estimator. The following proof mimics the approach of Lu and Zhang (2010).

Proof of Theorem* 4.1*: Let us first establish the local consistency of $\hat{H}(\cdot)$, $\hat{\beta}$ and $\hat{f}(\cdot)$ that result from the proposed iterative algorithm. Specifically, we want to show that the proposed estimating equations have unique solutions in small neighbourhoods of the true parameters β_0 and f_0, respectively. Furthermore, for β and f in this neighbourhood, we show that the estimator of $H(\cdot)$ is close to $H_0(\cdot)$. Based on the arguments of Carroll et al. (1997) and Lemma A.1, it is expected that $\hat{\beta}, \hat{\alpha}_0$ and $\hat{\alpha}_1$ are locally consistent estimators of β_0, f_0 and $\dot{f}_0$, respectively. Hence, it suffices to show the local consistency of $\hat{H}(\cdot)$.

For any fixed β and $f(\cdot)$, Equation (4) is monotone with respect to $H(\cdot)$, and thus there exists a unique solution to Equation (4). Let $\hat{H}(\cdot;\beta,f)$ be the function implicitly defined as the unique solution of Equation (4). Similar to the proof given in Kim et al. (2013), we first prove the consistency of $\hat{H}(\cdot;\beta_0,f_0)$ to $H_0(\cdot)$, i.e. $\sup_{t\in[0,\tau]} |\hat{H}(t;\beta_0,f_0) - H_0(t)| \to 0$ in probability as $n\to\infty$. By the monotonicity of $\hat{H}(\cdot;\beta_0,f_0)$, it suffices to show that $\bar{H}(\cdot)$ is identical to $H_0(\cdot)$, where $\bar{H}(\cdot)$ is a limit function of $\hat{H}(\cdot;\beta_0,f_0)$ defined on $[0,\tau]$. By the law of large numbers, we obtain

$$E[N(t)] = \int_0^t E[Y(s)\lambda_\epsilon(\bar{H}(s) + Z^\top\beta_0 + f_0(W))]\,\mathrm{d}\bar{H}(s)$$

from Equation (4). This indicates that $\bar{H}(\cdot)$ is differentiable and must therefore satisfy

$$\frac{\mathrm{d}\bar{H}(t)}{\mathrm{d}t} = \frac{\mathrm{d}E[N(t)]}{\mathrm{d}t}\{E[Y(t)\lambda_\epsilon(\bar{H}(t) + Z^\top\beta_0 + f_0(W))]\}^{-1}. \tag{A4}$$

Note that as Equation (A4) is a Cauchy problem, it results in a unique solution under some local smoothness assumptions (see Theorem 3.4.2 in Reinhard 1986, p. 40). Moreover, by the definition of $M(t)$, $H_0(\cdot)$ satisfies Equation (A4), hence we obtain $\bar{H}(\cdot) = H_0(\cdot)$, and $\hat{H}(\cdot;\beta_0,f_0)$ converges to $H_0(\cdot)$.

Similar to the proof of Lemma A.1, for t in a compact subset of the interior of the support of X, we can show that the derivatives of $\hat{H}(t;\beta,\alpha_0)$ with respect to β and α_0 are bounded in a small neighbourhood of β_0 and f_0, respectively. Thus, we have $\hat{H}(t;\hat{\beta},\hat{\alpha}_0) \to \hat{H}(t;\beta_0,f_0)$, provided that $\hat{\beta} \to \beta_0$ and $\hat{\alpha}_0 \to f_0$. This yields $\hat{H}(t;\hat{\beta},\hat{\alpha}_0) \to H_0(t)$ if $\hat{\beta} \to \beta_0$ and $\hat{\alpha}_0 \to f_0$ hold, meaning that $\hat{H}(t;\hat{\beta},\hat{\alpha}_0)$ is consistent. Next, we prove the asymptotic normality of $\hat{\beta}$. Our proof consists of 4 parts.

Part 1: By the definition of $M_i(t)$ and Equation (4), it follows from the law of large numbers that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \mathrm{d}M_i(t) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{d}N_i(t) - \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\,\mathrm{d}\Lambda_\epsilon(H_0(t) + Z_i^\top \beta_0 + f_0(W_i)) \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\,\mathrm{d}\left\{\frac{\lambda_\epsilon(H_0(t) + Z_i^\top\beta_0 + f_0(W_i))}{\lambda^*\{H_0(t)\}}(\Lambda^*\{\hat{H}(t;\beta_0,f_0)\} - \Lambda^*\{H_0(t)\})\right\} + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\frac{\lambda_\epsilon(H_0(t) + Z_i^\top\beta_0 + f_0(W_i))}{\lambda^*\{H_0(t)\}}\mathrm{d}[\Lambda^*\{\hat{H}(t;\beta_0,f_0)\} - \Lambda^*\{H_0(t)\}] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} Y_i(t)[\Lambda^*\{\hat{H}(t;\beta_0,f_0)\} - \Lambda^*\{H_0(t)\}]\,\mathrm{d}\frac{\lambda_\epsilon(H_0(t) + Z_i^\top\beta_0 + f_0(W_i))}{\lambda^*\{H_0(t)\}} + o_p(n^{-1/2}) \\
&= \frac{B_2(t)}{\lambda^*\{H_0(t)\}}\mathrm{d}[\Lambda^*\{\hat{H}(t;\beta_0,f_0)\} - \Lambda^*\{H_0(t)\}] \\
&\quad + [\Lambda^*\{\hat{H}(t;\beta_0,f_0)\} - \Lambda^*\{H_0(t)\}]\frac{B_1(t)\,\mathrm{d}H_0(t) - B_2(t)\frac{B_1(t)}{B_2(t)}\mathrm{d}H_0(t)}{\lambda^*\{H_0(t)\}} + o_p(n^{-1/2}) \\
&= \frac{B_2(t)}{\lambda^*\{H_0(t)\}}\mathrm{d}[\Lambda^*\{\hat{H}(t;\beta_0,f_0)\} - \Lambda^*\{H_0(t)\}] + o_p(n^{-1/2}),
\end{aligned}$$

which yields

$$\Lambda^*\{\hat{H}(t,\beta_0,f_0)\} - \Lambda^*\{H_0(t)\} = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)}\mathrm{d}M_i(s) + o_p(n^{-1/2}).$$

Note that the $o_p(n^{-1/2})$ term on the r.h.s. of the above equation is due to the $\sqrt{n}$-consistency of $\hat{H}(\cdot;\beta_0,f_0)$, which can be established by the empirical process theory for Z-estimators (van der Vaart and Wellner 1996).

Part 2: From Equation (4), note that

$$\sum_{i=1}^{n}\{\mathrm{d}N_i(t) - Y_i(t)\,\mathrm{d}\Lambda_\epsilon(\hat{H}(t;\beta,f) + Z_i^\top\beta + f(W_i))\} = 0. \tag{A5}$$

Taking derivative with respect to β on both sides of Equation (A5), we have

$$\begin{aligned}
&\sum_{i=1}^{n} Y_i(t)\lambda_\epsilon(\hat{H}(t;\beta,f) + Z_i^\top\beta + f(W_i))\,\mathrm{d}\frac{\partial\hat{H}(t;\beta,f)}{\partial\beta} \\
&\quad + \sum_{i=1}^{n} Y_i(t)\dot{\lambda}_\epsilon(\hat{H}(t;\beta,f) + Z_i^\top\beta + f(W_i))\left(\frac{\partial\hat{H}(t;\beta,f)}{\partial\beta} + Z_i\right)\mathrm{d}\hat{H}(t;\beta,f) = 0.
\end{aligned}$$

Using the law of large numbers and recognising that $\hat{H}(t;\beta_0,f_0)$ converges to $H_0(t)$, we obtain

$$\left.\frac{\partial\hat{H}(t;\beta,f)}{\partial\beta}\right|_{\beta=\beta_0,f=f_0} = -\int_0^t \frac{B(s,t)}{B_2(s)}B_1^Z(s)\,\mathrm{d}H_0(s) + o_p(1). \tag{A6}$$

Hence we have

$$\mathrm{d}\left.\frac{\partial \hat{H}(t;\beta,f)}{\partial \beta}\right|_{\beta=\beta_0,f=f_0} = -\frac{1}{B_2(t)}\left\{B_1^Z(t) + B_1(t)\left.\frac{\partial \hat{H}(t;\beta,f)}{\partial \beta}\right|_{\beta=\beta_0,f=f_0} + o_p(1)\right\}\mathrm{d}H_0(t). \quad \text{(A7)}$$

Denote

$$V_1(\beta,f) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i\left\{\mathrm{d}N_i(t) - Y_i(t)\,\mathrm{d}\Lambda_\epsilon(\hat{H}(t;\beta,f) + Z_i^T\beta + f(W_i))\right\}$$

obtained by substituting $\hat{H}(t;\beta,f)$ in Equation (5). By differentiating $V_1(\beta,f)$ with respect to β, setting $\beta = \beta_0$ and $f = f_0$, and using the law of large numbers and Equations (A6) and (A7), we obtain

$$\begin{aligned}
&\left.\frac{\partial V_1(\beta,f)}{\partial \beta}\right|_{\beta=\beta_0,f=f_0} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i Y_i(t)\lambda_\epsilon(\hat{H}(t;\beta_0,f_0) + Z_i^\top\beta_0 + f_0(W_i))\,\mathrm{d}\left.\frac{\partial \hat{H}(t;\beta,f)}{\partial \beta}\right|_{\beta=\beta_0,f=f_0} \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i Y_i(t)\left(Z_i + \left.\frac{\partial \hat{H}(t;\beta,f)}{\partial \beta}\right|_{\beta=\beta_0,f=f_0}\right)^\top \dot{\lambda}_\epsilon(\hat{H}(t;\beta_0,f_0) \\
&\quad + Z_i^\top\beta_0 + f_0(W_i))\,\mathrm{d}\hat{H}(t;\beta_0,f_0) \\
&= -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{Z_i - z(t)\}Z_i^\top Y_i(t)\dot{\lambda}_\epsilon(H_0(t) + Z_i^\top\beta_0 + f_0(W_i))\,\mathrm{d}H_0(t) + o_p(1) \\
&= -\int_0^{\tau} E[\{Z - z(t)\}Z^\top Y(t)\dot{\lambda}_\epsilon(H_0(t) + Z^\top\beta_0 + f_0(W))]\,\mathrm{d}H_0(t) + o_p(1) \\
&= -A_1 + o_p(1).
\end{aligned}$$

Part 3: For any $w \in \mathcal{W}$, denote

$$\begin{aligned}
V_2(\alpha_0,\alpha_1,H,\beta)(w) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(W_i - w)\begin{pmatrix} 1 \\ \frac{W_i - w}{h}\end{pmatrix}[\mathrm{d}N_i(t) - Y_i(t)\,\mathrm{d}\Lambda_\epsilon\{H(t) \\
+ Z_i^\top\beta + \alpha_0(w) + \alpha_1(w)(W_i - w)\}].
\end{aligned}$$

Then we have $V_2(\hat{\alpha}_0,\hat{\alpha}_1,\hat{H}(\cdot;\hat{\beta},\hat{\alpha}_0),\hat{\beta})(w) = 0$, where $(\hat{\alpha}_0,\hat{\alpha}_1)$ is the solution of Equation (7) at convergence, and $(\hat{\beta},\hat{H}(\cdot;\hat{\beta},\hat{\alpha}_0))$ is the solution of Equations (4) and (5) at convergence. Using the Taylor series expansion and the law of large numbers, we have

$$\begin{aligned}
&V_2(\hat{\alpha}_0,\hat{\alpha}_1,\hat{H}(\cdot;\hat{\beta},\hat{\alpha}_0),\hat{\beta})(w) \\
&= V_2(\hat{\alpha}_0,\hat{\alpha}_1,\hat{H}(\cdot;\beta_0,\hat{\alpha}_0),\beta_0)(w) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(W_i - w)Y_i(t)\begin{pmatrix} 1 \\ \frac{W_i - w}{h}\end{pmatrix}[\mathrm{d}\Lambda_\epsilon\{\hat{H}(t;\hat{\beta},\hat{\alpha}_0) \\
&\quad + Z_i^\top\hat{\beta} + \hat{\alpha}_0(w) + \hat{\alpha}_1(w)(W_i - w)\} \\
&\quad - \mathrm{d}\Lambda_\epsilon\{\hat{H}(t;\beta_0,\hat{\alpha}_0) + Z_i^\top\beta_0 + \hat{\alpha}_0(w) + \hat{\alpha}_1(w)(W_i - w)\}] \\
&= V_2(\hat{\alpha}_0,\hat{\alpha}_1,\hat{H}(\cdot;\beta_0,\hat{\alpha}_0),\beta_0)(w) - E_1(w) + o_p(n^{-1/2}),
\end{aligned}$$

where

$$
\begin{aligned}
&E_1(w)\\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(W_i - w)Y_i(t)\begin{pmatrix}1\\ \frac{W_i - w}{h}\end{pmatrix}\mathrm{d}\Bigg[\lambda_\epsilon\{\hat{H}(t;\beta_0,\hat{\alpha}_0) + Z_i^\top\beta_0 + \hat{\alpha}_0(w) + \hat{\alpha}_1(w)(W_i - w)\}\\
&\quad\times\left(Z_i + \left.\frac{\partial \hat{H}(t;\beta,\hat{\alpha}_0)}{\partial\beta}\right|_{\beta=\beta_0}\right)^\top(\hat{\beta}-\beta_0)\Bigg]\\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(W_i - w)Y_i(t)\begin{pmatrix}1\\ \frac{W_i - w}{h}\end{pmatrix}\lambda_\epsilon\{\hat{H}(t;\beta_0,\hat{\alpha}_0) + Z_i^\top\beta_0 + \hat{\alpha}_0(w) + \hat{\alpha}_1(w)(W_i - w)\}\\
&\quad\times\left(\mathrm{d}\left.\frac{\partial \hat{H}(t;\beta,\hat{\alpha}_0)}{\partial\beta}\right|_{\beta=\beta_0}\right)^\top(\hat{\beta}-\beta_0)\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(W_i - w)Y_i(t)\begin{pmatrix}1\\ \frac{W_i - w}{h}\end{pmatrix}\dot{\lambda}_\epsilon\{\hat{H}(t;\beta_0,\hat{\alpha}_0) + Z_i^\top\beta_0 + \hat{\alpha}_0(w) + \hat{\alpha}_1(w)(W_i - w)\}\\
&\quad\times\left(Z_i + \left.\frac{\partial \hat{H}(t;\beta,\hat{\alpha}_0)}{\partial\beta}\right|_{\beta=\beta_0}\right)^\top \mathrm{d}\hat{H}(t;\beta_0,\hat{\alpha}_0)(\hat{\beta}-\beta_0).
\end{aligned}
$$

Similar to Part 2, we can obtain

$$
\left.\frac{\partial \hat{H}(t;\beta,\hat{\alpha}_0)}{\partial\beta}\right|_{\beta=\beta_0} = -\int_0^t \frac{B(s,t)}{B_2(s)}B_1^Z(s)\,\mathrm{d}H_0(s) + o_p(1) \tag{A8}
$$

and

$$
\mathrm{d}\left.\frac{\partial \hat{H}(t;\beta,\hat{\alpha}_0)}{\partial\beta}\right|_{\beta=\beta_0} = -\frac{1}{B_2(t)}\left\{B_1^Z(t) + B_1(t)\left.\frac{\partial \hat{H}(t;\beta,\hat{\alpha}_0)}{\partial\beta}\right|_{\beta=\beta_0} + o_p(1)\right\}dH_0(t). \tag{A9}
$$

Using standard nonparametric techniques and the law of large numbers, and substituting Equations (A8) and (A9) into $E_1(w)$, we can show that $E_1(w)$ converges to the following deterministic function:

$$
\begin{aligned}
&E_1(w)\\
&= -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(W_i - w)Y_i(t)\begin{pmatrix}1\\ \frac{W_i - w}{h}\end{pmatrix}\lambda_\epsilon\{\hat{H}(t;\beta_0,\hat{\alpha}_0) + Z_i^\top\beta_0 + \hat{\alpha}_0(w) + \hat{\alpha}_1(w)(W_i - w)\}\\
&\quad\times\frac{1}{B_2(t)}\left\{B_1^Z(t) - B_1(t)\int_0^t \frac{B(s,t)}{B_2(s)}B_1^Z(s)\,\mathrm{d}H_0(s)\right\}\mathrm{d}H_0(t)(\hat{\beta}-\beta_0)\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(W_i - w)Y_i(t)\begin{pmatrix}1\\ \frac{W_i - w}{h}\end{pmatrix}\dot{\lambda}_\epsilon\{\hat{H}(t;\beta_0,\hat{\alpha}_0) + Z_i^\top\beta_0 + \hat{\alpha}_0(w) + \hat{\alpha}_1(w)(W_i - w)\}\\
&\quad\times\left(Z_i - \int_0^t \frac{B(s,t)}{B_2(s)}B_1^Z(s)\,\mathrm{d}H_0(s)\right)^\top \mathrm{d}\hat{H}(t;\beta_0,\hat{\alpha}_0)(\hat{\beta}-\beta_0) + o_p(n^{-1/2})\\
&:= \begin{pmatrix}e_1^\top(w)\\ 0^\top\end{pmatrix}(\hat{\beta}-\beta_0) + o_p(n^{-1/2}),
\end{aligned}
$$

where

$$e_1^\top(w) = g(w)\int_0^\tau \left\{B_{1w}^Z(t) - \frac{B_{2w}(t)}{B_2(t)}\left\{B_1^Z(t) - B_1(t)\int_0^t \frac{B(s,t)}{B_2(s)}B_1^Z(s)\,\mathrm{d}H_0(s)\right\}\right.$$
$$\left. - B_{1w}(t)\int_0^t \frac{B(s,t)}{B_2(s)}B_1^Z(s)\,\mathrm{d}H_0(s)\right\}\mathrm{d}H_0(t).$$

In addition, we can obtain

$$\begin{aligned}
&V_2(\hat\alpha_0, \hat\alpha_1, \hat H(\cdot;\beta_0,\hat\alpha_0), \beta_0)(w)\\
&\quad = V_2(\hat\alpha_0, \hat\alpha_1, H_0, \beta_0)(w)\\
&\qquad - \frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(W_i - w)Y_i(t)\begin{pmatrix}1\\ \frac{W_i - w}{h}\end{pmatrix}[\mathrm{d}\Lambda_\epsilon\{\hat H(t;\beta_0,\hat\alpha_0)\\
&\qquad + Z_i^\top\beta_0 + \hat\alpha_0(w) + \hat\alpha_1(w)(W_i - w)\}\\
&\qquad - \mathrm{d}\Lambda_\epsilon\{H_0(t) + Z_i^\top\beta_0 + \hat\alpha_0(w) + \hat\alpha_1(w)(W_i - w)\}]\\
&\quad = V_2(\hat\alpha_0, \hat\alpha_1, H_0, \beta_0)(w) - E_2(w) + o_p(n^{-1/2}),
\end{aligned}$$

where

$$\begin{aligned}
E_2(w) &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(W_i - w)Y_i(t)\begin{pmatrix}1\\ \frac{W_i - w}{h}\end{pmatrix}\mathrm{d}\left[\frac{\lambda_\epsilon\{H_0(t) + Z_i^\top\beta_0 + \hat\alpha_0(w) + \hat\alpha_1(w)(W_i - w)\}}{\lambda^*\{H_0(t)\}}\right.\\
&\quad \left.\times [\Lambda^*\{\hat H(t;\beta_0,\hat\alpha_0)\} - \Lambda^*\{H_0(t)\}]\right]\\
&= \frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(W_i - w)Y_i(t)\begin{pmatrix}1\\ \frac{W_i - w}{h}\end{pmatrix}\frac{\lambda_\epsilon\{H_0(t) + Z_i^\top\beta_0 + \hat\alpha_0(w) + \hat\alpha_1(w)(W_i - w)\}}{\lambda^*\{H_0(t)\}}\\
&\quad \times \mathrm{d}[\Lambda^*\{\hat H(t;\beta_0,\hat\alpha_0)\} - \Lambda^*\{H_0(t)\}]\\
&\quad + \frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(W_i - w)Y_i(t)\begin{pmatrix}1\\ \frac{W_i - w}{h}\end{pmatrix}[\Lambda^*\{\hat H(t;\beta_0,\hat\alpha_0)\} - \Lambda^*\{H_0(t)\}]\\
&\quad \times \mathrm{d}\frac{\lambda_\epsilon\{H_0(t) + Z_i^\top\beta_0 + \hat\alpha_0(w) + \hat\alpha_1(w)(W_i - w)\}}{\lambda^*\{H_0(t)\}}\\
&= g(w)\int_0^\tau \begin{pmatrix}\frac{B_{2w}(t)}{\lambda^*\{H_0(t)\}}\\ 0\end{pmatrix}\mathrm{d}[\Lambda^*\{\hat H(t;\beta_0,\hat\alpha_0)\} - \Lambda^*\{H_0(t)\}] + o_p(n^{-1/2})\\
&:= \int_0^\tau \begin{pmatrix}e_2(w,t)\\ 0\end{pmatrix}\mathrm{d}[\Lambda^*\{\hat H(t;\beta_0,\hat\alpha_0)\} - \Lambda^*\{H_0(t)\}] + o_p(n^{-1/2}),
\end{aligned}$$

with

$$e_2(w,t) = g(w)\frac{B_{2w}(t)}{\lambda^*\{H_0(t)\}}.$$

Furthermore,

$$\begin{aligned}
&V_2(\hat\alpha_0, \hat\alpha_1, H_0, \beta_0)(w)\\
&\quad = V_2(f_0, \dot f_0, H_0, \beta_0)(w)\\
&\qquad - \frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(W_i - w)Y_i(t)\begin{pmatrix}1\\ \frac{W_i - w}{h}\end{pmatrix}[\mathrm{d}\Lambda_\epsilon\{H_0(t) + Z_i^\top\beta_0 + \hat\alpha_0(w) + \hat\alpha_1(w)(W_i - w)\}
\end{aligned}$$

$$- \mathrm{d}\Lambda_\epsilon\{H_0(t) + Z_i^\top \beta_0 + f_0(w) + \dot{f}_0(w)(W_i - w)\}]$$
$$= V_2(f_0, \dot{f}_0, H_0, \beta_0)(w) - E_3(w) \begin{pmatrix} \hat{\alpha}_0(w) - f_0(w) \\ h(\hat{\alpha}_1(w) - \dot{f}_0(w)) \end{pmatrix} + o_p(n^{-1/2}),$$

where

$$E_3(w) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(W_i - w) Y_i(t) \begin{pmatrix} 1 \\ \frac{W_i - w}{h} \end{pmatrix} \begin{pmatrix} 1 & \frac{W_i - w}{h} \end{pmatrix} \dot{\lambda}_\epsilon\{H_0(t)$$
$$+ Z_i^\top \beta_0 + f_0(w) + \dot{f}_0(w)(W_i - w)\}\,\mathrm{d}H_0(t)$$
$$= g(w) \int_0^\tau \begin{pmatrix} 1 & 0 \\ 0 & k_2 \end{pmatrix} B_{1w}(t)\,\mathrm{d}H_0(t) + o_p(n^{-1/2})$$
$$:= e_3(w) + o_p(n^{-1/2}).$$

The above calculations lead to

$$E_3(w) \begin{pmatrix} \hat{\alpha}_0(w) - f_0(w) \\ h(\hat{\alpha}_1(w) - \dot{f}_0(w)) \end{pmatrix} = V_2(f_0, \dot{f}_0, H_0, \beta_0)(w) - E_1(w) - E_2(w) + o_p(n^{-1/2}).$$

Thus we have

$$\begin{pmatrix} \hat{\alpha}_0(w) - f_0(w) \\ h(\hat{\alpha}_1(w) - \dot{f}_0(w)) \end{pmatrix}$$
$$= e_3^{-1}(w) V_2\{f_0, \dot{f}_0, H_0, \beta_0\}(w) - e_3^{-1}(w) \begin{pmatrix} e_1^\top(w) \\ 0^\top \end{pmatrix} (\hat{\beta} - \beta_0)$$
$$- e_3^{-1}(w) \int_0^\tau \begin{pmatrix} e_2(w,t) \\ 0 \end{pmatrix} \mathrm{d}[\Lambda^*\{\hat{H}(t; \beta_0, \hat{\alpha}_0)\} - \Lambda^*\{H_0(t)\}] + o_p(n^{-1/2}),$$

where

$$e_3^{-1}(w) = \frac{1}{\int_0^\tau g(w) B_{1w}(t)\,\mathrm{d}H_0(t)} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{k_2} \end{pmatrix} := \begin{pmatrix} \frac{1}{e_{31}(w)} & 0 \\ 0 & \frac{1}{k_2 e_{31}(w)} \end{pmatrix}.$$

Specifically, for any $w \in \mathcal{W}$, the asymptotic representations of $\hat{\alpha}_0(w) - f_0(w)$ and $h(\hat{\alpha}_1(w) - \dot{f}_0(w))$ are

$$\hat{\alpha}_0(w) - f_0(w) = \frac{1}{e_{31}(w)} V_{21}\{f_0, \dot{f}_0, H_0, \beta_0\}(w) - \frac{e_1^\top(w)}{e_{31}(w)}(\hat{\beta} - \beta_0)$$
$$- \int_0^\tau \frac{e_2(w,t)}{e_{31}(w)} \mathrm{d}[\Lambda^*\{\hat{H}(t; \beta_0, \hat{\alpha}_0)\} - \Lambda^*\{H_0(t)\}] + o_p(n^{-1/2}) \qquad \text{(A10)}$$

and

$$h(\hat{\alpha}_1(w) - \dot{f}_0(w)) = \frac{1}{k_2 e_{31}(w)} V_{22}\{f_0, \dot{f}_0, H_0, \beta_0\}(w) + o_p(n^{-1/2}),$$

where $V_2(f_0, \dot{f}_0, H_0, \beta_0)(w) = (V_{21}\{f_0, \dot{f}_0, H_0, \beta_0\}(w), V_{22}\{f_0, \dot{f}_0, H_0, \beta_0\}(w))^\top$.

Part 4: For any fixed β, $\hat{H}(t; \beta, \hat{\alpha}_0)$ is the solution to the equation

$$\sum_{i=1}^n [\mathrm{d}N_i(t) - Y_i(t)\,\mathrm{d}\Lambda_\epsilon(H(t) + Z_i^\top \beta + \hat{\alpha}_0(W_i))] = 0, \qquad \text{(A11)}$$

and $\hat{\beta}$ is the solution to the estimating equation

$$\sum_{i=1}^n \int_0^\tau Z_i[\mathrm{d}N_i(t) - Y_i(t)\,\mathrm{d}\Lambda_\epsilon(\hat{H}(t; \beta, \hat{\alpha}_0) + Z_i^\top \beta + \hat{\alpha}_0(W_i))] = 0. \qquad \text{(A12)}$$

Using Equation (A11) and mimicking the procedures in Part 1, we have

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{d}M_i(t)$$
$$= \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\,\mathrm{d}\Lambda_\epsilon(\hat{H}(t;\beta_0,\hat{\alpha}_0) + Z_i^\top\beta_0 + \hat{\alpha}_0(W_i)) - \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\,\mathrm{d}\Lambda_\epsilon(H_0(t) + Z_i^\top\beta_0 + f_0(W_i))$$
$$= \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\,\mathrm{d}\left[\frac{\lambda_\epsilon(H_0(t) + Z_i^\top\beta_0 + f_0(W_i))}{\lambda^*\{H_0(t)\}}[\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\} - \Lambda^*\{H_0(t)\}]\right.$$
$$\left. + \lambda_\epsilon(H_0(t) + Z_i^\top\beta_0 + f_0(W_i))(\hat{\alpha}_0(W_i) - f_0(W_i))\right] + o_p(n^{-1/2})$$
$$= \frac{B_2(t)}{\lambda^*\{H_0(t)\}} d[\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\} - \Lambda^*\{H_0(t)\}]$$
$$+ \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\{\hat{\alpha}_0(W_i) - f_0(W_i)\}\,\mathrm{d}\lambda_\epsilon(H_0(t) + Z_i^\top\beta_0 + f_0(W_i)) + o_p(n^{-1/2}). \quad \text{(A13)}$$

Write the l.h.s. of Equation (A12) as $nU(\beta,\hat{H}(t;\beta,\hat{\alpha}_0),\hat{\alpha}_0)$, i.e.

$$U(\beta,\hat{H}(t;\beta,\hat{\alpha}_0),\hat{\alpha}_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau Z_i[\mathrm{d}N_i(t) - Y_i(t)\,\mathrm{d}\Lambda_\epsilon\{\hat{H}(t;\beta,\hat{\alpha}_0) + Z_i^\top\beta + \hat{\alpha}_0(W_i)\}].$$

As $U(\hat{\beta},\hat{H}(t;\hat{\beta},\hat{\alpha}_0),\hat{\alpha}_0) = 0$, by the Taylor series expansion, we have

$$U(\hat{\beta},\hat{H}(t;\hat{\beta},\hat{\alpha}_0),\hat{\alpha}_0)$$
$$= U(\beta_0,\hat{H}(t;\beta_0,\hat{\alpha}_0),f_0) - \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau Z_iY_i(t)\,\mathrm{d}[\Lambda_\epsilon(\hat{H}(t;\hat{\beta},\hat{\alpha}_0) + Z_i^\top\hat{\beta} + \hat{\alpha}_0(W_i))$$
$$- \Lambda_\epsilon(\hat{H}(t;\beta_0,\hat{\alpha}_0) + Z_i^\top\beta_0 + f_0(W_i))]$$
$$= U(\beta_0,\hat{H}(t;\beta_0,\hat{\alpha}_0),f_0) - \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau Z_iY_i(t)\,\mathrm{d}[\lambda_\epsilon(\hat{H}(t;\beta_0,\hat{\alpha}_0)$$
$$+ Z_i^\top\beta_0 + f_0(W_i))\{\hat{\alpha}_0(W_i) - f_0(W_i)\}]$$
$$- \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau Z_iY_i(t)\,\mathrm{d}[\lambda_\epsilon(\hat{H}(t;\beta_0,\hat{\alpha}_0) + Z_i^\top\beta_0 + f_0(W_i))$$
$$\times \left\{Z_i + \left.\frac{\partial\hat{H}(t;\beta,\hat{\alpha}_0)}{\partial\beta}\right|_{\beta=\beta_0}\right\}^\top (\hat{\beta} - \beta_0)] + o_p(n^{-1/2})$$
$$= U(\beta_0,\hat{H}(t;\beta_0,\hat{\alpha}_0),f_0) - \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau Z_iY_i(t)\{\hat{\alpha}_0(W_i) - f_0(W_i)\}\,\mathrm{d}\lambda_\epsilon(H_0(t) + Z_i^\top\beta_0 + f_0(W_i))$$
$$- \int_0^\tau E[\{Z - z(t)\}Z^\top Y(t)\dot{\lambda}_\epsilon(H_0(t) + Z^\top\beta_0 + f_0(W))]\,\mathrm{d}H_0(t)(\hat{\beta} - \beta_0) + o_p(n^{-1/2})$$

$$= U(\beta_0, \hat{H}(t;\beta_0,\hat{\alpha}_0), f_0) - \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i Y_i(t)\{\hat{\alpha}_0(W_i) - f_0(W_i)\}\,\mathrm{d}\lambda_{\epsilon}(H_0(t) + Z_i^{\top}\beta_0 + f_0(W_i))$$
$$- A_1(\hat{\beta} - \beta_0) + o_p(n^{-1/2}).$$

Moreover,

$$U(\beta_0, \hat{H}(t;\beta_0,\hat{\alpha}_0), f_0)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i[\mathrm{d}N_i(t) - Y_i(t)\,\mathrm{d}\Lambda_{\epsilon}\{\hat{H}(t;\beta_0,\hat{\alpha}_0) + Z_i^{\top}\beta_0 + f_0(W_i)\}]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i\,\mathrm{d}M_i(t) + \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i Y_i(t)\,\mathrm{d}\Lambda_{\epsilon}\{H_0(t) + Z_i^{\top}\beta_0 + f_0(W_i)\}$$
$$- \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i Y_i(t)\,\mathrm{d}\Lambda_{\epsilon}\{\hat{H}(t;\beta_0, f_0) + Z_i^{\top}\beta_0 + f_0(W_i)\}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i\,\mathrm{d}M_i(t)$$
$$- \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i Y_i(t)\,\mathrm{d}\left[\frac{\lambda_{\epsilon}(H_0(t) + Z_i^{\top}\beta_0 + f_0(W_i))}{\lambda^*\{H_0(t)\}}[\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\} - \Lambda^*\{H_0(t)\}]\right]$$
$$+ o_p(n^{-1/2})$$
$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i\,\mathrm{d}M_i(t) - \int_0^{\tau}\frac{B_2^Z(t)}{\lambda^*\{H_0(t)\}}\mathrm{d}[\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\} - \Lambda^*\{H_0(t)\}]$$
$$- \int_0^{\tau}[\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\} - \Lambda^*\{H_0(t)\}]\frac{B_1^Z(t) - B_2^Z(t)\frac{B_1(t)}{B_2(t)}}{\lambda^*\{H_0(t)\}}dH_0(t) + o_p(n^{-1/2}).$$

The above calculations lead to

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i dM_i(t)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i Y_i(t)\{\hat{\alpha}_0(W_i) - f_0(W_i)\}\,\mathrm{d}\lambda_{\epsilon}(H_0(t) + Z_i^{\top}\beta_0 + f_0(W_i)) + A_1(\hat{\beta} - \beta_0)$$
$$+ \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\frac{B_2^Z(t)}{\lambda^*\{H_0(t)\}}\mathrm{d}[\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\} - \Lambda^*\{H_0(t)\}]$$
$$+ \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}[\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\} - \Lambda^*\{H_0(t)\}]\frac{B_1^Z(t) - B_2^Z(t)\frac{B_1(t)}{B_2(t)}}{\lambda^*\{H_0(t)\}}\mathrm{d}H_0(t) + o_p(n^{-1/2}). \quad \text{(A14)}$$

Substituting Equation (A10), the asymptotic representations of $\hat{\alpha}_0(w) - f_0(w)$, into Equation (A13), we obtain

$$\frac{B_2(t)}{\lambda^*\{H_0(t)\}}\mathrm{d}[\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\} - \Lambda^*\{H_0(t)\}]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\mathrm{d}M_i(t) - \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\frac{V_{21}\{f_0, \dot{f}_0, H_0, \beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_{\epsilon}(H_0(t) + Z_i^{\top}\beta_0 + f_0(W_i))$$

$$
\begin{aligned}
&+ \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\frac{e_1^\top(W_i)(\hat\beta-\beta_0)}{e_{31}(W_i)}\mathrm{d}\lambda_\epsilon(H_0(t)+Z_i^\top\beta_0+f_0(W_i))\\
&+ \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\int_0^\tau \frac{e_2(W_i,t)}{e_{31}(W_i)}\mathrm{d}[\Lambda^*\{\hat H(t;\beta_0,\hat\alpha_0)\}-\Lambda^*\{H_0(t)\}]\,\mathrm{d}\lambda_\epsilon(H_0(t)+Z_i^\top\beta_0+f_0(W_i))\\
&+ o_p(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{d}M_i(t) - \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\frac{V_{21}\{f_0,\dot f_0,H_0,\beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_\epsilon(H_0(t)+Z_i^\top\beta_0+f_0(W_i))\\
&+ \mathrm{d}\{c_1(t)\}(\hat\beta-\beta_0) + \int_0^\tau c_2(t,s)\,\mathrm{d}[\Lambda^*\{\hat H(s;\beta_0,\hat\alpha_0)\}-\Lambda^*\{H_0(s)\}]\,\mathrm{d}H_0(t) + o_p(n^{-1/2}),
\end{aligned}
$$

where

$$
\mathrm{d}\{c_1(t)\} = E\left\{Y(t)\frac{e_1^\top(W)}{e_{31}(W)}\dot\lambda_\epsilon(H_0(t)+Z^\top\beta_0+f_0(W))\right\}\mathrm{d}H_0(t) \quad \text{and}
$$

$$
c_2(t,s) = E\left\{Y(t)\frac{e_2(W,s)}{e_{31}(W)}\dot\lambda_\epsilon(H_0(t)+Z^\top\beta_0+f_0(W))\right\}.
$$

Hence we have

$$
\begin{aligned}
&\frac{B_2(t)}{\lambda^*\{H_0(t)\}}\mathrm{d}[\Lambda^*\{\hat H(t;\beta_0,\hat\alpha_0)\}-\Lambda^*\{H_0(t)\}] - \mathrm{d}\{c_1(t)\}(\hat\beta-\beta_0)\\
&\quad - \int_0^\tau c_2(t,s)\,\mathrm{d}[\Lambda^*\{\hat H(s;\beta_0,\hat\alpha_0)\}-\Lambda^*\{H_0(s)\}]\,\mathrm{d}H_0(t)\\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\mathrm{d}M_i(t)\\
&\quad - \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\frac{V_{21}\{f_0,\dot f_0,H_0,\beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_\epsilon(H_0(t)+Z_i^\top\beta_0+f_0(W_i)) + o_p(n^{-1/2}). \qquad \text{(A15)}
\end{aligned}
$$

Multiplying $m_Z(t)$ on the both sides of Equation (A15) and integrating both sides of the resultant equation with respect to t from 0 to τ, we obtain

$$
\begin{aligned}
&\int_0^\tau \left[q(t) - \int_0^\tau m_Z(s)c_2(s,t)\,\mathrm{d}H_0(s)\right]\mathrm{d}[\Lambda^*\{\hat H(t;\beta_0,\hat\alpha_0)\}-\Lambda^*\{H_0(t)\}]\\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau m_Z(t)\,\mathrm{d}M_i(t) + A_{21}(\hat\beta-\beta_0)\\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau m_Z(t)Y_i(t)\frac{V_{21}\{f_0,\dot f_0,H_0,\beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_\epsilon(H_0(t)+Z_i^\top\beta_0+f_0(W_i))\\
&\quad + o_p(n^{-1/2}), \qquad \text{(A16)}
\end{aligned}
$$

where

$$
A_{21} = \int_0^\tau m_Z(t)\,\mathrm{d}\{c_1(t)\}.
$$

Similarly, substituting Equation (A10) into Equation (A14), we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i\,\mathrm{d}M_i(t)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_iY_i(t)\frac{V_{21}\{f_0,\dot{f}_0,H_0,\beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_\epsilon(H_0(t)+Z_i^\top\beta_0+f_0(W_i))$$
$$-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_iY_i(t)\frac{e_1^\top(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_\epsilon(H_0(t)+Z_i^\top\beta_0+f_0(W_i))(\hat{\beta}-\beta_0)$$
$$-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_iY_i(t)\int_0^{\tau}\frac{e_2(W_i,s)}{e_{31}(W_i)}\mathrm{d}[\Lambda^*\{\hat{H}(s;\beta_0,\hat{\alpha}_0)\}-\Lambda^*\{H_0(s)\}]\,\mathrm{d}\lambda_\epsilon(H_0(t)$$
$$+Z_i^\top\beta_0+f_0(W_i))$$
$$+A_1(\hat{\beta}-\beta_0)+\int_0^{\tau}\frac{B_2^Z(t)}{\lambda^*\{H_0(t)\}}\mathrm{d}[\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\}-\Lambda^*\{H_0(t)\}]$$
$$+\int_0^{\tau}\int_t^{\tau}\frac{B_1^Z(s)-B_2^Z(s)\frac{B_1(s)}{B_2(s)}}{\lambda^*\{H_0(s)\}}\mathrm{d}H_0(s)\mathrm{d}[\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\}-\Lambda^*\{H_0(t)\}]+o_p(n^{-1/2}),$$

which is equivalent to

$$(A_1-A_{22})(\hat{\beta}-\beta_0)-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i\,\mathrm{d}M_i(t)$$
$$+\int_0^{\tau}\left[\frac{B_2(t)z(t)}{\lambda^*\{H_0(t)\}}-c_3(t)\right]\mathrm{d}[\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\}-\Lambda^*\{H_0(t)\}]$$
$$=-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_iY_i(t)\frac{V_{21}\{f_0,\dot{f}_0,H_0,\beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_\epsilon(H_0(t)$$
$$+Z_i^\top\beta_0+f_0(W_i))+o_p(n^{-1/2}), \tag{A17}$$

where

$$A_{22}=\int_0^{\tau} E\left[ZY(t)\frac{e_1^\top(W)}{e_{31}(W)}\dot{\lambda}_\epsilon(H_0(t)+Z^\top\beta_0+f_0(W))\right]\mathrm{d}H_0(t).$$

Note that $A_2 = A_{22} - A_{21}$ and $q(t)$ is the solution to the following integral equation:

$$q(t)-\int_0^{\tau} q(s)D_1(s,t)\,\mathrm{d}H_0(s)=\frac{B_2(t)z(t)}{\lambda^*\{H_0(t)\}}-c_3(t),$$

where

$$D_1(s,t)=\frac{\lambda^*\{H_0(s)\}}{B_2(s)}c_2(s,t) \quad \text{and}$$
$$c_3(t)=\int_0^{\tau} E\left[ZY(s)\frac{e_2(W,t)}{e_{31}(W)}\dot{\lambda}_\epsilon(H_0(s)+Z^\top\beta_0+f_0(W))\right]\mathrm{d}H_0(s).$$

Thus, by subtracting Equation (A16) from Equation (A17), we obtain

$$
\begin{aligned}
&(A_1 - A_2)(\hat{\beta} - \beta_0) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} [Z_i - m_Z(t)]\,\mathrm{d}M_i(t) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i Y_i(t)\frac{V_{21}\{f_0,\dot{f}_0,H_0,\beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_{\epsilon}(H_0(t) + Z_i^{\top}\beta_0 + f_0(W_i)) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} m_Z(t) Y_i(t)\frac{V_{21}\{f_0,\dot{f}_0,H_0,\beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_{\epsilon}(H_0(t) + Z_i^{\top}\beta_0 + f_0(W_i)) + o_p(n^{-1/2}) \\
&:= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} [Z_i - m_Z(t)]\,\mathrm{d}M_i(t) - (G_1 - G_2) + o_p(n^{-1/2}),
\end{aligned}
$$

where

$$
G_1 = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i Y_i(t)\frac{V_{21}\{f_0,\dot{f}_0,H_0,\beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_{\epsilon}(H_0(t) + Z_i^{\top}\beta_0 + f_0(W_i)) \quad \text{and}
$$

$$
G_2 = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} m_Z(t) Y_i(t)\frac{V_{21}\{f_0,\dot{f}_0,H_0,\beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_{\epsilon}(H_0(t) + Z_i^{\top}\beta_0 + f_0(W_i)).
$$

Applying standard nonparametric techniques together with the Taylor series expansion, we have

$$
\begin{aligned}
G_1 &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \frac{Z_i Y_i(t)}{e_{31}(W_i)}\frac{1}{n}\sum_{j=1}^{n}\int_0^{\tau} K_h(W_j - W_i)[\mathrm{d}N_j(t) - Y_j(t)\,\mathrm{d}\Lambda_{\epsilon}\{H_0(t) \\
&\qquad + Z_j^{\top}\beta_0 + f_0(W_i) + \dot{f}_0(W_i)(W_j - W_i)\}]\,\mathrm{d}\lambda_{\epsilon}(H_0(t) + Z_i^{\top}\beta_0 + f_0(W_i)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \frac{\int_0^{\tau} E[ZY(t)\dot{\lambda}_{\epsilon}\{H_0(t) + Z^{\top}\beta_0 + f_0(W)\} \mid W = W_i]\,\mathrm{d}H_0(t)}{\int_0^{\tau} E[Y(t)\dot{\lambda}_{\epsilon}\{H_0(t) + Z^{\top}\beta_0 + f_0(W)\} \mid W = W_i]\,\mathrm{d}H_0(t)}\mathrm{d}M_i(t) + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Z_i^{*}\mathrm{d}M_i(t) + o_p(n^{-1/2})
\end{aligned}
$$

and

$$
\begin{aligned}
G_2 &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} m_Z(t)\frac{Y_i(t)}{e_{31}(W_i)}\frac{1}{n}\sum_{j=1}^{n}\int_0^{\tau} K_h(W_j - W_i)[\mathrm{d}N_j(t) - Y_j(t)\,\mathrm{d}\Lambda_{\epsilon}\{H_0(t) \\
&\quad + Z_j^{\top}\beta_0 + f_0(W_i) + \dot{f}_0(W_i)(W_j - W_i)\}]\,\mathrm{d}\lambda_{\epsilon}(H_0(t) + Z_i^{\top}\beta_0 + f_0(W_i)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \frac{\int_0^{\tau} m_Z(t)E[Y(t)\dot{\lambda}_{\epsilon}\{H_0(t) + Z^{\top}\beta_0 + f_0(W)\} \mid W = W_i]\,\mathrm{d}H_0(t)}{\int_0^{\tau} E[Y(t)\dot{\lambda}_{\epsilon}\{H_0(t) + Z^{\top}\beta_0 + f_0(W)\} \mid W = W_i]\,\mathrm{d}H_0(t)}\mathrm{d}M_i(t) + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} m_{Z_i^{*}}\,\mathrm{d}M_i(t) + o_p(n^{-1/2}).
\end{aligned}
$$

Combining the above results, we obtain

$$\begin{aligned}
&\sqrt{n}(\hat{\beta} - \beta_0) \\
&\quad = (A_1 - A_2)^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}[Z_i - m_Z(t)]\,\mathrm{d}M_i(t) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}[Z_i^* - m_{Z_i^*}]\,\mathrm{d}M_i(t)\right\} + o_p(1) \\
&\quad = (A_1 - A_2)^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\{[Z_i - m_Z(t)] - [Z_i^* - m_{Z_i^*}]\}\,\mathrm{d}M_i(t)\right\} + o_p(1). \qquad \text{(A18)}
\end{aligned}$$

The asymptotic normality of $\sqrt{n}(\hat{\beta} - \beta_0)$ follows immediately and the proof is completed. ■

We next establish the asymptotic representation of $\sqrt{n}(\hat{H}(t;\hat{\beta},\hat{\alpha}_0) - H_0(t))$. We first give the following Lemma A.2 that is useful for proving Theorem 4.2.

Lemma A.2: *Under the regularity conditions* (C1)–(C7), *if* $nh^2/\{\log(1/h)\} \to \infty$ *and* $nh^4 \to 0$ *as* $n \to \infty$, *then* $\hat{S}_n(t) = \sqrt{n}\{\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\} - \Lambda^*\{H_0(t)\}\}$ *satisfies the following integral equation asymptotically:*

$$\hat{S}_n(t) - \int_0^{\tau} p(t,s)\,\mathrm{d}\hat{S}_n(s) = W_n(t), \quad t \in [0,\tau], \qquad \text{(A19)}$$

where $p(t,s)$ *is a deterministic function* (*to be defined ahead in the proof that follows*), *and* $W_n(t)$ *is a summation of independent mean zero functions, i.e.* $W_n(t) = n^{-1/2}\sum_{i=1}^{n} w_i(t)$, *which converges weakly to a mean zero Gaussian process as* $n \to \infty$.

Proof: By Equation (A15) from the proof of Theorem 4.1, we have

$$\begin{aligned}
&\frac{B_2(t)}{\lambda^*\{H_0(t)\}}\mathrm{d}\hat{S}_n(t) - \int_0^{\tau} c_2(t,s)\,\mathrm{d}\hat{S}_n(s)\,\mathrm{d}H_0(t) \\
&\quad = \mathrm{d}\{c_1(t)\}\sqrt{n}(\hat{\beta} - \beta_0) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{d}M_i(t) \\
&\qquad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i(t)\frac{V_{21}\{f_0,\dot{f}_0,H_0,\beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_\epsilon(H_0(t) + Z_i^\top\beta_0 + f_0(W_i)) + o_p(1).
\end{aligned}$$

Multiplying $\lambda^*\{H_0(t)\}/B_2(t)$ on both sides of the above equation and integrating the equation with respect to t from 0 to t, we have

$$\begin{aligned}
&\hat{S}_n(t) - \int_0^{\tau}\int_0^{t}\frac{\lambda^*\{H_0(u)\}}{B_2(u)}c_2(u,s)\,\mathrm{d}H_0(u)\,\mathrm{d}\hat{S}_n(s) \\
&\quad = \int_0^{t}\frac{\lambda^*\{H_0(u)\}}{B_2(u)}\mathrm{d}\{c_1(u)\}\sqrt{n}(\hat{\beta} - \beta_0) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{t}\frac{\lambda^*\{H_0(u)\}}{B_2(u)}\mathrm{d}M_i(u) \\
&\qquad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{t} Y_i(u)\frac{\lambda^*\{H_0(u)\}}{B_2(u)}\frac{V_{21}\{f_0,\dot{f}_0,H_0,\beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_\epsilon(H_0(u) + Z_i^\top\beta_0 + f_0(W_i)) + o_p(1).
\end{aligned}$$

It follows from Equation (A18) that

$$\begin{aligned}
&\int_0^{t}\frac{\lambda^*\{H_0(u)\}}{B_2(u)}\mathrm{d}\{c_1(u)\}\sqrt{n}(\hat{\beta} - \beta_0) \\
&\quad = \int_0^{t}\frac{\lambda^*\{H_0(u)\}}{B_2(u)}\mathrm{d}\{c_1(u)\}(A_1 - A_2)^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\{[Z_i - m_Z(t)] - [Z_i^* - m_{Z_i^*}]\}\,\mathrm{d}M_i(t)\right\} \\
&\qquad + o_p(1).
\end{aligned}$$

Using standard nonparametric techniques, we obtain

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t Y_i(u)\frac{\lambda^*\{H_0(u)\}}{B_2(u)}\frac{V_{21}\{f_0,\dot{f}_0,H_0,\beta_0\}(W_i)}{e_{31}(W_i)}\mathrm{d}\lambda_\epsilon(H_0(u)+Z_i^\top\beta_0+f_0(W_i))$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t\frac{\lambda^*\{H_0(u)\}}{B_2(u)}\frac{Y_i(u)}{e_{31}(W_i)}\frac{1}{n}\sum_{j=1}^{n}\int_0^\tau K_h(W_j-W_i)[\mathrm{d}N_j(s)-Y_j(s)\,\mathrm{d}\Lambda_\epsilon\{H_0(s)$$

$$+Z_j^\top\beta_0+f_0(W_i)+\dot{f}_0(W_i)(W_j-W_i)\}]\,\mathrm{d}\lambda_\epsilon(H_0(u)+Z_i^\top\beta_0+f_0(W_i))$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau\frac{\int_0^t\frac{\lambda^*\{H_0(u)\}}{B_2(u)}E[Y(u)\dot{\lambda}_\epsilon\{H_0(u)+Z^\top\beta_0+f_0(W)\}\mid W=W_i]\mathrm{d}H_0(u)}{\int_0^\tau E[Y(u)\dot{\lambda}_\epsilon\{H_0(u)+Z^\top\beta_0+f_0(W)\}\mid W=W_i]\,\mathrm{d}H_0(u)}\mathrm{d}M_i(s)+o_p(1)$$

$$:=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau\tilde{m}_{Z_i^*}(t)\,\mathrm{d}M_i(s)+o_p(1),$$

where $\tilde{m}_{Z_i^*}(t)=\int_0^t(\lambda^*\{H_0(u)\}\,/\,B_2(u))E[Y(u)\dot{\lambda}_\epsilon\{H_0(u)+Z^\top\beta_0+f_0(W)\}\mid W=W_i]\,\mathrm{d}H_0(u)\,/\int_0^\tau E[Y(u)\dot{\lambda}_\epsilon\{H_0(u)+Z^\top\beta_0+f_0(W)\}\mid W=W_i]\,\mathrm{d}H_0(u)$. Combining these results, we can show the following result is true asymptotically:

$$\hat{S}_n(t)-\int_0^\tau p(t,s)\,\mathrm{d}\hat{S}_n(s)=W_n(t)=n^{-1/2}\sum_{i=1}^{n}w_i(t),$$

where $p(t,s)=\int_0^t(\lambda^*\{H_0(u)\}/B_2(u))c_2(u,s)\,\mathrm{d}H_0(u)$, and

$$w_i(t)=\int_0^t\frac{\lambda^*\{H_0(u)\}}{B_2(u)}\mathrm{d}\{c_1(u)\}(A_1-A_2)^{-1}\left\{\int_0^\tau\{[Z_i-m_Z(t)]-[Z_i^*-m_{Z_i^*}]\}\,\mathrm{d}M_i(t)\right\}$$

$$+\int_0^t\frac{\lambda^*\{H_0(u)\}}{B_2(u)}\mathrm{d}M_i(u)-\int_0^\tau\tilde{m}_{Z_i^*}(t)\,\mathrm{d}M_i(s),\quad i=1,\dots,n,$$

which are independent mean zero functions. Thus, by the functional central limit theorem, $W_n(t)$ converges weakly to a mean zero Gaussian process as $n\to\infty$. This completes the proof. ■

Proof of Theorem* 4.2*: We will now establish the asymptotic representation of $\sqrt{n}\{\hat{H}(t;\hat{\beta},\hat{\alpha}_0)-H_0(t)\}$, where $(\hat{\alpha}_0,\hat{\alpha}_1)$ are the solutions of Equation (7) at convergence. First, by using the Taylor series expansion, for any $t\in[0,\tau]$, we have

$$\Lambda^*\{\hat{H}(t;\hat{\beta},\hat{\alpha}_0)\}$$

$$=\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\}+\lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\}\left(\left.\frac{\partial\hat{H}(t;\beta,\hat{\alpha}_0)}{\partial\beta}\right|_{\beta=\beta_0}\right)^\top(\hat{\beta}-\beta_0)+o_p(n^{-1/2})$$

$$=\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\}-\int_0^t\frac{B_1^Z(s)}{B_2(s)}\mathrm{d}\Lambda^*\{H_0(s)\}(\hat{\beta}-\beta_0)+o_p(n^{-1/2}),$$

where the last equality follows from Part 3 of Theorem 4.1.

By Lemma A.2, we proved that $\hat{S}_n(t)=\sqrt{n}[\Lambda^*\{\hat{H}(t;\beta_0,\hat{\alpha}_0)\}-\Lambda^*\{H_0(t)\}]$ satisfies the integral equation (A19) for any $t\in[0,\tau]$. Using integration by part, we can rewrite Equation (A19) as a Fredholm integral equation of the second kind with the kernel $\partial p(t,s)/\partial s$, i.e.

$$\hat{S}_n(t)+\int_0^\tau\hat{S}_n(s)\frac{\partial p(t,s)}{\partial s}\mathrm{d}s=W_n(t)+p(t,s)\hat{S}_n(s)\Big|_{s=0}^{\tau}.$$

The uniqueness of the solution to the integral equation (A19) can be guaranteed by the condition

$$\sup_{t\in[0,\tau]} \int_0^\tau \left| \frac{\partial p(t,s)}{\partial s} \right| \mathrm{d}s < \infty. \tag{A20}$$

Moreover, we can construct a solution to Equation (A19) as follows:

$$\hat{S}_n(t) = W_n(t) + \int_0^\tau r(t,s)\,\mathrm{d}W_n(s). \tag{A21}$$

By substituting Equation (A21) into Equation (A19), we can obtain that $r(t,s)$ is the solution to the following equation:

$$r(t,s) = p(t,s) + \int_0^\tau p(t,u)\frac{\partial r(u,s)}{\partial u}\mathrm{d}u, \quad t,s \in [0,\tau], \tag{A22}$$

which can be written as a Fredholm integral equation of the second kind with the kernel $\partial p(t,s)/\partial s$. Thus, given Equation (A20), Equation (A22) also has a unique solution, and $\hat{S}_n(t)$ defined in Equation (A21) is thus a solution to the integral equation (A19).

Based on the above derivations and Equation (A18), the asymptotic representation of $\sqrt{n}(\hat{\beta} - \beta_0)$ established in the proof of Theorem 4.1, we can obtain

$$\sqrt{n}[\Lambda^*\{\hat{H}(t;\hat{\beta},\hat{\alpha}_0)\} - \Lambda^*\{H_0(t)\}] = \frac{1}{\sqrt{n}}\sum_{i=1}^n \kappa_i(t) + o_p(1),$$

where

$$\begin{aligned}\kappa_i(t) = w_i(t) &+ \int_0^\tau r(t,s)\,\mathrm{d}w_i(s) - \int_0^t \frac{B_1^Z(s)}{B_2(s)}\mathrm{d}\Lambda^*\{H_0(s)\} \\ &\times (A_1 - A_2)^{-1}\left\{\int_0^\tau \{[Z_i - m_Z(t)] - [Z_i^* - m_{Z_i^*}]\}\,\mathrm{d}M_i(t)\right\} + o_p(1)\end{aligned} \tag{A23}$$

are independent mean zero functions for $i = 1,\ldots,n$. Thus we have

$$\sqrt{n}[\hat{H}(t;\hat{\beta},\hat{\alpha}_0) - H_0(t)] = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\kappa_i(t)}{\lambda^*\{H_0(t)\}} + o_p(1),$$

which can be shown to converge weakly to a mean zero Gaussian process by the functional central limit theorem (see Theorem 10.6 in Pollard 1990). Thus, the proof is complete. ■

Proof of Theorem* 4.3*: Our goal is to establish the asymptotic representations of $\sqrt{nh}(\hat{\alpha}_0(w) - f_0(w))$ and $\sqrt{nh}(h\hat{\alpha}_1(w) - h\dot{f}_0(w))$. Note that $V_2(\hat{\alpha}_0,\hat{\alpha}_1,\hat{H}(\cdot;\hat{\beta},\hat{\alpha}_0),\hat{\beta})(w) = 0$, $\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1)$ and $\sqrt{n}|\hat{H}(t;\hat{\beta},\hat{\alpha}_0) - H_0(t)| = O_p(1)$ for any $t \in [0,\tau]$. Thus, we can readily obtain $V_2(\hat{\alpha}_0,\hat{\alpha}_1,H_0,\beta_0)(w) = O_p(n^{-1/2}) = o_p(1/\sqrt{nh})$.

Denote $\alpha(w) = (\alpha_0(w), h\alpha_1(w))^\top, \hat{\alpha}(w) = (\hat{\alpha}_0(w), h\hat{\alpha}_1(w))^\top$ and $\tilde{f}(w) = (f_0(w), h\dot{f}_0(w))^\top$. By the Taylor series expansion, we have

$$\begin{aligned}V_2(\hat{\alpha}_0,\hat{\alpha}_1,H_0,\beta_0)(w) &= V_2(f_0,\dot{f}_0,H_0,\beta_0)(w) + \frac{\partial V_2(\alpha_0^*,\alpha_1^*,H_0,\beta_0)(w)}{\partial\alpha(w)}\{\hat{\alpha}(w) - \tilde{f}(w)\} \\ &= o_p\left(\frac{1}{\sqrt{nh}}\right),\end{aligned} \tag{A24}$$

where $\alpha^*(w) = (\alpha_0^*(w), h\alpha_1^*(w))$ lies between $\hat{\alpha}(w)$ and $\tilde{f}(w)$. Thus, we have $\alpha^*(w) \to \tilde{f}(w)$ in probability. Moreover, consider

$$\begin{aligned}
&\frac{\partial V_2(\alpha_0, \alpha_1, \hat{H}(\cdot; \hat{\beta}, \hat{\alpha}_0), \hat{\beta})(w)}{\partial \alpha(w)} \\
&\quad = -\frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} K_h(W_i - w) Y_i(t) \dot{\lambda}_\epsilon \{\hat{H}(\cdot; \hat{\beta}, \hat{\alpha}_0) + Z_i^\top \hat{\beta} + \alpha_0(w) + \alpha_1(w)(W_i - w)\} \\
&\quad\quad \times \begin{pmatrix} 1 \\ \frac{W_i - w}{h} \end{pmatrix} \begin{pmatrix} 1 & \frac{W_i - w}{h} \end{pmatrix} \mathrm{d}\hat{H}(\cdot; \hat{\beta}, \hat{\alpha}_0),
\end{aligned}$$

which is negative definite. By the strong law of large numbers and standard nonparametric techniques, we can show that $\partial V_2(\alpha_0, \alpha_1, \hat{H}(\cdot; \hat{\beta}, \hat{\alpha}_0), \hat{\beta})(w)/\partial \alpha(w)$ converges to $-\dot{v}_\alpha(\alpha_0, H_0, \beta_0)$, a deterministic negative definite matrix, where

$$\dot{v}_\alpha(\alpha_0, H_0, \beta_0) = g(w) \int_0^{\tau} E[Y(t) \dot{\lambda}_\epsilon \{H_0(t) + Z^\top \beta_0 + \alpha_0(w)\} \mid W = w] dH_0(t) \begin{pmatrix} 1 & 0 \\ 0 & k_2 \end{pmatrix}.$$

Let

$$\Sigma_1(w) = -\lim_{n \to \infty} \frac{\partial V_2(f_0, \dot{f}_0, H_0, \beta_0)(w)}{\partial \alpha(w)} = \dot{v}_\alpha(f_0, H_0, \beta_0).$$

By the definition of $M_i(t)$, we have

$$\begin{aligned}
&V_2(f_0, \dot{f}_0, H_0, \beta_0)(w) \\
&\quad = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} K_h(W_i - w) \begin{pmatrix} 1 \\ \frac{W_i - w}{h} \end{pmatrix} \mathrm{d}N_i(t) \\
&\quad\quad - \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} K_h(W_i - w) Y_i(t) \begin{pmatrix} 1 \\ \frac{W_i - w}{h} \end{pmatrix} \mathrm{d}\Lambda_\epsilon \{H_0(t) + Z_i^\top \beta_0 + f_0(w) + \dot{f}_0(w)(W_i - w)\} \\
&\quad = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} K_h(W_i - w) \begin{pmatrix} 1 \\ \frac{W_i - w}{h} \end{pmatrix} \mathrm{d}M_i(t) \\
&\quad\quad + \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} K_h(W_i - w) Y_i(t) \begin{pmatrix} 1 \\ \frac{W_i - w}{h} \end{pmatrix} \mathrm{d}[\Lambda_\epsilon \{H_0(t) + Z_i^\top \beta_0 + f_0(W_i)\} \\
&\quad\quad - \Lambda_\epsilon \{H_0(t) + Z_i^\top \beta_0 + f_0(w) + \dot{f}_0(w)(W_i - w)\}] \\
&\quad := C_1 + C_2,
\end{aligned} \tag{A25}$$

where

$$C_1 = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} K_h(W_i - w) \begin{pmatrix} 1 \\ \frac{W_i - w}{h} \end{pmatrix} \mathrm{d}M_i(t)$$

and

$$\begin{aligned}
C_2 = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} K_h(W_i - w) Y_i(t) \begin{pmatrix} 1 \\ \frac{W_i - w}{h} \end{pmatrix} &\mathrm{d}[\Lambda_\epsilon \{H_0(t) + Z_i^\top \beta_0 + f_0(W_i)\} \\
&- \Lambda_\epsilon \{H_0(t) + Z_i^\top \beta_0 + f_0(w) + \dot{f}_0(w)(W_i - w)\}].
\end{aligned}$$

Similar to the proof of Theorem 4 of Cai et al. (2007), using the central limit theorem, we obtain

$$(nh)^{1/2} C_1 \xrightarrow{\mathcal{D}} N\{0, \Sigma_2(w)\} \quad \text{as } n \to \infty,$$

where

$$\Sigma_2(w) = h\begin{pmatrix}1 & 0\\ 0 & k_2\end{pmatrix} g(w)E\left\{\int_0^\tau \mathrm{d}M(t) \mid W = w\right\}^2.$$

$$\begin{aligned}C_2 &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(W_i - w)Y_i(t)\begin{pmatrix}1\\ \dfrac{W_i - w}{h}\end{pmatrix}\dot{\lambda}_\epsilon\{H_0(t) + Z_i^\top\beta_0 + f_0(w)\}\\ &\quad\times\left[\ddot{f}_0(w)\frac{(W_i - w)^2}{2}\right]\mathrm{d}H_0(t) + o_p(h^2)\\ &= \frac{h^2}{2}\ddot{f}_0(w)g(w)\begin{pmatrix}k_2\\ 0\end{pmatrix}\int_0^\tau E[Y(t)\dot{\lambda}_\epsilon\{H_0(t) + Z^\top\beta_0 + f_0(w)\} \mid W = w]\,\mathrm{d}H_0(t) + o_p(h^2)\\ &:= \Sigma_1(w)b_n(w) + o_p(h^2),\end{aligned}$$

where

$$b_n(w) = \frac{h^2}{2}\ddot{f}_0(w)g(w)\Sigma_1^{-1}(w)\begin{pmatrix}k_2\\ 0\end{pmatrix}\int_0^\tau E[Y(t)\dot{\lambda}_\epsilon\{H_0(t) + Z^\top\beta_0 + f_0(w)\} \mid W = w]\,\mathrm{d}H_0(t).$$

Combining the above derivations and Equations (A24) and (A25), we have

$$\Sigma_1(w)(nh)^{1/2}\{[\hat{\alpha}(w) - \tilde{f}(w)] - b_n(w) + o_p(h^2)\} = (nh)^{1/2}C_1.$$

Hence $(nh)^{1/2}\{[\hat{\alpha}(w) - \tilde{f}(w)] - b_n(w)\}$ weakly converges to a mean zero Gaussian Process with covariance matrix $\Sigma_1^{-1}(w)\Sigma_2(w)\Sigma_1^{-1}(w)$. This completes the proof. ■