Electronic Journal of Statistics
Vol. 18 (2024) 77–118
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2195

Regression analysis of partially linear transformed mean residual life models

Haijin He

School of Mathematical Sciences, Shenzhen University, Shenzhen, China
e-mail: hehj@szu.edu.cn

Jingheng Cai

Department of Statistics, Sun Yat-sen University, Guangzhou, China
e-mail: caijheng@mail.sysu.edu.cn

Xinyuan Song*

The Chinese University of Hong Kong, Hong Kong, China
e-mail: xysong@sta.cuhk.edu.hk

Abstract: We propose a novel class of partially linear transformed mean residual life (TMRL) models to investigate linear and nonlinear covariate effects on survival outcomes of interest. A martingale-based estimating equation approach with global and kernel-weighted local estimating equations is developed to estimate the parametric and nonparametric components. Unlike the existing inverse probability of censoring weighting estimating equation approach on TMRL models, the newly proposed method avoids estimating or modeling the distribution of the censoring time, thereby enhancing model capability and computational efficiency. Furthermore, we establish the asymptotic properties for the estimators of parametric and nonparametric components and develop an efficient iterative algorithm to implement the proposed procedure. Simulation studies demonstrate the satisfactory finite sample performance of the proposed method. Finally, our model is applied to the studies of lung cancer and type 2 diabetic complications.

Keywords and phrases: Estimating equation, local polynomial, martingale, partially linear model, TMRL model.

Received October 2022.

1. Introduction

The mean residual life (MRL) function is defined as $E(T-t|T>t)$ $(t \geq 0)$ for a nonnegative survival time T. It provides the remaining life expectancy of a subject surviving up to time t. As a valuable alternative to hazard function, the MRL function has been widely applied in biomedical sciences, actuarial studies, industrial reliability research, and other disciplines. Interested readers can refer to [35] for a detailed discussion of the MRL function.

*Corresponding Author.

Much research efforts have been devoted to the MRL regression analysis. Oakes and Dasu [31] proposed the proportional mean residual life (PMRL) model for dichotomous covariates. Maguluri and Zhang [28] extended the PMRL model to accommodate continuous covariates without censoring and developed an estimation procedure based on the hazards of the forward recurrence times in the renewal processes formed by T. Chen et al. [6] and Chen and Cheng [8] further studied the PMRL model in the presence of censoring and proposed an inverse probability of censoring weighting (IPCW) approach and a quasi partial score estimating equation method, respectively. Chen and Cheng [9] and Chen [7] proposed the additive mean residual life (AMRL) model and discussed various estimation procedures with or without censoring. Sun and Zhang [35] further proposed a general class of transformed mean residual life (TMRL) models that subsumes the PMRL and AMRL models as exceptional cases. They also developed an IPCW approach and its variant, namely, an augmented IPCW (AIPCW) approach, both requiring estimating or modeling the distribution of censoring. Sun, Song and Zhang [34] investigated the class of TMRL models with time-dependent coefficients, again under the IPCW framework. Mansourvar, Martinussen and Scheike [29] considered semiparametric regression for the restricted MRL model under right censoring. Mansourvar, Martinussen and Scheike [30] and Cai et al. [3] proposed different additive-multiplicative restricted MRL models and developed martingale estimating equation methods for estimation. He et al. [16, 17] explored the PMRL and AMRL models with latent variables through the corrected estimating equation approach.

Despite the valuable developments in MRL modeling above, they all assumed that all covariates linearly affect the MRL of interest. In some situations, however, such a linearity assumption may be unrealistic; some covariates might have nonlinear effects on the MRL of study subjects. For instance, in the Veteran's Administration lung cancer trial study of [20], patients' age nonlinearly affects their survival time [25]. Thus, the MRL analysis assuming linear effects of age might be erroneous or incomprehensive. Therefore, there is a need to consider nonlinear covariate effects in MRL regression from theoretical and practical viewpoints. Partially linear models originally proposed by Engle [12] enjoy the flexibility of modeling nonlinear covariate effects and share the parsimony and interpretability of ordinary regression models by allowing some covariates to have linear effects. Partially linear models have been widely studied in multivariate analysis [4, 5, 11, 15, 21, 22, 33] and hazard regression [1, 14, 18, 23, 25, 27]. However, no existing studies have investigated partially linear models in the MRL context. Hence, to fill this research gap and provide a comprehensive modeling framework, this study considers a novel class of partially linear TMRL models to investigate linear and nonlinear covariate effects on survival outcomes.

The inference procedures for partially linear models in the multivariate analysis have been systematically established; for example, see the monograph by Härdle et al. [15]. In particular, Gray [14] proposed the penalized partial likelihood with B-splines for the partially linear proportional hazards model. Huang [18] investigated the asymptotic properties of the partial likelihood estimators using polynomial splines for the same model. Cai et al. [1] proposed a local and profile pseudo partial likelihood method for the partially linear proportional hazards model with multivariate failure time data. Lu and McMahan [23] used monotone splines to approximate the baseline cumulative hazard function and adopted B-splines to accommodate nonlinear covariate effects in the partially linear proportional hazards model with current status data. In the context of transformation models, Ma and Kosorok [27] proposed a nonparametric maximum penalized log-likelihood method for the partially linear transformation models with current status data. Lu and Zhang [25] developed martingale-based global and local estimating equations for the partially linear transformation models. Nonetheless, the prior studies never considered the inference of partially linear TMRL models.

The TMRL models have distinctive features compared with hazard models; there are no analogies of partial likelihood or nonparametric likelihood in the field of MRL regression analysis. Therefore, the various methodologies developed for partially linear hazard models are not directly applicable to the proposed partially linear TRML models. This paper develops martingale-based global and local estimating equations to overcome the difficulties. The global equations are used to estimate the baseline function and unknown parameters indicating linear effects, whereas the local equations are adopted to estimate the nonlinear covariate effects. We develop an iterative algorithm to implement the proposed estimation procedure. We establish the root-n consistency of the estimator for linear parameters under suitable regularity conditions and proper choices of the kernel bandwidth. We also obtain the asymptotic normality for estimating the linear and nonlinear effects. Finally, we propose an easy-to-implement resampling method to estimate the asymptotic variances of the estimated linear effects.

The contributions of this study are three-fold. First, we consider a novel class of partially linear TMRL models. The proposed models are general in the following senses: (i) both linear and nonlinear covariate effects are considered, and (ii) the TMRL models are general and encompass PMRL, AMRL, and the Box-Cox transformation MRL models as exceptional cases. This general model class has never been investigated in the literature. Second, we develop a novel martingale-based global and local estimating equation method to overcome the theoretical and computational challenges incurred by the distinctive structure of the proposed model. Notably, the proposed approach only assumes conditional independence between the survival and censoring times given covariates. In contrast, the IPCW approach for linear TMRL models [35] either assumes independence between the censoring time and covariates and thus requires the estimation of the censoring distribution or modeling how the censoring time depends on covariates. Thus, the proposed approach provides a convenient alternative to the IPCW approach of [35] in that it avoids estimating or modeling the distribution of censoring time. Such an appealing feature of the proposed approach is due to the use of martingale representations. Even without accommodating partially linear covariate effects, the proposed approach is new to the inference of the TMRL models. Third, we develop an iterative algorithm to implement the proposed estimation procedure and show the asymptotic normality for the estimators of the linear and nonlinear effects.

The rest of the article is structured as follows. Section 2 describes the proposed partially linear TMRL models, presents the proposed martingale-based global and local estimating equations, and provides an iterative algorithm for implementation. Section 3 establishes the asymptotic properties of the proposed estimators. Section 4 reports simulation studies for evaluating the finite sample performance of the proposed method. Section 5 presents applications to two real-life datasets, one from the Veteran's Administration lung cancer trial and the other from chronic kidney disease (CKD) study of type 2 diabetic patients, to demonstrate the utility of the proposed methodology. Section 6 provides some concluding remarks. All technical proofs are relegated to the appendices.

2. Model and method

Let T_i $(i = 1, \ldots, n)$ be the failure time of interest. The MRL function of T_i is $m_i(t) = E[T_i - t|T_i > t]$, with $m_i(t) = 0$ whenever $P(T_i > t) = 0$. Let $\mathbf{Z}_i$ be a $p \times 1$ vector of covariates, and X_i be a uni-variate covariate which has nonlinear effect. We consider the partially linear TMRL models as follows:

$$m(t|\mathbf{Z}_i, X_i) = g\big\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + f(X_i)\big\}, \tag{2.1}$$

where $m(t|\mathbf{Z}_i, X_i)$ is the MRL function of subject i, $g(\cdot)$ is a prespecified transfomation or link function, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters, $m_0(t)$ is an unknown baseline function, and $f(x)$ is an unknown smooth function. For a constant a, $(m_0(t) + a, \boldsymbol{\beta}, f(x) - a)$ and $(m_0(t), \boldsymbol{\beta}, f(x))$ represent the same model. Thus, a restriction $f(0) = 0$ is set to ensure the model identifiability. Notably, $g\{m_0(t)\}$ is the baseline MRL function. Without the presence of the nonparametric component, Model (2.1) reduces to the TMRL models of [35]. Moreover, Model (2.1) defines a rich family of models through the link function $g(\cdot)$. It becomes the partially linear AMRL model if $g(u) = u$, the partially linear PMRL model if $g(u) = \exp(u)$, and encompasses the partially linear Box-Cox transformation MRL models with $g(u) = [(u+1)^q - 1]/q$, where $q = 0$ means that $g(u) = \log(u + 1)$. To our best knowledge, none of the preceding three special cases of (2.1) has been investigated in the literature.

In practice, the transformation function $g(\cdot)$ should be properly chosen to fit the data or to achieve desired interpretation of the regression coefficients. Theoretical restrictions for $g(\cdot)$ are as follows: (i) $g(\cdot)$ is twice continuously differentiable, (ii) $g(\cdot)$ is strictly increasing, and (iii) $g\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + f(X_i)\}$ is a proper MRL function for all possible values of $\mathbf{Z}_i$ and X_i. Model (2.1) may prove helpful in model selection if we insist that $g(\cdot)$ lies in the Box-Cox transformation family but allow q to be estimated. For more discussion on the transformation function $g(\cdot)$, see [35].

In model (2.1), we only consider one covariate in the nonlinear component for simplicity. However, extending the proposed model to accommodate multiple nonlinear covariates is straightforward and can be implemented as in [4]. Another practical issue is deciding which covaraites should be considered in the nonlinear part. To address this issue, we may include the covariates of interest into the model one at a time as the nonparametric component and check whether they have nonlinear effects.

Let C_i be the censoring time, $\tilde{T}_i = \min\{T_i, C_i\}$ be the observed time, $\Delta_i = I(T_i \leq C_i)$ denote the censoring indicator, $N_i(t) = I(\tilde{T}_i \leq t, \Delta_i = 1)$ stand for the counting process recording the number of events that have occurred by time t, and $Y_i(t) = I(\tilde{T}_i \geq t)$ be the at-risk process. The observed data consist of independent copies $(\tilde{T}_i, \Delta_i, \mathbf{Z}_i, X_i)$ $(i = 1, \ldots, n)$.

Let $\Lambda(t|\mathbf{Z}_i, X_i)$ be the cumulative hazard function of T_i. Using the inversion formula (e.g., [31]), we have

$$d\Lambda(t|\mathbf{Z}_i, X_i) = \frac{d(m(t|\mathbf{Z}_i, X_i) + t)}{m(t|\mathbf{Z}_i, X_i)} = \frac{d(g\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + f(X_i)\} + t)}{g\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + f(X_i)\}}.$$

Define

$$dM_i(t; m_0, \boldsymbol{\beta}, f) = dN_i(t) - Y_i(t)\frac{d(g\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + f(X_i)\} + t)}{g\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + f(X_i)\}}. \tag{2.2}$$

By definition, $M_i(t; m_0, \boldsymbol{\beta}, f)$ $(i = 1, \ldots, n)$ are martingales under the true model indexed by $m_0(t)$, $\boldsymbol{\beta}$, and $f(\cdot)$.

For fixed $f(\cdot)$, we construct the following global estimating equations for $m_0(t)$ and β as

$$\sum_{i=1}^{n} g\big\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + f(X_i)\big\} dM_i(t; m_0, \boldsymbol{\beta}, f) = 0, \tag{2.3}$$

and

$$\sum_{i=1}^{n} \int_0^\tau \mathbf{Z}_i g\big\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + f(X_i)\big\} dM_i(t; m_0, \boldsymbol{\beta}, f) = 0, \tag{2.4}$$

where τ is the end of the study. Notably, the proposed estimating equations (2.3) and (2.4) are novel for linear TMRL models, as they are different from the IPCW-based estimating equations of [35]. The rationale for constructing (2.3) and (2.4) are as follows. A common practice of survival analysis suggests using estimating equations

$$\sum_{i=1}^{n} dM_i(t; m_0, \boldsymbol{\beta}, f) = 0 \tag{2.5}$$

and

$$\sum_{i=1}^{n} \int_0^\tau \mathbf{Z}_i dM_i(t; m_0, \boldsymbol{\beta}, f) = 0 \tag{2.6}$$

to estimate $m_0(t)$ and β with $f(\cdot)$ fixed. However, due to the particular structure of $M_i(t; m_0, \boldsymbol{\beta}, f)$ in the proposed model (see (2.2)), i.e., the unknown parameters and functions are involved in the denominator, and the estimating equations (2.5) and (2.6) are computationally infeasible. Chen and Cheng

[8] and Chen and Cheng [9] proposed estimating equations different from (2.5) and (2.6) for PMRL and AMRL models to address the difficulty. The proposed estimating equations in the present study can be regarded as an extension of Chen and Cheng [8] and Chen and Cheng [9] to TMRL models. Multiplying by the weight $g\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + f(X_i)\}$ on both sides of (2.5) and (2.6) can eliminate the unknown parameters and functions in the denominator of (2.2), thereby enabling a feasible estimation of $m_0(t)$ and $\boldsymbol{\beta}$. Furthermore, unlike the existing IPCW approach that requires estimating or modeling the distribution of the censoring time, the martingale-based estimating equations (2.3) and (2.4) only assume independence between the survival and censoring times given covariates, simplifying the estimation procedure and enhancing model capability and computational efficiency.

Note that $f(x) \approx \gamma_0(u) + \gamma_1(u)(x-u)$ for x in a neighborhood of u, where $\gamma_0(u) = f(u)$, and $\gamma_1(u) = \dot{f}(u)$ is the first order derivative of $f(u)$. Let the kernel $K(x)$ be a symmetric probability density function and $K_h(x) = K(x/h)/h$, with $h > 0$ as the bandwidth parameter. Some commonly used kernel functions, such as the Gaussian kernel function and Epanechnikov kernel function [1], can be adopted. Since the choice of the kernel function is not crucial, we employ the Gaussian kernel function in numerical studies. For fixed $m_0(t)$ and $\boldsymbol{\beta}$, we propose the following local estimating equation for $f(\cdot)$ (i.e., γ_0 and γ_1) as

$$\sum_{i=1}^{n}\int_0^{\tau}(1, X_i - x)^T K_h(X_i - x) \times g\big\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + \gamma_0(x) + \gamma_1(x)(X_i - x)\big\} d\tilde{M}_i(t; m_0, \boldsymbol{\beta}, x) = 0, \quad (2.7)$$

where

$$d\tilde{M}_i(t; m_0, \boldsymbol{\beta}, x) = dN_i(t) - Y_i(t)\frac{d(g\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + \gamma_0(x) + \gamma_1(x)(X_i - x)\} + t)}{g\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + \gamma_0(x) + \gamma_1(x)(X_i - x)\}}.$$

$\tilde{M}_i(t; m_0, \beta, x)$ is unnecessarily a martingale but is still approximately mean-zero. The validity of (2.7) shares the same spirit with the local estimating equations of Carroll, Ruppert and Welsh [5], which employs the approximate mean-zero property. See also the local estimating equations of Lu and Zhang [25].

The MRL function may not be estimable at the upper tail of the survival distribution due to right censoring. To avoid such non-identifiability, Chen and Cheng [6, 9] and Sun and Zhang [35] imposed an assumption that the support of the censoring time is longer than that of the survival time, while Mansourvar, Martinussen and Scheike [30] considered the restricted MRL function. In this article, we adopt the assumption of [6, 9, 35], see condition (C3) in Section 3. The restricted MRL function of [30] can also be considered without difficulties.

To enhance the implementation of the proposed method, we develop an algorithm to estimate $m_0(t)$, $\boldsymbol{\beta}$, and $f(\cdot)$ in a recursive manner.

Computing algorithm Let $t_1 < t_2 < \cdots < t_K$ be the K distinct observed failure times, and $\hat{f}^{(0)}(\cdot)$ and $\hat{\boldsymbol{\beta}}^{(0)}$ be the initial value of $f(\cdot)$ and $\boldsymbol{\beta}$, respectively.

The computing algorithm consists of three steps (a)–(c). In step (a), we solve for $m_0(\cdot)$ at the observed failure times and $t = 0$ based on the current values of $\boldsymbol{\beta}$ and $f(\cdot)$. In step (b), we update $\boldsymbol{\beta}$ based on the current value of $m_0(\cdot)$ and $f(\cdot)$. In step (c), we update $\gamma_0(\cdot)$ and $\gamma_1(\cdot)$ at the observed covariates X_j $(j = 1, \cdots, n)$ based on the current value of $m_0(\cdot)$ and $\boldsymbol{\beta}$. We employed the R function "dfsane" to solve for the roots of the corresponding equations. Specifically, at the ath iteration, we implement three steps (a)–(c) as follows:

Step (a). Update $\hat{m}_0^{(a)}(t_k)$ by solving

$$
\begin{aligned}
0 = &\sum_{i=1}^{n} I(\tilde{T}_i > t_k) g\big\{\hat{m}_0^{(a)}(t_k) + \mathbf{Z}_i^T \hat{\boldsymbol{\beta}}^{(a-1)} + \hat{f}^{(a-1)}(X_i)\big\} \\
&- \sum_{i=1}^{n} I(\tilde{T}_i \geq t_k) g\big\{\hat{m}_0^{(a)}(t_{k-1}) + \mathbf{Z}_i^T \hat{\boldsymbol{\beta}}^{(a-1)} + \hat{f}^{(a-1)}(X_i)\big\} \\
&+ \sum_{i=1}^{n} I(\tilde{T}_i \geq t_k)(t_k - t_{k-1}) + \sum_{i=1}^{n} I(t_{k-1} < \tilde{T}_i < t_k)(\tilde{T}_i - t_{k-1}).
\end{aligned} \tag{2.8}
$$

Step (b). Update $\hat{\boldsymbol{\beta}}^{(a-1)}$ to $\hat{\boldsymbol{\beta}}^{(a)}$ by solving

$$
\begin{aligned}
0 = &\sum_{i=1}^{n} (\Delta - 1) \mathbf{Z}_i g\big\{\hat{m}_0^{(a)}(\tilde{T}_i) + \boldsymbol{\beta}^T \mathbf{Z}_i + \hat{f}^{(a-1)}(X_i)\big\} \\
&+ \sum_{i=1}^{n} \mathbf{Z}_i g\big\{\hat{m}_0^{(a)}(0) + \boldsymbol{\beta}^T \mathbf{Z}_i + \hat{f}^{(a-1)}(X_i)\big\} - \sum_{i=1}^{n} \mathbf{Z}_i \tilde{T}_i.
\end{aligned} \tag{2.9}
$$

Step (c). Update $\hat{\gamma}_0^{(a-1)}(X_j)$ and $\hat{\gamma}_1^{(a-1)}(X_j)$ to $\hat{\gamma}_0^{(a)}(X_j)$ and $\hat{\gamma}_1^{(a)}(X_j)$ $(j = 1, 2, \ldots, n)$ by solving

$$
\begin{aligned}
0 = &\sum_{i=1}^{n} (\Delta_i - 1)(1, X_i - X_j)^T K_h(X_i - X_j) g \\
&\times \big\{\hat{m}_0^{(a)}(\tilde{T}_i) + \mathbf{Z}_i^T \hat{\boldsymbol{\beta}}^{(a)} + \gamma_0(X_j) + \gamma_1(X_j)(X_i - X_j)\big\} \\
&+ \sum_{i=1}^{n} (1, X_i - X_j)^T K_h(X_i - X_j) \\
&\times g\big\{\hat{m}_0^{(a)}(0) + \mathbf{Z}_i^T \hat{\boldsymbol{\beta}}^{(a)} + \gamma_0(X_j) + \gamma_1(X_j)(X_i - X_j)\big\} \\
&- \sum_{i=1}^{n} (1, X_i - X_j)^T K_h(X_i - X_j) \tilde{T}_i.
\end{aligned} \tag{2.10}
$$

When returning from step (c) to (a), we set $\hat{f}^{(a)}(X_i) = \hat{\gamma}_0^{(a)}(X_i)$. In step (a), we obtain $\hat{m}_0^{(a)}(t_k)$ in a backward fashion; that is, we first obtain $\hat{m}_0^{(a)}(t_{K-1})$ from $\hat{m}_0^{(a)}(t_K) = 0$, then obtain $\hat{m}_0^{(a)}(t_{K-2})$ from $\hat{m}_0^{(a)}(t_{K-1})$, and finally obtain

$\hat{m}_0^{(a)}(0)$ from $\hat{m}_0^{(a)}(t_1)$. Steps (a), (b), and (c) are repeated until convergence. The estimates of $m_0(t)$ and $\boldsymbol{\beta}$ at convergence are denoted as $\hat{m}_0(t)$ and $\hat{\boldsymbol{\beta}}$. In the above updating, (2.8) is obtained by taking integration over the interval $(t_{k-1}, t_k]$ ($k = 1, \ldots, K$, $t_0 = 0$) on both sides of (2.3) and using some simple algebraic manipulations. Moreover, from (2.4) we have

$$\int_0^\tau g\big\{m_0(t) + \boldsymbol{\beta}^T \mathbf{Z}_i + f(X_i)\big\} dN_i(t) = g\big\{m_0(\tilde{T}_i) + \boldsymbol{\beta}^T \mathbf{Z}_i + f(X_i)\big\}\Delta_i,$$

and

$$\begin{aligned}\int_0^\tau & \mathbf{Z}_i Y_i(t) d\big(g\big\{m_0(t) + \boldsymbol{\beta}^T \mathbf{Z}_i + f(X_i)\big\} + t\big) \\ &= \mathbf{Z}_i\big[g\big\{m_0(\tilde{T}_i) + \boldsymbol{\beta}^T \mathbf{Z}_i + f(X_i)\big\} - g\big\{m_0(0) + \boldsymbol{\beta}^T \mathbf{Z}_i + f(X_i)\big\}\big] + \mathbf{Z}_i \tilde{T}_i,\end{aligned}$$

and there are similar expressions for (2.7). Then, we can obtain (2.9) and (2.10) from (2.4) and (2.7), respectively, after some simple manipulations.

Step (d). The final estimate of $f(x)$ is obtained by solving Equation (2.10) with $\hat{m}_0^{(a)}(t)$ and $\hat{\boldsymbol{\beta}}^{(a)}$ replaced by $\hat{m}_0(t)$ and $\hat{\boldsymbol{\beta}}$, respectively.

The asymptotic properties of the estimators of $m_0(t)$, $\boldsymbol{\beta}$, and $f(\cdot)$ are presented in Section 3 with proofs in Appendix A. In Appendix B, we propose a one-step estimator as the initial value $\hat{f}^{(0)}(\cdot)$ and show its local consistency. Practically, one can employ a parametric form of $f(\cdot)$ and estimate it using the proposed global estimating equations for computational convenience.

In implementing the computing algorithm, we must select the bandwidth parameter h in steps (a)–(d). Notably, h plays different roles in different steps. In steps (a)–(c), the selection of h should ensure the proper estimation of $m_0(t)$ and $\boldsymbol{\beta}$, whereas in step (d), h should be optimal for estimating the nonparametric function $f(\cdot)$. Therefore, the value of h in steps (a)–(c) is different from that in step (d). In the simulation section, we provide more discussion on the selection of h based on the theoretical convergence rate or some data-adaptive tuning criteria.

3. Asymptotic properties

Let $m_*(t)$ and $f_*(x)$ be the true values of the functions $m_0(t)$ and $f(x)$, respectively, and $\boldsymbol{\beta}_0$ be the true value of $\boldsymbol{\beta}$. To investigate the asymptotic properties of the proposed estimators $\hat{m}_0(t)$, $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}_0(x)$, and $\hat{\gamma}_1(x)$, we define

$$\begin{aligned}&dQ_i(t; m_0, \boldsymbol{\beta}, f) \\ &\quad = \dot{g}\big\{m_0(t) + \boldsymbol{\beta}^T \mathbf{Z}_i + f(X_i)\big\} dN_i(t) - Y_i(t) d\dot{g}\big\{m_0(t) + \boldsymbol{\beta}^T \mathbf{Z}_i + f(X_i)\big\},\end{aligned}$$

where $\dot{g}(u)$ denotes the derivative of $g(u)$. For simplicity, we denote $dQ_i(t) = dQ_i(t; m_*, \boldsymbol{\beta}_0, f_*)$. In the following, we omit the subscripts representing subjects when taking expectations for convenience of presentation. Define

$$B(t,s) = \exp\left(-\int_s^t \frac{E[dQ(u)]}{E[Y(u)\dot{g}\{m_*(u)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)\}]}\right),$$

$$\boldsymbol{\mu}_1(t) = \frac{E[(\Delta-1)\mathbf{Z}\dot{g}\{m_*(\tilde{T})+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)\}B(t,\tilde{T})I(\tilde{T}<t)]}{E[Y(t)\dot{g}\{m_*(t)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)\}]},$$

$$\boldsymbol{\mu}_2(t) = \frac{E[\mathbf{Z}\dot{g}\{m_*(0)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)\}B(t,0)]}{E[Y(t)\dot{g}\{m_*(t)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)\}]}.$$

Set $v\{m_*(t)\} = B(t,\tau)$, $\mathbf{A}_1 = E[\int_0^\tau \{\mathbf{Z}-\boldsymbol{\mu}_1(t)-\boldsymbol{\mu}_2(t)\}\mathbf{Z}^T dQ(t)]$,

$$\mathbf{A}_2 = E\left[\int_0^\tau \big\{\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}}(t)\big\}\boldsymbol{\rho}(X)^T dQ(t)\right],$$

$$\boldsymbol{\Sigma} = E\left[\int_0^\tau \big\{\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}}(t)-\big(\mathbf{Z}^*-\boldsymbol{\mu}_{\mathbf{Z}^*}\big)\big\}^{\otimes 2} g\big\{m_*(t)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)\big\}^2 dN(t)\right],$$

where $\mathbf{Z}^{\otimes 2} = \mathbf{Z}\mathbf{Z}^T$ for any vector $\mathbf{Z}$, and

$$\boldsymbol{\mu}_{\mathbf{Z}}(t) = \frac{\boldsymbol{\alpha}(t)v\{m_*(t)\}}{E[Y(t)\dot{g}\{m_*(t)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)\}]},$$

$$\mathbf{Z}_i^* = \int_0^\tau \frac{E[\mathbf{Z}dQ(t)|X=X_i]}{E[(\Delta-1)\dot{g}\{m_*(\tilde{T})+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)+\dot{g}\{m_*(0)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)|X=X_i]},$$

$$\boldsymbol{\mu}_{\mathbf{Z}^*,i} = \int_0^\tau \frac{\boldsymbol{\mu}_{\mathbf{Z}}(t)E[dQ(t)|X=X_i]}{E[(\Delta-1)\dot{g}\{m_*(\tilde{T})+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)+\dot{g}\{m_*(0)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)|X=X_i]}.$$

In the definition of $\boldsymbol{\mu}_{\mathbf{Z}}(t)$, $\boldsymbol{\alpha}(t)$ is the solution to the following Fredholm integral equation of the second kind

$$\boldsymbol{\alpha}(t) - \int_0^\tau \boldsymbol{\alpha}(s)D_1(t,ds) = \mathbf{D}_2(t), \quad t\in[0,\tau] \tag{3.1}$$

with $D_1(t,ds)$, $\mathbf{D}_2(t)$, and $\boldsymbol{\rho}(\cdot)$ provided in Appendix A.

We need the following regularity conditions:

(C1) The covariates $\mathbf{Z}_i$ and X_i are of compact support, and the density of X_i, denoted by $r(x)$, has a bounded second derivative.
(C2) $\boldsymbol{\beta}_0$ belongs to the interior of a known compact set, $m_*(t)$ is continuously differentiable, $f_*(x)$ has a continuous second derivative, $\dot{g}(\cdot)$ is continuous, positive and bounded away from zero.
(C3) The support of the censoring time C_i is longer than that of the survival time T_i.
(C4) τ is finite and $P(T_i \geq \tau) > 0$ and $P(C_i \geq \tau) > 0$.
(C5) The kernel $D_1(\cdot,\cdot)$ satisfies $\sup_{t\in[0,\tau]}\int_0^\tau |D_1(t,ds)| < \infty$.
(C6) The matrices $\mathbf{A} = \mathbf{A}_1 - \mathbf{A}_2$ and $\boldsymbol{\Sigma}$ are nondegenerate.

Conditions (C1)–(C4) are similar to those in [35] for establishing asymptotic properties for TMRL models. Condition (C5) guarantees the uniqueness of the solution to the Fredholm integral equation (3.1). Condition (C6) is needed for establishing the asymptotic normality of the proposed estimators. Condition (C6) is mainly a technical condition to ensure the invertibility of $\mathbf{A}$ and the non-degeneracy of $\mathbf{\Sigma}$, so that the asymptotic representation of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ can be obtained and the asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ will not be degenerate. This condition usually holds when the covariates do not centered on a lower-dimensional subspace.

The following three theorems establish the asymptotic properties of $\hat{m}_0(t)$, $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}_0(x)$, and $\hat{\gamma}_1(x)$. The proofs are relegated to Appendix A.

Theorem 3.1. *Under the regularity conditions (C1)–(C6), if $nh^2/\{\log(1/h)\} \to \infty$ and $nh^4 \to 0$, then $\hat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$, and $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to N(\mathbf{0}, \boldsymbol{A}^{-1}\mathbf{\Sigma}(\boldsymbol{A}^{-1})^T)$ in distribution as $n \to \infty$.*

Theorem 3.2. *Under the regularity conditions (C1)–(C6), if $nh^2/\{\log(1/h)\} \to \infty$ and $nh^4 \to 0$, we have the following asymptotic representation*

$$n^{1/2}\big\{\hat{m}_0(t) - m_*(t)\big\} = n^{-1/2}\sum_{i=1}^{n} \frac{\kappa_i(t)}{v\{m_*(t)\}} + o_p(1),$$

for $t \in [0, \tau)$, where $\kappa_i(t)$'s are independent mean zero functions and their definitions are given in Appendix A.

Theorem 3.3. *Under the regularity conditions (C1)–(C4), if nh^5 is bounded, and $\boldsymbol{\beta}$ and $m_0(t)$ are estimated at the order $O_p(n^{-1/2})$, then for any x in the compact support of X_i, we have*

$$(nh)^{1/2}\left(\begin{bmatrix} \hat{\gamma}_0(x) - f_*(x) \\ h\{\hat{\gamma}_1(x) - \dot{f}_*(x)\}\end{bmatrix} - \boldsymbol{b}_n(x)\right) \to N\big(\mathbf{0}, \mathbf{\Omega}(x)\big)$$

in distribution, where $\mathbf{\Omega}(x) = \mathbf{\Omega}_1^{-1}(x)\mathbf{\Omega}_2(x)(\mathbf{\Omega}_1^{-1}(x))^T$, and the definitions of $\mathbf{\Omega}_1(x)$, $\mathbf{\Omega}_2(x)$, and $\boldsymbol{b}_n(x)$ are provided in Appendix A.

The asymptotic variance of $\hat{\boldsymbol{\beta}}$ has a standard sandwich form $\mathbf{A}^{-1}\mathbf{\Sigma}(\mathbf{A}^{-1})^T$. However, the matrices $\mathbf{A}$ and $\mathbf{\Sigma}$ are complicated, and their computation requires solving a Fredholm integral equation, which is often difficult and unstable even for moderate sample size. Therefore, we propose using a resampling method [19, 25] to approximate the asymptotic distribution of $\hat{\boldsymbol{\beta}}$.

First, we generate n i.i.d. exponential random variables $\{\xi_i, i = 1, \ldots, n\}$ with mean 1 and variance 1. Fixing the data at their observed values, we solve the following ξ_i-weighted estimating equations

$$\sum_{i=1}^{n} \xi_i g\big\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + f(X_i)\big\}dM_i(t; m_0, \boldsymbol{\beta}, f) = 0, \tag{3.2}$$

$$\sum_{i=1}^{n} \xi_i \int_0^{\tau} \mathbf{Z}_i g\big\{m_0(t) + \boldsymbol{\beta}^T\mathbf{Z}_i + f(X_i)\big\}dM_i(t; m_0, \boldsymbol{\beta}, f) = 0, \tag{3.3}$$

$$\sum_{i=1}^{n} \xi_i \int_0^{\tau} (1, X_i - x)^T K_h(X_i - x) g\big\{m_0(t) + \boldsymbol{\beta}^T \mathbf{Z}_i + \gamma_0(x) + \gamma_1(x)(X_i - x)\big\}$$
$$d\tilde{M}_i(t; m_0, \boldsymbol{\beta}, x) = 0. \tag{3.4}$$

Equations (3.2), (3.3), and (3.4) can be solved using the same recursive algorithm as in Section 2. Denote the solutions as $\hat{\boldsymbol{\beta}}^*$, $\hat{m}_0^*(t)$, and $\hat{f}^*(x)$.

Theorem 3.4. *Under the regularity conditions (C1)–(C6), if $nh^2/\{\log(1/h)\} \to \infty$ and $nh^4 \to 0$, then the conditional distribution of $n^{1/2}(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}})$ given the observed data converges to the asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$*

By repeatedly generating $(\xi_1, \ldots, \xi_n)$ many times, we may obtain a large number of realizations of $\hat{\boldsymbol{\beta}}^*$. The asymptotic variance of $\hat{\boldsymbol{\beta}}$ can be estimated by the empirical variance of $\hat{\boldsymbol{\beta}}^*$.

4. Simulation

We conduct simulations to assess the finite sample performance of the proposed method. The survival times T_i's are generated from the partially linear TMRL models (2.1). Two independent covariates $\mathbf{Z} = (Z_{i1}, Z_{i2})^T$ are considered, where $Z_{i1} \sim \text{Bernoulli}(0.5)$ and $Z_{i2} \sim \text{Uniform}[-0.5, 0.5]$. The regression coefficients are taken as $\boldsymbol{\beta} = (\beta_1, \beta_2)^T = (-1, 1)^T$. For the nonlinear effect, we let $f(x) = 3(x - x^3)$ and $X_i \sim \text{Uniform}[0, 1]$, where X_i is independent of $\mathbf{Z}_i$. Following the practice of [35], we consider several transformation functions $g(\cdot)$'s, but only report the results for $g(u) = u$ and $g(u) = \exp(u)$, which correspond to the partially linear AMRL and PMRL models, respectively. The baseline MRL function $g\{m_0(t)\}$ is taken from the Hall-Wellner family [31], such that $g\{m_0(t)\} = (D_1 t + D_2)^+$, where $D_1 > -1$, $D_2 > 0$, and $d^+ = dI(d \geq 0)$ for any d. We set $D_1 = -1/6$ and $D_2 = 1.5$ for the partially linear PMRL model and $D_1 = -0.9$ and $D_2 = 1.5$ for the partially linear AMRL model. For generating T_i's, we use the formula

$$S(t|\mathbf{Z}_i, X_i) = \frac{m(0|\mathbf{Z}_i, X_i)}{m(t|\mathbf{Z}_i, X_i)} \exp\bigg\{-\int_0^t \frac{du}{m(u|\mathbf{Z}_i, X_i)}\bigg\},$$

where $S(t|\mathbf{Z}_i, X_i)$ is the survivor function of subject i, and $m(t|\mathbf{Z}_i, X_i)$ is given by (2.1).

Two censoring schemes (CS) are considered: (i) Covariate-independent censoring. The censoring times C_i's are generated from $\text{Uniform}(0, c_0)$, where c_0 is chosen to achieve a censoring rate (CR) of 10% and 30%. (ii) Covariate-dependent censoring. The censoring times C_i's are generated from exponential distributions with means $\exp(c_0 + c_1 Z_{i1})$, where c_0 and c_1 are chosen to achieve a CR of 10% and 30%. In most scenarios considered, the support of C_i is longer than the support of T_i for all $\mathbf{Z}_i$ and X_i. For a few exceptional cases violating this condition, the proposed method and algorithm still work and produce fairly good estimation results.

To start the computing algorithm, we set the initial value of $f(\cdot)$ as $\hat{f}^{(0)}(\cdot) \equiv 0$. We further discuss the selection on the bandwidth parameter h. In Theorems 3.1 and 3.2, a condition on h is provided to ensure the convergence of the estimates of $m_0(t)$ and $\boldsymbol{\beta}$ derived from the global equations (2.3) and (2.4). In Theorem 3.3, another condition on h is given to ensure the convergence of the estimates of $f(\cdot)$ derived from the local equation (2.7). These two conditions are different. Such difference should be reflected on the implementation of the algorithm. Therefore, h plays different roles in different steps. For the estimation of the parametric component, we set the bandwidth parameter $h = \alpha_1 n^{-1/3}$ according to the asymptotic properties. We find that $\alpha_1 = 0.5$ works well among various values of α_1 between 0.1 and 1 under all configurations. For the estimation of the nonparametric function $f(\cdot)$, we set $h = \alpha_2 n^{-1/5}$ and select $\alpha_2 = 0.2$ from various values among 0.1 to 1 as a trade off between bias and the mean integrated squared error (MISE) under all settings considered. To assess the performance of the proposed resampling method for variance estimation, we generated $M = 500$ sets of ξ's for each simulated data and computed the empirical variance of $\hat{\boldsymbol{\beta}}^*$'s as the estimated asymptotic variance of $\hat{\boldsymbol{\beta}}$. All simulations are based on 500 replications with sample sizes 200 and 400.

Table 1 reports the simulation results on the estimate of $\boldsymbol{\beta}$. Bias is the sampling mean of the estimate minus the true value, SE is the estimated asymptotic standard error based on the resampling method, SD is the sample standard deviation of the parameter estimate, CP is the 95% empirical coverage probability for the parameter based on SE, and RMSE is the root mean square error of the estimate. It is clear that the proposed estimator for $\boldsymbol{\beta}$ performs well for all the situations under consideration. Specifically, the bias is small, the estimated and empirical standard errors match well, and the 95% empirical coverage probabilities are reasonably close to the nominal level. The performance of the proposed estimator is improved as the sample size increases from 200 and 400.

For the nonparametric component, Figures 1 and 2 depict the estimated nonparametric function $f(\cdot)$ and the 95% point-wise confidence interval for covariate-independent censoring. The bounds of the confidence interval are given by the 97.5th and 2.5th percentiles of the estimated function at each grid point among 500 replications. The results indicate that the nonparametric function $f(\cdot)$ is estimated with good accuracy, and the confidence interval becomes narrower as the sample size increases. The results for covariate-dependent censoring are similar and thus not reported here.

To examine the sensitivity of the computing algorithm to initial values, we repeat the above analysis by using different initial values, such as $\hat{f}^{(0)}(\cdot) = -1$ or 1. The results are similar to those reported in Table 1 and Figures 1 and 2.

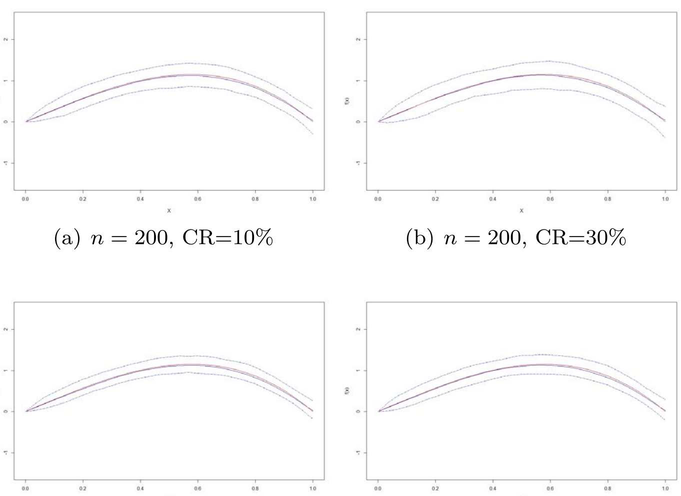
5. Application

5.1. Veteran's Administration lung cancer study

We apply the proposed methodology to a dataset from the Veteran's Administration lung cancer trial [20]. A total of 137 patients were randomized to two

TABLE 1
Simulation results.

$g(u) = u$			$\beta = -1$				$\beta = 1$			
n	CS	CR	Bias	RMSE	SE/SD	CP	Bias	RMSE	SE/SD	CP
200	(i)	10%	0.010	0.069	0.915	0.924	0.002	0.115	0.951	0.944
		30%	0.006	0.079	0.961	0.948	0.007	0.137	0.992	0.932
	(ii)	10%	0.006	0.074	0.905	0.918	0.004	0.121	0.944	0.948
		30%	−0.003	0.088	0.925	0.922	0.005	0.142	0.963	0.939
400	(i)	10%	0.013	0.049	0.953	0.934	−0.001	0.080	0.984	0.944
		30%	0.010	0.055	0.992	0.930	0.002	0.097	0.999	0.930
	(ii)	10%	0.013	0.052	0.933	0.934	0.000	0.083	0.979	0.952
		30%	0.011	0.061	0.966	0.934	0.002	0.096	1.013	0.942
$g(u) = \exp(u)$			$\beta = -1$				$\beta = 1$			
n	CS	CR	Bias	RMSE	SE/SD	CP	Bias	RMSE	SE/SD	CP
200	(i)	10%	0.005	0.123	0.912	0.920	0.005	0.197	0.950	0.934
		30%	0.053	0.147	0.990	0.897	−0.046	0.221	1.171	0.951
	(ii)	10%	0.005	0.123	0.906	0.907	−0.003	0.198	0.966	0.930
		30%	0.029	0.145	1.108	0.933	−0.013	0.239	1.322	0.947
400	(i)	10%	0.003	0.085	0.929	0.920	0.006	0.136	0.961	0.936
		30%	0.029	0.100	1.055	0.915	−0.025	0.158	0.123	0.942
	(ii)	10%	0.003	0.084	0.934	0.924	0.008	0.137	0.967	0.944
		30%	−0.001	0.104	1.076	0.938	0.013	0.171	1.314	0.970

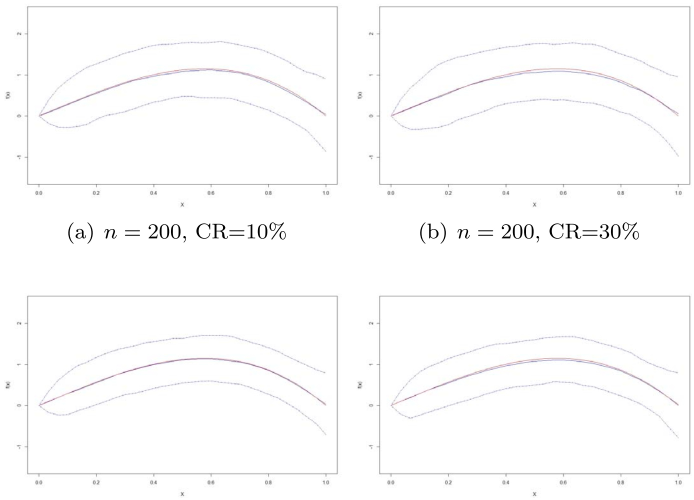


(a) $n = 200$, CR=10% (b) $n = 200$, CR=30%

(c) $n = 400$, CR=10% (d) $n = 400$, CR=30%

FIG 1. *The estimated nonlinear function and 95% point-wise confidence interval for $g(u) = u$ and covariate-independent censoring.*

chemotherapeutic agents (0, standard; 1, test). The response was the survival time of each patient, with a censoring rate of 6.6%. The dataset includes six covariates: treatment indicator, tumor type with four levels (large, adeno, small, and squamous), karnofsky score, months from diagnosis, age, and prior therapy (0 = no, 10 = yes). Many authors have analyzed this dataset in the hazard



(a) $n = 200$, CR=10% (b) $n = 200$, CR=30%

(c) $n = 400$, CR=10% (d) $n = 400$, CR=30%

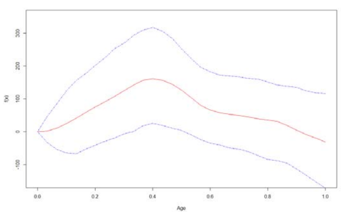
FIG 2. *The estimated nonlinear function and 95% point-wise confidence interval for* $g(u) = \exp(u)$ *and covariate-independent censoring.*

model framework, for example, [24, 25, 36]. In particular, Lu and Zhang [25] pointed out that the covariate age is a complex confounding factor and usually has a nonlinear effect. To compare with [25], we fitted the data to the proposed partially linear TMRL models with three covariates: treatment, tumor type, and age, where age is assumed to have a nonlinear effect. Following the choice of the link function $g(u)$ presented in [35], three special cases of the class of partially linear TMRL models were considered: partially linear AMRL model ($g(u) = u$), partially linear PMRL model ($g(u) = \exp(u)$), and partially linear Box-Cox transformation MRL models ($g(u) = [(u + 1)^2 - 1]/2$). We rescaled age between 0 and 1 and treated tumor type as a categorical variable with the large type as the reference. In estimating the parametric component and the nonparametric function $f(\cdot)$, we set the bandwidth parameter $h = \alpha_1 n^{-1/3}$ and $h = \alpha_2 n^{-1/5}$ and tried various values of α_1 and α_2 among 0.1 to 1. The results are similar. Thus, we only report the results for $\alpha_1 = 0.5$ and $\alpha_2 = 0.2$ as those in the simulation studies. Using a single PC with an Intel(R) Core(TM) i7-4790 CPU @3.60GHz and 16.00 GB RAM, the computing times for $g(u) = u$, $g(u) = \exp(u)$, and $g(u) = [(u + 1)^2 - 1]/2$ are 30, 45, and 22 minutes; respectively.

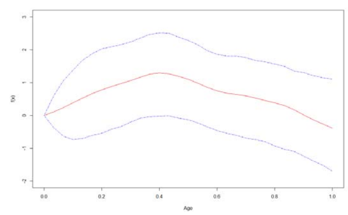
Table 2 presents the estimated regression coefficients along with the standard error estimates based on 500 resamplings for the three models under consideration. Similar to the findings of [25], tumor type (small versus large, adeno versus large) is significant for $g(u) = u$ and $g(u) = \exp(u)$. However, tumor type is not significant for $g(u) = [(u + 1)^2 - 1]/2$. Treatment is not significant in all the three models. Furthermore, Figure 3 displays the estimates of the nonparamet-

TABLE 2
Estimates of linear coefficients for lung cancer data.

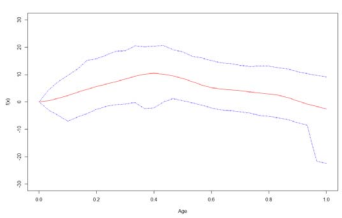
	$g(u) = u$		$g(u) = \exp(u)$		$g(u) = \frac{(1+u)^2-1}{2}$	
Covariate	Est	SE	Est	SE	Est	SE
Treatment	4.180	26.453	0.004	0.173	0.101	3.039
Squamous vs large	47.129	47.439	0.226	0.222	2.206	5.276
Small vs large	−102.150	29.199	−0.849	0.232	−6.733	4.146
Adeno vs large	−106.229	28.757	−0.949	0.214	−7.233	3.971



(a) $g(u) = u$



(b) $g(u) = \exp(u)$



(c) $g(u) = \frac{(1+u)^2-1}{2}$

FIG 3. *The estimated nonlinear function and 95% point-wise confidence interval for the lung cancer data.*

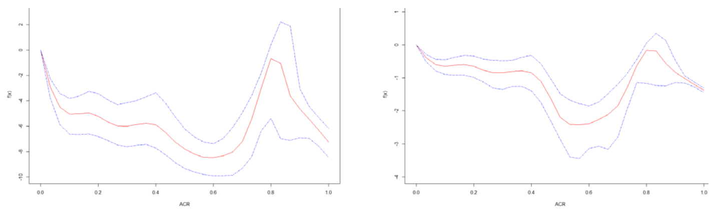
ric components. The red curves are estimated nonparametric functions, and the blue curves are the resampling-based 95% point-wise confidence intervals. The plots clearly show a nonlinear age effect (bell-shape) on the MRL function. The MRL function first increases and then decreases in the covariate age. This result is in agreement with the U-shape nonlinear effect of age on the hazard function [25]. In the present study, the zero line is included in the 95% confidence interval for $g(u) = \exp(u)$ and $g(u) = [(u+1)^2 - 1]/2$, but not included in the confidence interval for $g(u) = u$.

5.2. CKD study of type 2 diabetic patients

This section applied our methodology to a study concerning CKD for type 2 diabetic patients. This study was based on the Hong Kong Diabetes Registry established in 1995 as part of a continuous quality improvement program at the Prince of Wales Hospital, Hong Kong. The detailed descriptions of patient recruitment and characterization are provided in [26]. The event (clinical

Table 3
Estimates of linear coefficients for diabetic data set.

Covariate	$g(u)=u$ Est	$g(u)=u$ SE	$g(u)=\exp(u)$ Est	$g(u)=\exp(u)$ SE
Gender	−0.560	0.646	−0.040	0.057



(a) $g(u) = u$ (b) $g(u) = \exp(u)$

Fig 4. *The estimated nonlinear function and 95% point-wise confidence interval for the CKD data.*

endpoint) of CKD was defined as estimated glomerular filtration rate (eGFR) < 60 ml/min per 1.7 m^2 [26]. The survival time of CKD was calculated as the duration from enrollment to the first clinical endpoint. A subject was censored if its first clinical endpoint occurred beyond January 31, 2009. In the current study, we are mainly interested in type 2 diabetic patients with cardiovascular and renal disease history who had low health conditions among the cohort. This sub-cohort consists of 429 patients, and the censoring rate is around 51%.

Albuminuria levels measured by albumin-creatinine ratio (ACR) are important for evaluating the risk of CKD [26]. Luk et al. [26] used two categorical variables, microalbuminuria (ACR of 2.5–30 mg/mmol for women or 3.5–30 mg/mmol for men) and macroalbuminuria (ACR $>$ 30 mg/mmol), to assess the effect of albuminuria levels on the progression of CKD. To be more comprehensive, in the present study, we directly investigate the effect of ACR on the MRL function of CKD, while adjusting for the covariate gender. We first rescaled ACR between 0 and 1. As in the Veteran's Administration lung cancer study presented in Section 5.1, we report the results based on the bandwidth parameter $h = 0.5n^{-1/3}$ and $h = 0.2n^{-1/5}$ for estimating the parametric effect and the nonparametric part $f(\cdot)$, respectively. Using a single PC with an Intel(R) Core(TM) i7-4790 CPU @3.60GHz and 16.00 GB RAM, the computing times for $g(u) = u$ and $g(u) = \exp(u)$ are 30 and 60 minutes; respectively. Table 3 summarizes the estimated coefficients and their standard errors obtained based on 500 resamplings for $g(u) = u$ and $g(u) = \exp(u)$. The MRL function of CKD is not significantly affected by gender.

Figure 4 gives the estimates of the nonlinear components along with the resampling-based 95% point-wise confidence intervals. Based on the plots, the effects of ACR on the MRL function of CKD have a clear nonlinear pattern, and the nonlinear effects are evidenced by the 95% confidence intervals, which do not include the zero line. Considering that there are only five participants with rescaled ACR > 0.6 (ACR > 446.8), the estimated nonparametric functions are unreliable in the region with sparse data. Therefore, we focus on the estimated curves with rescaled ACR in $[0, 0.6]$. We have the following findings. (i) The MRL function of CKD is roughly decreasing as ACR gets large, which is compatible with the results of [26] that patients with microalbuminuria or macroalbuminuria have significantly greater hazards of developing CKD than patients with normal ACR. (ii) Although the trend of the estimated curves is generally decreasing, it is not monotone. There are some intervals on which the estimated nonparametric functions are slightly increasing. The biological meaning for such a pattern is unclear and worth further investigating. (iii) The estimated curves decrease drastically with rescaled ACR in the intervals $[0, 0.1]$ and $[0.4, 0.6]$ (ACR in $[0, 74.5]$ and $[297.8, 446.8]$, respectively), but are relatively flat with rescaled ACR in the interval $[0.1, 0.4]$ (ACR in $[74.5, 297.8]$). This result suggests that patients with rescaled ACR in $[0, 0.1]$ or $[0.4, 0.6]$ should be more cautious in taking surgeries or medicines that may increase ACR substantially. Overall, our analysis is compatible with the research of [26], but we reveal the nonlinear effect of ACR on CKD progression more comprehensively. To see the influence of outliers, we rescaled ACR between 0 and 1 after discarding the data points with ACR > 446.8 and reconducted the analysis. Again, gender is insignificant, and the pattern of the estimated nonparametric functions is similar to that in Figure 4 and not reported.

6. Discussion

In this paper, we have proposed a class of partially linear TMRL models and developed a martingale-based approach by solving a system of global and local estimating equations. We have established the root-n consistency and asymptotic normality for the estimates of regression coefficients and obtained the asymptotic properties for the estimated nonparametric components, including the baseline function and the nonlinear covariate effect. We have proposed an iterative algorithm for computing the estimates and provided a resampling method for estimating the asymptotic variances of the regression coefficients.

The current study has several extensions. First, the proposed estimating equations are given in a somewhat ad hoc fashion and thus might not be efficient. Improvement of the proposed approach to enhance efficiency merit further research efforts. Second, Cortese, Holmboe and Scheike [10] considered the restricted MRL model for right-censored and left-truncated data. The proposed models and method may be extended to cope with left-truncated data. Finally, multivariate survival outcomes are frequently encountered in substantive studies. Generalizing the proposed methodology to analyze multivariate survival data is worthy of further investigation.

Appendix A: Proofs of asymptotic results

For fixed $\boldsymbol{\beta}$ and f, denote the estimate of $m_0(t)$ derived from estimating equation (2.3) as $\hat{m}_0(t;\boldsymbol{\beta},f)$, and denote the global estimating function for $\boldsymbol{\beta}$ as

$$\begin{aligned}\mathbf{U}_G(\boldsymbol{\beta},f) = &\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i g\big\{\hat{m}_0(t;\boldsymbol{\beta},f)+\boldsymbol{\beta}^T\mathbf{Z}_i+f(X_i)\big\}dN_i(t)\\ &-\sum_{i=1}^n \int_0^\tau Y_i(t)\mathbf{Z}_i d\big(g\big\{\hat{m}_0(t;\boldsymbol{\beta},f)+\boldsymbol{\beta}^T\mathbf{Z}_i+f(X_i)\big\}+t\big).\end{aligned} \tag{A.1}$$

For fixed $m_0(t)$ and $\boldsymbol{\beta}$, denote the local estimating functions for $\gamma_0(x)$ and $\gamma_1(x)$ as

$$\begin{aligned}&\mathbf{U}_L(m_0,\boldsymbol{\beta},\gamma_0,\gamma_1)(x)\\ &\quad=\sum_{i=1}^n \int_0^\tau \left(1,\frac{X_i-x}{h}\right)^T K_h(X_i-x)\\ &\qquad\times g\big\{m_0(t)+\boldsymbol{\beta}^T\mathbf{Z}_i+\gamma_0(x)+\gamma_1(x)(X_i-x)\big\}dN_i(t)\\ &\qquad-\sum_{i=1}^n \int_0^\tau \left(1,\frac{X_i-x}{h}\right)^T K_h(X_i-x)Y_i(t)\\ &\qquad\times d\big(g\big\{m_0(t)+\boldsymbol{\beta}^T\mathbf{Z}_i+\gamma_0(x)+\gamma_1(x)(X_i-x)\big\}+t\big).\end{aligned} \tag{A.2}$$

Proof of Theorem 3.1. To establish the asymptotic results in Theorem 3.1, we first need to prove the consistency of $\hat{m}_0(\cdot)$, $\hat{\boldsymbol{\beta}}$ and $\hat{f}(\cdot)$ obtained from estimating equations (2.3)–(2.7). Because the global consistency is difficult to derive, we only prove the local consistency. That is, we only consider a small neighborhood of the true parameters $\boldsymbol{\beta}_*$ and $f_*(\cdot)$, but not the possible domain of $\boldsymbol{\beta}$ and $f(\cdot)$.

The one-step estimators can be used as the initial estimators for the fully-iterated estimators. For ease of exposition, we defer the derivation and the proof of the local consistency of the one-step estimator to Appendix B. Following the arguments of [4], the fully-iterated estimators $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}_0(x)$ and $\hat{\gamma}_1(x)$ are also locally consistent.

To prove the consistency of $\hat{m}_0(t)$, we first need to show the consistency of $\hat{m}_0(t;\boldsymbol{\beta}_0,f_*)$. Let $\mathcal{S}$ be the proper space of all the possible baseline mean residual life functions. For any m_1, $m_2 \in \mathcal{S}$, define

$$d(m_1,m_2)=\sup_{t\in[0,\tau]}|m_1(t)-m_2(t)|$$

as a metric on $\mathcal{S}$.

Let H be a mapping on $\mathcal{S}$ defined by

$$\begin{aligned}H(m)(t)=\frac{1}{n}\sum_{i=1}^n\int_0^t &\big[g\big\{m(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}dN_i(s)\\ &-Y_i(s)d\big(g\big\{m(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}+s\big)\big].\end{aligned}$$

Simple algebraic manipulation yields

$$
\begin{aligned}
&H(m_1)(t) - H(m_2)(t)\\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^t \Big[\big(g\big\{m_1(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\} - g\big\{m_2(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}\big)dN_i(s)\\
&\quad - Y_i(s)d\big(g\big\{m_1(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\} - g\big\{m_2(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}\big)\Big]\\
&= \frac{1}{n}\sum_{i=1}^{n}\Big[\big(g\big\{m_1(\tilde{T}_i)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\} - g\big\{m_2(\tilde{T}_i)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}\big)\Delta_i I(\tilde{T}_i<t)\\
&\quad - \big(g\big\{m_1(t\wedge\tilde{T}_i)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\} - g\big\{m_2(t\wedge\tilde{T}_i)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}\big)\Big]\\
&= \frac{1}{n}\sum_{i=1}^{n}\big(\Delta_i I(\tilde{T}_i<t)-1\big)\Big[g\big\{m_1(t\wedge\tilde{T}_i)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}\\
&\quad - g\big\{m_2(t\wedge\tilde{T}_i)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}\Big].
\end{aligned}
$$

In the above derivation, note that by definition, $g\{m_1(0)\} = g\{m_2(0)\} = E[T_i|X_i = 0, \mathbf{Z}_i = 0]$, thus the term involving $m_1(0)$ and $m_2(0)$ vanishes. For any fixed $\varepsilon > 0$, if $d(m_1, m_2) > \varepsilon$, then $d(H(m_1), H(m_2)) \geq c\varepsilon$ holds almost surely, where c is a positive constant. We elaborate this point as follows. If $d(m_1, m_2) > \epsilon$, then there exists a $t_* \in [0, \tau]$ such that $|m_2(t_*) - m_1(t_*)| > \epsilon$, now we have

$$
\begin{aligned}
&d\big(H(m_1), H(m_2)\big)\\
&\geq \left|\frac{1}{n}\sum_{i=1}^{n}\big(\Delta_i I(\tilde{T}_i<t_*)-1\big)\int_{m_2(t_*\wedge\tilde{T}_i)}^{m_1(t_*\wedge\tilde{T}_i)} \dot{g}\big\{s+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}ds\right|\\
&\geq \left|\frac{1}{n}\sum_{i=1}^{n}\big(\Delta_i I(\tilde{T}_i<t_*)-1\big)I(\tilde{T}_i\geq\tau)\int_{m_2(t_*)}^{m_1(t_*)} \dot{g}\big\{s+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}ds\right|\\
&= \frac{1}{n}\sum_{i=1}^{n} I(\tilde{T}_i\geq\tau)\dot{g}\big\{s_*+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}|m_2(t_*)-m_1(t_*)|,
\end{aligned}
$$

where s_* lies in between $m_1(t_*)$ and $m_2(t_*)$, and the equality holds by noting that $I(\tilde{T}_i < t_*)I(\tilde{T}_i \geq \tau) = 0$. Then, in view of conditions (C2) and (C4), using the law of large numbers, $\frac{1}{n}\sum_{i=1}^n I(\tilde{T}_i \geq \tau)\dot{g}\{s_* + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\}$ converges to a positive constant. Therefore, for large n, there exists a positive constant c such that $d(H(m_1), H(m_2)) \geq c\epsilon$. Thus, the linear bounded operator H is injective, this ensures that the inverse of H exists. Furthermore, the inverse of H is continuous and bounded by the arbitrariness of ε [32]. Using the uniform strong law of large numbers and the continuity of $m_*(t)$ (condition (C2)), we can show that $\sup_{t\in[0,\tau]} |H(\hat{m}_0(\cdot; \boldsymbol{\beta}_0, f_*))(t) - H(m_*)(t)| \to 0$ almost surely. Therefore, with probability 1, $\hat{m}_0(t; \boldsymbol{\beta}_0, f_*)$ is in the neighborhood of m_* with arbitrarily small radius ε under the metric $d(\cdot, \cdot)$. It follows that $\hat{m}_0(t; \boldsymbol{\beta}_0, f_*)$ converges to $m_*(t)$ almost surely. In view of the local consistency of $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}_0$ and $\hat{\gamma}_1$, $\hat{m}_0(t) = \hat{m}_0(t; \hat{\boldsymbol{\beta}}, \hat{\gamma}_0)$ converges to $m_*(t)$ in probability.

Given the consistency of $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}_0$, $\hat{\gamma}_1$ and $\hat{m}_0(t;\boldsymbol{\beta}_0, f_*)$, we will establish the following asymptotic representation of $\hat{\boldsymbol{\beta}}$

$$\begin{aligned} & n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ = & - \mathbf{A}^{-1} n^{-1/2} \sum_{i=1}^n \int_0^\tau \big\{ \mathbf{Z}_i - \boldsymbol{\mu}_{\mathbf{Z}}(t) - \big(\mathbf{Z}_i^* - \boldsymbol{\mu}_{\mathbf{Z}^*,i}\big)\big\} \\ & \times g\big\{m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\} dM_i(t) + o_p(1), \end{aligned} \tag{A.3}$$

where $\mathbf{A}$, $\boldsymbol{\mu}_{\mathbf{Z}}(t)$, $\mathbf{Z}_i^*$ and $\boldsymbol{\mu}_{\mathbf{Z}^*,i}$ are defined in Section 3, and $dM_i(t) = dM_i(t; m_*, \boldsymbol{\beta}_0, f_*)$.

Assuming (A.3) holds, by the regularity conditions given in Theorem 3.1, and the martingale central limit theorem, $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges to a normal random vector with zero-mean and covariance matrix $\mathbf{A}^{-1}\boldsymbol{\Sigma}(\mathbf{A}^{-1})^T$. To prove (A.3), we proceed in seven steps.

Step A1. We first derive an asymptotic representation of $\hat{m}_0(t;\boldsymbol{\beta}_0, f_*)$. Some manipulation yields that $-\frac{1}{n}\sum_{i=1}^n g\{m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\} dM_i(t)$ equals

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n \big[g\big\{\hat{m}_0(t;\boldsymbol{\beta}_0, f_*) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\} - g\big\{m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\}\big] dN_i(t) \\ & - \frac{1}{n}\sum_{i=1}^n Y_i(t) d\big[g\big\{\hat{m}_0(t;\boldsymbol{\beta}_0, f_*) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\} - g\big\{m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\}\big]. \end{aligned}$$

Define $V(x) = \int_0^x v(s) ds$, where $v\{m_*(t)$ is define in Section 3. It can be verified that

$$v\big\{m_*(t)\big\} E\big[dQ(t)\big] + E\big[Y(t)\dot{g}\big\{m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z} + f_*(X)\big\}\big] dv\big\{m_*(t)\big\} = 0. \tag{A.4}$$

Using the Taylor expansion, we have

$$\begin{aligned} & - \frac{1}{n}\sum_{i=1}^n \int_0^t g\big\{m_*(s) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\} dM_i(s) \\ = & \frac{1}{n}\sum_{i=1}^n \int_0^t \dot{g}\big\{m_*(s) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\} \frac{V\{\hat{m}_0(s;\boldsymbol{\beta}_0, f_*)\} - V\{m_*(s)\}}{v\{m_*(s)\}} dN_i(s) \\ & - \frac{1}{n}\sum_{i=1}^n \int_0^t Y_i(s) d\Big[\frac{\dot{g}\{m_*(s) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\}}{v\{m_*(s)\}} \\ & \times \big(V\big\{\hat{m}_0(s;\boldsymbol{\beta}_0, f_*)\big\} - V\big\{m_*(s)\big\}\big)\Big] + o_p\big(n^{-1/2}\big) \\ = & \frac{1}{n}\sum_{i=1}^n \int_0^t \dot{g}\big\{m_*(s) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\} \frac{V\{\hat{m}_0(s;\boldsymbol{\beta}_0, f_*)\} - V\{m_*(s)\}}{v\{m_*(s)\}} dN_i(s) \\ & - \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{Y_i(s)\dot{g}\{m_*(s) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\}}{v\{m_*(s)\}} d\big[V\big\{\hat{m}_0(s;\boldsymbol{\beta}_0, f_*)\big\} - V\big\{m_*(s)\big\}\big] \end{aligned}$$

$$
\begin{aligned}
& -\frac{1}{n}\sum_{i=1}^{n}\int_0^t Y_i(s)\big[V\big\{\hat{m}_0(s;\boldsymbol{\beta}_0,f_*)\big\}-V\big\{m_*(s)\big\}\big] \\
& \times \frac{v\{m_*(s)\}d\dot{g}\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\}-\dot{g}\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\}dv\{m_*(s)\}}{v\{m_*(s)\}^2} \\
& + o_p\big(n^{-1/2}\big).
\end{aligned}
$$

By the uniform strong law of large numbers and (A.4), we have

$$
\begin{aligned}
& -\frac{1}{n}\sum_{i=1}^{n}\int_0^t g\big\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}dM_i(s) \\
=& \int_0^t \frac{V\{\hat{m}_0(s;\boldsymbol{\beta}_0,f_*)\}-V\{m_*(s)\}}{v\{m_*(s)\}^2} \\
& \times \left(v\big\{m_*(s)\big\}\frac{\sum_{i=1}^n dQ_i(s)}{n}+\frac{\sum_{i=1}^n Y_i(s)\dot{g}\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\}}{n}dv\big\{m_*(s)\big\}\right) \\
& -\int_0^t \frac{n^{-1}\sum_{i=1}^n Y_i(s)\dot{g}\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\}}{v\{m_*(s)\}} \\
& \quad d\big[V\big\{\hat{m}_0(s;\boldsymbol{\beta}_0,f_*)\big\}-V\big\{m_*(s)\big\}\big]+o_p\big(n^{-1/2}\big) \\
=& \int_0^t \frac{V\{\hat{m}_0(s;\boldsymbol{\beta}_0,f_*)\}-V\{m_*(s)\}}{v\{m_*(s)\}^2} \\
& \quad \times \big(v\big\{m_*(s)\big\}E\big[dQ(s)\big]+E\big[Y(s)\dot{g}\big\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)\big\}\big]dv\big\{m_*(s)\big\}\big) \\
& -\int_0^t \frac{E[Y(s)\dot{g}\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)\}]}{v\{m_*(s)\}}d\big[V\big\{\hat{m}_0(s;\boldsymbol{\beta}_0,f_*)\big\}-V\big\{m_*(s)\big\}\big] \\
& + o_p\big(n^{-1/2}\big) \\
=& -\int_0^t \frac{E[Y(s)\dot{g}\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)\}]}{v\{m_*(s)\}}d\big[V\big\{\hat{m}_0(s;\boldsymbol{\beta}_0,f_*)\big\}-V\big\{m_*(s)\big\}\big] \\
& + o_p\big(n^{-1/2}\big).
\end{aligned}
$$

Thus, we have the representation

$$
\begin{aligned}
& V\big\{\hat{m}_0(t;\boldsymbol{\beta}_0,f_*)\big\}-V\big\{m_*(t)\big\} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\int_t^\tau \frac{v\{m_*(s)\}g\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\}}{E[Y(s)\dot{g}\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)\}]}dM_i(s)+o_p(1). \qquad \text{(A.5)}
\end{aligned}
$$

Step A2. We now find the limit of the derivative of $\hat{m}_0(t;\boldsymbol{\beta},f)$ with respect to $\boldsymbol{\beta}$ at $\boldsymbol{\beta}=\boldsymbol{\beta}_0$ and $f=f_*$. Replacing $m_0(t)$ with $\hat{m}_0(t;\boldsymbol{\beta},f)$ in (2.3), and taking derivatives with respect to $\boldsymbol{\beta}$ in both sides, we have

$$
\begin{aligned}
& \sum_{i=1}^{n}\dot{g}\big\{\hat{m}_0(t;\boldsymbol{\beta},f)+\boldsymbol{\beta}^T\mathbf{Z}_i+f_*(X_i)\big\}\left(\frac{\partial\hat{m}_0(t;\boldsymbol{\beta},f)}{\partial\boldsymbol{\beta}}+\mathbf{Z}_i\right)dN_i(t) \\
& -\sum_{i=1}^{n}Y_i(t)d\left[\dot{g}\big\{\hat{m}_0(t;\boldsymbol{\beta},f)+\boldsymbol{\beta}^T\mathbf{Z}_i+f_*(X_i)\big\}\left(\frac{\partial\hat{m}_0(t;\boldsymbol{\beta},f)}{\partial\boldsymbol{\beta}}+\mathbf{Z}_i\right)\right]=0.
\end{aligned}
$$

Using a similar argument as in the derivation of (A.5), we obtain

$$\frac{\partial \hat{m}_0(t;\boldsymbol{\beta},f)}{\partial \boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0, f=f_*} = -\int_t^\tau \frac{B(s,t)E[\mathbf{Z}dQ(s)]}{E[Y(s)\dot{g}\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(X)\}]} + o_p(1)$$
$$\triangleq -\mathbf{a}(t) + o_p(1). \tag{A.6}$$

Step A3. In this step, we need to find the limit of $n^{-1}\partial \mathbf{U}_G(\boldsymbol{\beta},f)/\partial\boldsymbol{\beta}$ at $\boldsymbol{\beta}=\boldsymbol{\beta}_0$ and $f=f_*$. Some manipulation yields

$$\begin{aligned}
&\frac{1}{n}\frac{\partial \mathbf{U}_G(\boldsymbol{\beta},f)}{\partial\boldsymbol{\beta}}\\
&= \frac{1}{n}\sum_{i=1}^n\int_0^\tau \mathbf{Z}_i\dot{g}\big\{\hat{m}_0(t;\boldsymbol{\beta},f)+\boldsymbol{\beta}^T\mathbf{Z}_i+f(X_i)\big\}\left(\frac{\partial \hat{m}_0(t;\boldsymbol{\beta},f)}{\partial\boldsymbol{\beta}}+\mathbf{Z}_i\right)^T dN_i(t)\\
&\quad - \frac{1}{n}\sum_{i=1}^n\int_0^\tau \mathbf{Z}_iY_i(t)d\left[\dot{g}\big\{\hat{m}_0(t;\boldsymbol{\beta},f)+\boldsymbol{\beta}^T\mathbf{Z}_i+f(X_i)\big\}\left(\frac{\partial \hat{m}_0(t;\boldsymbol{\beta},f)}{\partial\boldsymbol{\beta}}+\mathbf{Z}_i\right)^T\right]\\
&= \frac{1}{n}\sum_{i=1}^n\int_0^\tau \mathbf{Z}_i\mathbf{Z}_i^T\big[\dot{g}\big\{\hat{m}_0(t;\boldsymbol{\beta},f)+\boldsymbol{\beta}^T\mathbf{Z}_i+f(X_i)\big\}dN_i(t)\\
&\quad - Y_i(t)d\dot{g}\big\{\hat{m}_0(t;\boldsymbol{\beta},f)+\boldsymbol{\beta}^T\mathbf{Z}_i+f(X_i)\big\}\big]\\
&\quad + \frac{1}{n}\sum_{i=1}^n\int_0^\tau \mathbf{Z}_i\dot{g}\big\{\hat{m}_0(t;\boldsymbol{\beta},f)+\boldsymbol{\beta}^T\mathbf{Z}_i+f(X_i)\big\}\left(\frac{\partial \hat{m}_0(t;\boldsymbol{\beta},f)}{\partial\boldsymbol{\beta}}\right)^T dN_i(t)\\
&\quad - \frac{1}{n}\sum_{i=1}^n\int_0^\tau \mathbf{Z}_iY_i(t)d\left[\dot{g}\big\{\hat{m}_0(t;\boldsymbol{\beta})+\boldsymbol{\beta}^T\mathbf{Z}_i+f(X_i)\big\}\left(\frac{\partial \hat{m}_0(t;\boldsymbol{\beta},f)}{\partial\boldsymbol{\beta}}\right)^T\right].
\end{aligned}$$

Since $\int_0^\tau h(t)dN_i(t) = h(\tilde{T}_i)\Delta_i$ and $\int_0^\tau Y_i(t)dh(t) = h(\tilde{T}_i) - h(0)$, where $h(\cdot)$ is a function, we rewrite the above expression as

$$\begin{aligned}
&\frac{1}{n}\frac{\partial \mathbf{U}_G(\boldsymbol{\beta},f)}{\partial\boldsymbol{\beta}}\\
&= \frac{1}{n}\sum_{i=1}^n\int_0^\tau \mathbf{Z}_i\mathbf{Z}_i^T dQ_i\big(t;\hat{m}_0(\cdot;\boldsymbol{\beta},f),\boldsymbol{\beta},f\big)\\
&\quad + \frac{1}{n}\sum_{i=1}^n(\Delta_i-1)\mathbf{Z}_i\dot{g}\big\{\hat{m}_0(\tilde{T}_i;\boldsymbol{\beta},f)+\boldsymbol{\beta}^T\mathbf{Z}_i+f(X_i)\big\}\left(\frac{\partial \hat{m}_0(\tilde{T}_i;\boldsymbol{\beta},f)}{\partial\boldsymbol{\beta}}\right)^T\\
&\quad + \frac{1}{n}\sum_{i=1}^n\mathbf{Z}_i\dot{g}\big\{\hat{m}_0(0;\boldsymbol{\beta},f)+\boldsymbol{\beta}^T\mathbf{Z}_i+f(X_i)\big\}\left(\frac{\partial \hat{m}_0(0;\boldsymbol{\beta},f)}{\partial\boldsymbol{\beta}}\right)^T\\
&\triangleq I_1(\boldsymbol{\beta},f)+I_2(\boldsymbol{\beta},f)+I_3(\boldsymbol{\beta},f).
\end{aligned}$$

By the consistency of $\hat{m}_0(\cdot;\boldsymbol{\beta}_0,f_*)$ to $m_*(\cdot)$, and the uniform law of large numbers, we have

$$I_1(\boldsymbol{\beta}_0,f_*) = E\left[\int_0^\tau \mathbf{Z}\mathbf{Z}^T dQ(t)\right] + o_p(1).$$

Using (A.6), we obtain

$$\begin{aligned}
I_2(\boldsymbol{\beta}_0, f_*) &= \frac{1}{n}\sum_{i=1}^{n}(\Delta_i - 1)\mathbf{Z}_i \dot{g}\big\{\hat{m}_0(\tilde{T}_i; \boldsymbol{\beta}_0, f_*) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\} \\
&\quad \times \left[-\int_{\tilde{T}_i}^{\tau} \frac{B(s, \tilde{T}_i)E[\mathbf{Z}^T dQ(s)]}{E[Y(s)\dot{g}\{m_*(s) + \boldsymbol{\beta}_0^T \mathbf{Z} + f_*(X)\}]}\right] + o_p(1) \\
&= -E\left[\int_0^{\tau} \boldsymbol{\mu}_1(s)\mathbf{Z}^T dQ(s)\right] + o_p(1).
\end{aligned}$$

Similarly,

$$I_3(\boldsymbol{\beta}_0, f_*) = -E\left[\int_0^{\tau} \boldsymbol{\mu}_2(s)\mathbf{Z}^T dQ(s)\right] + o_p(1),$$

where $\boldsymbol{\mu}_1(s)$ and $\boldsymbol{\mu}_2(s)$ are defined in Section 3. Thus, we have

$$\begin{aligned}
\frac{1}{n}\frac{\partial \mathbf{U}_G(\boldsymbol{\beta}, f)}{\partial \boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0, f=f_*} &= E\left[\int_0^{\tau}\big\{\mathbf{Z} - \boldsymbol{\mu}_1(t) - \boldsymbol{\mu}_2(t)\big\}\mathbf{Z}^T dQ(t)\right] + o_p(1) \\
&= \mathbf{A}_1 + o_p(1). \qquad (A.7)
\end{aligned}$$

Step A4. We now find the asymptotic representations of $\hat{\gamma}_0$ and $\hat{\gamma}_1$, which are the solutions to the local equation (2.7) at convergence. Let $\hat{m}_\#(t;\boldsymbol{\beta}) = \hat{m}_0(t;\boldsymbol{\beta},\hat{\gamma}_0)$, recall (A.2), we have

$$\begin{aligned}
&\mathbf{U}_L\big(\hat{m}_\#(\cdot;\hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}, \hat{\gamma}_0, \hat{\gamma}_1\big)(x) - \mathbf{U}_L\big(\hat{m}_\#(\cdot;\boldsymbol{\beta}_0), \boldsymbol{\beta}_0, \hat{\gamma}_0, \hat{\gamma}_1\big)(x) \\
=&\sum_{i=1}^{n}\int_0^{\tau}\left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x) \\
&\quad \times g\big\{\hat{m}_\#(t;\hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\beta}}^T \mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\big\} dN_i(t) \\
&- \sum_{i=1}^{n}\int_0^{\tau}\left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x) \\
&\quad \times g\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\big\} dN_i(t) \\
&- \sum_{i=1}^{n}\int_0^{\tau}\left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x) Y_i(t) \\
&\quad \times dg\big\{\hat{m}_\#(t;\hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\beta}}^T \mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\big\} \\
&+ \sum_{i=1}^{n}\int_0^{\tau}\left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x) Y_i(t) \\
&\quad \times dg\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\big\} \\
=&n\mathbf{W}_1(x)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p\big(n^{1/2}\big), \qquad (A.8)
\end{aligned}$$

where

$$\mathbf{W}_1(x) = \frac{1}{n}(\Delta_i - 1)\left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x)$$

$$\times \dot{g}\{\hat{m}_{\#}(\tilde{T}_i;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\}$$
$$\times \left(\left.\frac{\partial \hat{m}_{\#}(\tilde{T}_i;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + \mathbf{Z}_i^T \right)$$
$$+ \frac{1}{n}\left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x)$$
$$\times \dot{g}\{\hat{m}_{\#}(0;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\}$$
$$\times \left(\left.\frac{\partial \hat{m}_{\#}(0;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + \mathbf{Z}_i^T \right).$$

Similarly, we have

$$\mathbf{U}_L\big(\hat{m}_{\#}(\cdot;\boldsymbol{\beta}_0), \boldsymbol{\beta}_0, \hat{\gamma}_0, \hat{\gamma}_1\big)(x) - \mathbf{U}_L(m_*, \boldsymbol{\beta}_0, \hat{\gamma}_0, \hat{\gamma}_1)(x) = n\mathbf{W}_2(x) + o_p\big(n^{1/2}\big), \tag{A.9}$$

where

$$\begin{aligned}\mathbf{W}_2(x) =& \frac{1}{n}(\Delta_i - 1)\left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x)\\ &\times \dot{g}\{m_*(\tilde{T}_i) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\}\\ &\times \frac{V\{\hat{m}_{\#}(\tilde{T}_i;\boldsymbol{\beta}_0)\} - V\{m_*(\tilde{T}_i)\}}{v\{m_*(\tilde{T}_i)\}}\\ &+ \frac{1}{n}\left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x)\dot{g}\{m_*(0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\}\\ &\times \frac{V\{\hat{m}_{\#}(0;\boldsymbol{\beta}_0)\} - V\{m_*(0)\}}{v\{m_*(0)\}}.\end{aligned}$$

Moreover, we can show

$$\begin{aligned}&\mathbf{U}_L(m_*, \boldsymbol{\beta}_0, \hat{\gamma}_0, \hat{\gamma}_1)(x) - \mathbf{U}_L(m_*, \boldsymbol{\beta}_0, f_*, \dot{f}_*)(x)\\ &\quad = n\mathbf{W}_3(x)\begin{bmatrix} \hat{\gamma}_0(x) - f_*(x) \\ h\{\hat{\gamma}_1(x) - \dot{f}_*(x)\}\end{bmatrix} + o_p\big(n^{1/2}\big),\end{aligned} \tag{A.10}$$

where

$$\begin{aligned}\mathbf{W}_3(x) = &\frac{1}{n}\sum_{i=1}^n (\Delta_i - 1)\begin{bmatrix} 1 \\ (X_i - x)/h\end{bmatrix}^{\otimes 2} K_h(X_i - x)\\ &\times \dot{g}\{m_*(\tilde{T}_i) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(x) + \dot{f}_*(x)(X_i - x)\}\\ &+ \frac{1}{n}\sum_{i=1}^n \begin{bmatrix} 1 \\ (X_i - x)/h\end{bmatrix}^{\otimes 2} K_h(X_i - x)\\ &\times \dot{g}\{m_*(0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(x) + \dot{f}_*(x)(X_i - x)\}.\end{aligned}$$

Similar to the derivation of (A.6), we can show

$$\left.\frac{\partial \hat{m}_{\#}(t;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = -\mathbf{a}(t) + o_p(1),$$

where $\mathbf{a}(t)$ is defined in (A.6). It can be checked that $\mathbf{W}_1(x)$ converges to

$$\begin{bmatrix}\mathbf{w}_{11}(x)^T\\ \mathbf{0}_{1\times p}\end{bmatrix},$$

where

$$\begin{aligned}\mathbf{w}_{11}(x) &= r(x)E\big[(\Delta-1)\big\{\mathbf{Z}-\mathbf{a}(\tilde{T}\big\}\dot{g}\big\{m_*(\tilde{T})+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(x)\big\}|X=x\big]\\ &\quad + r(x)E\big[\big\{\mathbf{Z}-\mathbf{a}(0)\big\}\dot{g}\big\{m_*(0)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(x)\big\}|X=x\big],\end{aligned}$$

and $r(x)$ is the density of X_i as stated in condition (C1). Note that

$$\begin{aligned}V\big\{\hat{m}_\#(\tilde{T}_i;\boldsymbol{\beta}_0)\big\}-V\big\{m_*(\tilde{T}_i)\big\} &= -\int_{\tilde{T}_i}^{\tau} d\big[V\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\big\}-V\big\{m_*(t)\big\}\big]\\ &= -\int_0^{\tau} I(\tilde{T}_i<t)d\big[V\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\big\}-V\big\{m_*(t)\big\}\big],\end{aligned} \tag{A.11}$$

and

$$V\big\{\hat{m}_\#(0;\boldsymbol{\beta}_0)\big\}-V\big\{m_*(0)\big\} = -\int_0^{\tau} d\big[V\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\big\}-V\big\{m_*(t)\big\}\big]. \tag{A.12}$$

Thus, using (A.11) and (A.12), we obtain

$$\mathbf{W}_2(x) = -\int_0^{\tau}\begin{bmatrix}w_{21}(x,t)\\ 0\end{bmatrix} d\big[V\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\big\}-V\big\{m_*(t)\big\}\big]+o_p\big(n^{-1/2}\big), \tag{A.13}$$

where

$$\begin{aligned}w_{21}(x,t) &= r(x)E\big[(\Delta-1)\dot{g}\big\{m_*(\tilde{T})+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(x)\big\}I(\tilde{T}<t)v\big\{m_*(\tilde{T})\big\}^{-1}|X=x\big]\\ &\quad + r(x)E\big[\dot{g}\big\{m_*(0)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(x)\big\}v\big\{m_*(0)\big\}^{-1}|X=x\big].\end{aligned}$$

We can also check that $\mathbf{W}_3(x)$ converges to

$$w_{31}(x)\begin{bmatrix}1 & 0\\ 0 & k_2\end{bmatrix},$$

where $k_2=\int x^2K(x)dx$, and

$$w_{31}(x)=r(x)E\big[(\Delta-1)\dot{g}\big\{m_*(\tilde{T})+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(x)\big\}+\dot{g}\big\{m_*(0)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(x)\big\}|X=x\big].$$

Combining (A.8)–(A.10) and (A.13), and the convergence of $\mathbf{W}_1(x)$ and $\mathbf{W}_3(x)$, we have

$$\begin{aligned}\hat{\gamma}_0(x)-f_*(x) &= -\frac{1}{w_{31}(x)}\frac{1}{n}U_{L1}(m_*,\boldsymbol{\beta}_0,f_*,\dot{f}_*)(x)-\frac{\mathbf{w}_{11}(x)^T}{w_{31}(x)}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\ &\quad +\int_0^{\tau}\frac{w_{21}(x,t)}{w_{31}(x)}d\big[V\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\big\}-V\big\{m_*(t)\big\}\big]+o_p\big(n^{-1/2}\big),\end{aligned} \tag{A.14}$$

and

$$h\{\hat{\gamma}_1(x) - \dot{f}_*(x)\} = -\frac{1}{k_2 w_{31}(x)}\frac{1}{n}U_{L2}(m_*, \boldsymbol{\beta}_0, f_*, \dot{f}_*)(x) + o_p(n^{-1/2}), \quad \text{(A.15)}$$

where $U_{L1}(m_*, \boldsymbol{\beta}_0, f_*, \dot{f}_*)(x)$ and $U_{L2}(m_*, \boldsymbol{\beta}_0, f_*, \dot{f}_*)(x)$ are the first and second component of $\mathbf{U}_L(m_*, \boldsymbol{\beta}_0, f_*, \dot{f}_*)(x)$, respectively.

Step A5. Note that for fixed $\boldsymbol{\beta}$, $\hat{m}_\#(t;\boldsymbol{\beta})$ is the solution to the following estimating equation

$$\sum_{i=1}^n \big[g\{\hat{m}_\#(t;\boldsymbol{\beta}) + \boldsymbol{\beta}^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\}dN_i(t) \\ - Y_i(t)d\big(g\{\hat{m}_\#(t;\boldsymbol{\beta}) + \boldsymbol{\beta}^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\} + t\big)\big] = 0, \quad \text{(A.16)}$$

and $\hat{\boldsymbol{\beta}}$ solves the following estimating equation

$$\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i \big[g\{\hat{m}_\#(t;\boldsymbol{\beta}) + \boldsymbol{\beta}^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\}dN_i(t) \\ - Y_i(t)d\big(g\{\hat{m}_\#(t;\boldsymbol{\beta}) + \boldsymbol{\beta}^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\} + t\big)\big] = 0. \quad \text{(A.17)}$$

Simple algebraic manipulation yields

$$\begin{aligned} &-\frac{1}{n}\sum_{i=1}^n g\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\}dM_i(t) \\ =& \frac{1}{n}\sum_{i=1}^n \big[g\{\hat{m}_\#(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\} - g\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\}\big]dN_i(t) \\ &- \frac{1}{n}\sum_{i=1}^n Y_i(t)d\big[g\{\hat{m}_\#(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\} - g\{m_*(t)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+\hat{\gamma}_0(X_i)\}\big] \\ &+ \frac{1}{n}\sum_{i=1}^n \big[g\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\} - g\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\}\big]dN_i(t) \\ &- \frac{1}{n}\sum_{i=1}^n Y_i(t)d\big[g\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\} - g\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\}\big]. \end{aligned}$$

Using the Taylor expansion, similar to the the derivation of (A.5), we have

$$\begin{aligned} &-\frac{1}{n}\sum_{i=1}^n g\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\}dM_i(t) \\ &= -\frac{E[Y(t)\dot{g}\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z} + f_*(X)\}]}{v\{m_*(t)\}}d\big[V\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\} - V\{m_*(t)\}\big] \\ &\quad + \frac{1}{n}\sum_{i=1}^n \{\hat{\gamma}_0(X_i) - f_*(X_i)\}dQ_i(t) + d\{o_p\big(\hat{m}_\#(t;\boldsymbol{\beta}_0) - m_*(t)\big)\}. \end{aligned} \quad \text{(A.18)}$$

Denote the left side of (A.17) as $\mathbf{U}_G(\boldsymbol{\beta}, \hat{m}_{\#}(\cdot;\boldsymbol{\beta}), \hat{\gamma}_0)$, using the Taylor expansion, we have

$$\begin{aligned}
&\mathbf{U}_G\big(\hat{\boldsymbol{\beta}}, \hat{m}_{\#}(\cdot;\hat{\boldsymbol{\beta}}), \hat{\gamma}_0\big) - \mathbf{U}_G\big(\boldsymbol{\beta}_0, \hat{m}_{\#}(\cdot;\boldsymbol{\beta}_0), \hat{\gamma}_0\big)\\
=&\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i\big[g\big\{\hat{m}_{\#}(t;\hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\beta}}^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\big\}\\
&- g\big\{\hat{m}_{\#}(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\big\}\big]dN_i(t)\\
&- \sum_{i=1}^n \int_0^\tau \mathbf{Z}_iY_i(t)d\big[g\big\{\hat{m}_{\#}(t;\hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\beta}}^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\big\}\\
&- g\big\{\hat{m}_{\#}(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\big\}\big]\\
=&\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i\dot{g}\big\{\hat{m}_{\#}(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\big\}\\
&\times\left(\frac{\partial\hat{m}_{\#}(t;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + \mathbf{Z}_i\right)^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)dN_i(t)\\
&- \sum_{i=1}^n \int_0^\tau \mathbf{Z}_iY_i(t)d\bigg[\dot{g}\big\{\hat{m}_{\#}(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\big\}\\
&\times\left(\frac{\partial\hat{m}_{\#}(t;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + \mathbf{Z}_i\right)^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\bigg] + o_p\big(n^{1/2}\big)\\
=&\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i\mathbf{Z}_i^TdQ_i\big(t;\hat{m}_{\#}(\cdot;\boldsymbol{\beta}_0), \boldsymbol{\beta}_0, \hat{\gamma}_0\big)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\
&+ \sum_{i=1}^n(\Delta_i - 1)\mathbf{Z}_i\dot{g}\big\{\hat{m}_{\#}(\tilde{T}_i;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\big\}\\
&\times\left(\frac{\partial\hat{m}_{\#}(\tilde{T}_i;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\
&+ \sum_{i=1}^n\mathbf{Z}_i\dot{g}\big\{\hat{m}_{\#}(0;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + \hat{\gamma}_0(X_i)\big\}\\
&\times\left(\frac{\partial\hat{m}_{\#}(0;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + o_p\big(n^{1/2}\big).
\end{aligned}$$

Using a similar argument as in deriving (A.7), and by the consistency of $\hat{m}_{\#}(t;\boldsymbol{\beta})$ and $\hat{\gamma}_0(X_i)$, we have

$$\mathbf{U}_G\big(\hat{\boldsymbol{\beta}}, \hat{m}_{\#}(\cdot;\hat{\boldsymbol{\beta}}), \hat{\gamma}_0\big) - \mathbf{U}_G\big(\boldsymbol{\beta}_0, \hat{m}_{\#}(\cdot;\boldsymbol{\beta}_0), \hat{\gamma}_0\big) = n\mathbf{A}_1(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + o_p\big(n^{1/2}\big), \tag{A.19}$$

where $\mathbf{A}_1$ is defined in Section 3, also appears in (A.7).

Similarly,

$$\mathbf{U}_G\big(\boldsymbol{\beta}_0, \hat{m}_{\#}(\cdot;\boldsymbol{\beta}_0), \hat{\gamma}_0\big) - \mathbf{U}_G\big(\boldsymbol{\beta}_0, \hat{m}_{\#}(\cdot;\boldsymbol{\beta}_0), f_*\big)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \int_0^{\tau} \mathbf{Z}_i \big[g\big\{ \hat{m}_{\#}(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + \hat{\gamma}_0(X_i) \big\} \\
&\quad - g\big\{ \hat{m}_{\#}(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i) \big\} \big] dN_i(t) \\
&\quad - \sum_{i=1}^{n} \int_0^{\tau} \mathbf{Z}_i Y_i(t) d \big[g\big\{ \hat{m}_{\#}(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + \hat{\gamma}_0(X_i) \big\} \\
&\quad - g\big\{ \hat{m}_{\#}(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i) \big\} \big] \\
&= \sum_{i=1}^{n} \int_0^{\tau} \mathbf{Z}_i \big\{ \hat{\gamma}_0(X_i) - f_*(X_i) \big\} dQ_i\big(t; \hat{m}_{\#}(\cdot;\boldsymbol{\beta}_0), \boldsymbol{\beta}_0, f_*\big) + o_p\big(n^{1/2}\big). \qquad \text{(A.20)}
\end{aligned}
$$

Moreover, similar to the derivation of (A.19), we have

$$
\begin{aligned}
&\mathbf{U}_G\big(\boldsymbol{\beta}_0, \hat{m}_{\#}(\cdot;\boldsymbol{\beta}_0), f_*\big) \\
&= \sum_{i=1}^{n} \int_0^{\tau} \mathbf{Z}_i g\big\{ m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i) \big\} dM_i(t) \\
&\quad + \sum_{i=1}^{n} (\Delta_i - 1) \mathbf{Z}_i \big[g\big\{ \hat{m}_{\#}(\tilde{T}_i;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i) \big\} \\
&\quad - g\big\{ m_*(\tilde{T}_i) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i) \big\} \big] \\
&\quad + \sum_{i=1}^{n} \mathbf{Z}_i \big[g\big\{ \hat{m}_{\#}(0;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i) \big\} - g\big\{ m_*(0) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i) \big\} \big] \\
&= \sum_{i=1}^{n} \int_0^{\tau} \mathbf{Z}_i g\big\{ m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i) \big\} dM_i(t) \\
&\quad + \sum_{i=1}^{n} (\Delta_i - 1) \mathbf{Z}_i \dot{g}\big\{ m_*(\tilde{T}_i) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i) \big\} \frac{V\{\hat{m}_{\#}(\tilde{T}_i;\boldsymbol{\beta}_0)\} - V\{m_*(\tilde{T}_i)\}}{v\{m_*(\tilde{T}_i)\}} \\
&\quad + \sum_{i=1}^{n} \mathbf{Z}_i \dot{g}\big\{ m_*(0) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i) \big\} \frac{V\{\hat{m}_{\#}(0;\boldsymbol{\beta}_0)\} - V\{m_*(0)\}}{v\{m_*(0)\}} + o_p\big(n^{1/2}\big).
\end{aligned}
$$

In view of (A.11) and (A.12), it follows that

$$
\begin{aligned}
&\mathbf{U}_G\big(\boldsymbol{\beta}_0, \hat{m}_{\#}(\cdot;\boldsymbol{\beta}_0), f_*\big) \\
&= \sum_{i=1}^{n} \int_0^{\tau} \mathbf{Z}_i g\big\{ m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i) \big\} dM_i(t) \\
&\quad - n \int_0^{\tau} \mathbf{W}_4(t) d\big[V\big\{ \hat{m}_{\#}(t;\boldsymbol{\beta}_0) \big\} - V\big\{ m_*(t) \big\} \big] + o_p\big(n^{1/2}\big) \\
&= \sum_{i=1}^{n} \int_0^{\tau} \mathbf{Z}_i g\big\{ m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i) \big\} dM_i(t) \\
&\quad - n \int_0^{\tau} \mathbf{w}_4(t) d\big[V\big\{ \hat{m}_{\#}(t;\boldsymbol{\beta}_0) \big\} - V\big\{ m_*(t) \big\} \big] + o_p\big(n^{1/2}\big), \qquad \text{(A.21)}
\end{aligned}
$$

where

$$\begin{aligned}\mathbf{W}_4(t) &= \frac{1}{n}\sum_{i=1}^{n}(\Delta_i - 1)\mathbf{Z}_i\dot{g}\big\{m_*(\tilde{T}_i) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\big\}I(\tilde{T}_i < t)v\big\{m_*(\tilde{T}_i)\big\}^{-1}\\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i\dot{g}\big\{m_*(0) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\big\}v\big\{m_*(0)\big\}^{-1},\\ \mathbf{w}_4(t) &= E\big[(\Delta - 1)\mathbf{Z}\dot{g}\big\{m_*(\tilde{T}) + \boldsymbol{\beta}_0^T\mathbf{Z} + f_*(X)\big\}I(\tilde{T} < t)v\big\{m_*(\tilde{T})\big\}^{-1}\big]\\ &\quad + E\big[\mathbf{Z}\dot{g}\big\{m_*(0) + \boldsymbol{\beta}_0^T\mathbf{Z} + f_*(X)\big\}v\big\{m_*(0)\big\}^{-1}\big],\end{aligned}$$

and $\mathbf{w}_4(t)$ is the limit of $\mathbf{W}_4(t)$.

Adding (A.19), (A.20) and (A.21), and note that $\mathbf{U}_G(\hat{\boldsymbol{\beta}}, \hat{m}_\#(\cdot;\hat{\boldsymbol{\beta}}), \hat{\gamma}_0) = 0$, we obtain

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i g\big\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\big\}dM_i(t)\\ &= -\mathbf{A}_1(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i\big\{\hat{\gamma}_0(X_i) - f_*(X_i)\big\}dQ_i\big(t;\hat{m}_\#(\cdot;\boldsymbol{\beta}_0),\boldsymbol{\beta}_0, f_*\big)\\ &\quad + \int_0^\tau \mathbf{w}_4(t)d\big[V\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\big\} - V\big\{m_*(t)\big\}\big] + o_p\big(n^{-1/2}\big). \qquad \text{(A.22)}\end{aligned}$$

Step A6. We further refine (A.18) and (A.22) by using the expression of $\hat{\gamma}_0(x) - f_*(x)$ given by (A.14). For ease of presentaion, we set $B_2(t) = E[Y(t)\dot{g}\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z} + f_*(X)\}]$ hereafter. Plugging (A.14) into (A.18), we obtain

$$\begin{aligned}&-\frac{1}{n}\sum_{i=1}^{n} g\big\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\big\}dM_i(t)\\ &= -\frac{B_2(t)}{v\{m_*(t)\}}d\big[V\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\big\} - V\big\{m_*(t)\big\}\big]\\ &\quad - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_{31}(X_i)}\frac{1}{n}U_{L1}(m_*, \boldsymbol{\beta}_0, f_*, \dot{f}_*)(X_i)dQ_i(t)\\ &\quad - \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{w}_{11}(X_i)^T}{w_{31}(X_i)}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)dQ_i(t)\\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \frac{w_{21}(X_i, s)}{w_{31}(X_i)}d\big[V\big\{\hat{m}_\#(s;\boldsymbol{\beta}_0)\big\} - V\big\{m_*(s)\big\}\big]dQ_i(t)\\ &\quad + d\big\{o_p\big(\hat{m}_\#(t;\boldsymbol{\beta}_0) - m_*(t)\big)\big\}.\end{aligned}$$

Multiplying the above expression by $\frac{\boldsymbol{\alpha}(t)v\{m_*(t)\}}{B_2(t)}$ on both sides, then taking integral from 0 to τ, it follows that

$$-\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \frac{\boldsymbol{\alpha}(t)v\{m_*(t)\}}{B_2(t)}g\big\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\big\}dM_i(t)$$

$$
\begin{aligned}
&= -\int_0^\tau \boldsymbol{\alpha}(t)d\big[V\big\{\hat{m}_{\#}(t;\boldsymbol{\beta}_0)\big\} - V\big\{m_*(t)\big\}\big] \\
&\quad - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \frac{\boldsymbol{\alpha}(t)v\{m_*(t)\}}{B_2(t)} \frac{\mathbf{w}_{11}(X_i)^T}{w_{31}(X_i)} dQ_i(t)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\quad + \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left(\int_0^\tau \frac{\boldsymbol{\alpha}(t)v\{m_*(t)\}}{B_2(t)} dQ_i(t)\right) \\
&\quad \times \frac{w_{21}(X_i,s)}{w_{31}(X_i)} d\big[V\big\{\hat{m}_{\#}(s;\boldsymbol{\beta}_0)\big\} - V\big\{m_*(s)\big\}\big] \\
&\quad - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \frac{1}{w_{31}(X_i)} \frac{1}{n} U_{L1}(m_*, \boldsymbol{\beta}_0, f_*, \dot{f}_*)(X_i) \frac{\boldsymbol{\alpha}(t)v\{m_*(t)\}}{B_2(t)} dQ_i(t) \\
&\quad + o_p\big(n^{-1/2}\big).
\end{aligned} \tag{A.23}
$$

Plugging (A.14) into (A.22), we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i g\big\{m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\} dM_i(t) \\
&= -\mathbf{A}_1(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i \frac{1}{w_{31}(X_i)} \frac{1}{n} U_{L1}(m_*, \boldsymbol{\beta}_0, f_*, \dot{f}_*)(X_i) dQ_i(t) \\
&\quad + \frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i \frac{\mathbf{w}_{11}(X_i)^T}{w_{31}(X_i)} dQ_i(t)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\quad + \int_0^\tau \mathbf{w}_4(t) d\big[V\big\{\hat{m}_{\#}(t;\boldsymbol{\beta}_0)\big\} - V\big\{m_*(t)\big\}\big] \\
&\quad - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i \left(\int_0^\tau \frac{w_{21}(X_i,t)}{w_{31}(X_i)} d\big[V\big\{\hat{m}_{\#}(t;\boldsymbol{\beta}_0)\big\} - V\big\{m_*(t)\big\}\big]\right) dQ_i(s) \\
&\quad + o_p\big(n^{-1/2}\big).
\end{aligned} \tag{A.24}
$$

Let

$$
\begin{aligned}
D_1(t, ds) &= \frac{v\{m_*(s)\}}{B_2(s)} E\left[\frac{w_{21}(X,t)}{w_{31}(X)} dQ(s)\right], \\
\mathbf{D}_2(t) &= \mathbf{w}_4(t) - E\left[\int_0^\tau \frac{\mathbf{Z} w_{21}(X,t)}{w_{31}(X)} dQ(s)\right],
\end{aligned}
$$

and

$$
\mathbf{A}_{21} = E\left[\int_0^\tau \frac{\mathbf{Z}\mathbf{w}_{11}(X)^T}{w_{31}(X)} dQ(t)\right], \quad \mathbf{A}_{22} = E\left[\frac{\boldsymbol{\alpha}(t)v\{m_*(t)\}}{B_2(t)} \frac{\mathbf{w}_{11}(X)^T}{w_{31}(X)} dQ(t)\right].
$$

Then (A.23) can be rewritten as

$$
\int_0^\tau \left(\boldsymbol{\alpha}(t) - \int_0^\tau \boldsymbol{\alpha}(s) D_1(t, ds)\right) d\big[V\big\{\hat{m}_{\#}(t;\boldsymbol{\beta}_0)\big\} - V\big\{m_*(t)\big\}\big]
$$

$$+\mathbf{A}_{22}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)-\frac{1}{n}\sum_{i=1}^n\int_0^\tau\frac{\boldsymbol{\alpha}(t)v\{m_*(t)\}}{B_2(t)}g\big\{m_*(t)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}dM_i(t)$$
$$+\frac{1}{n}\sum_{i=1}^n\int_0^\tau\frac{1}{w_{31}(X_i)}\frac{1}{n}U_{L1}(m_*,\boldsymbol{\beta}_0,f_*,\dot{f}_*)(X_i)\frac{\boldsymbol{\alpha}(t)v\{m_*(t)\}}{B_2(t)}dQ_i(t)=o_p\big(n^{-1/2}\big), \tag{A.25}$$

and (A.24) can be rewritten as

$$\frac{1}{n}\sum_{i=1}^n\int_0^\tau\mathbf{Z}_ig\big\{m_*(t)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}dM_i(t)$$
$$+(\mathbf{A}_1-\mathbf{A}_{21})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)-\int_0^\tau\mathbf{D}_2(t)d\big[V\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\big\}-V\big\{m_*(t)\big\}\big]$$
$$-\frac{1}{n}\sum_{i=1}^n\int_0^\tau\mathbf{Z}_i\frac{1}{w_{31}(X_i)}\frac{1}{n}U_{L1}(m_*,\boldsymbol{\beta}_0,f_*,\dot{f}_*)(X_i)dQ_i(t)=o_p\big(n^{-1/2}\big). \tag{A.26}$$

Define $\boldsymbol{\rho}(X)=\frac{\mathbf{w}_{11}(X)}{w_{31}(X)}$, we observe that $\mathbf{A}_2=\mathbf{A}_{21}-\mathbf{A}_{22}$, where $\mathbf{A}_2$ is given is Section 3. Adding (A.25) and (A.26), and note that $\boldsymbol{\alpha}(t)$ solves equation (3.1), we have

$$\begin{aligned}&(\mathbf{A}_1-\mathbf{A}_2)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\ =&-\frac{1}{n}\sum_{i=1}^n\int_0^\tau\bigg\{\mathbf{Z}_i-\frac{\boldsymbol{\alpha}(t)v\{m_*(t)\}}{B_2(t)}\bigg\}g\big\{m_*(t)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}dM_i(t)\\ &+\mathbf{G}_1-\mathbf{G}_2+o_p\big(n^{-1/2}\big),\end{aligned} \tag{A.27}$$

where

$$\mathbf{G}_1=\frac{1}{n}\sum_{i=1}^n\int_0^\tau\mathbf{Z}_i\frac{1}{w_{31}(X_i)}\frac{1}{n}U_{L1}(m_*,\boldsymbol{\beta}_0,f_*,\dot{f}_*)(X_i)dQ_i(t)$$

and

$$\mathbf{G}_2=\frac{1}{n}\sum_{i=1}^n\int_0^\tau\frac{1}{w_{31}(X_i)}\frac{1}{n}U_{L1}(m_*,\boldsymbol{\beta}_0,f_*,\dot{f}_*)(X_i)\frac{\boldsymbol{\alpha}(t)v\{m_*(t)\}}{B_2(t)}dQ_i(t).$$

Step A7. In this step, we will find the asymptotic representations of $\mathbf{G}_1$ and $\mathbf{G}_2$. Recall the expression of $\mathbf{U}_L(m_0,\boldsymbol{\beta},\gamma_0,\gamma_1)(x)$ (see (A.2)), we obtain

$$\mathbf{G}_1=\frac{1}{n}\sum_{i=1}^n\int_0^\tau\frac{\mathbf{Z}_i}{w_{31}(X_i)}dQ_i(s)\Bigg(\frac{1}{n}\sum_{j=1}^n\int_0^\tau K_h(X_j-X_i)g\big\{m_*(t)+\boldsymbol{\beta}_0^T\mathbf{Z}_j$$
$$+f_*(X_i)+\dot{f}_*(X_i)(X_j-X_i)\big\}dN_j(t)\Bigg)$$

$$- \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \frac{\mathbf{Z}_i}{w_{31}(X_i)}dQ_i(s)\Bigg(\frac{1}{n}\sum_{j=1}^{n}\int_0^\tau K_h(X_j - X_i)Y_j(t)d\big(g\{m_*(t)$$
$$+ \boldsymbol{\beta}_0^T\mathbf{Z}_j + f_*(X_i) + \dot{f}_*(X_i)(X_j - X_i) + t\big)\Bigg).$$

Using the Taylor expansion and some standard nonparametric techniques, we have

$$\mathbf{G}_1 = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i^* g\big\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\big\}dM_i(t) + o_p\big(n^{-1/2}\big).$$

Similarly,

$$\mathbf{G}_2 = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \boldsymbol{\mu}_{\mathbf{Z}^*,i} g\big\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\big\}dM_i(t) + o_p\big(n^{-1/2}\big),$$

where $\mathbf{Z}_i^*$ and $\boldsymbol{\mu}_{\mathbf{Z}^*,i}$ are defined in Section 3.

By (A.27) and the asymptotic expressions of $\mathbf{G}_1$ and $\mathbf{G}_2$, (A.3) holds. Thus, using the martingale central limit theorem, Theorem 3.1 is proved. $\square$

Proof of Theorem 3.2. Let $\hat{m}_0(t) = \hat{m}_\#(t;\hat{\boldsymbol{\beta}})$, we need to find the asymptotic representation of $\sqrt{n}\{\hat{m}_0(t) - m_*(t)\}$. Recall that $\hat{m}_\#(t;\boldsymbol{\beta}) = \hat{m}_0(t;\boldsymbol{\beta},\hat{\gamma}_0)$, we first consider $\sqrt{n}[V\{\hat{m}_\#(t;\hat{\boldsymbol{\beta}})\} - V\{m_*(t)\}]$. Note that

$$\begin{aligned}&V\big\{\hat{m}_\#(t;\hat{\boldsymbol{\beta}})\big\} - V\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\big\}\\ &= v\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\big\}\left(\frac{\partial \hat{m}_\#(t;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|).\end{aligned}$$

Recall that $\frac{\partial \hat{m}_\#(t;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = -\mathbf{a}(t) + o_p(1)$, and that the asymptotic representation of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ has been established in (A.3), we only need to deal with

$$\hat{\pi}_n(t) \stackrel{\triangle}{=} \sqrt{n}\big[V\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\big\} - V\big\{m_*(t)\big\}\big].$$

To establish the asymptotic representation of $\hat{\pi}_n(t)$, we have the following lemma.

Lemma A1. *Under the regularity conditions (C1)–(C6), if* $nh^2/\{\log(1/h)\} \to \infty$ *and* $nh^4 \to 0$*, we have that* $\hat{\pi}_n(t)$ *satisfies the following integral equation*

$$\hat{\pi}_n(t) - \int_0^\tau a(t,s)d\hat{\pi}_n(s) = \Phi_n(t), \qquad t \in [0,\tau), \tag{A.28}$$

where $a(t,s)$ *is a deterministic function defined later in* (A.33). $\Phi_n(t)$ *can be written as a summation of independent mean zero functions, say,* $n^{-1/2}\sum_{i=1}^n \varphi_i(t)$*, which converges weakly to a mean-zero Gaussian process.* $\varphi_i(t)$ *is defined later in* (A.34).

Proof of Lemma A1. Plugging (A.14) into (A.18) and multiplying by $n^{1/2}$, we obtain

$$\begin{aligned}
&-n^{-1/2}\sum_{i=1}^{n} g\big\{m_*(t)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}dM_i(t)\\
=&-\frac{B_2(t)}{v\{m_*(t)\}}d\hat{\pi}_n(t)-n^{-1/2}\sum_{i=1}^{n}\frac{1}{w_{31}(X_i)}\frac{1}{n}U_{L1}(m_*,\boldsymbol{\beta}_0,f_*,\dot{f}_*)(X_i)dQ_i(t)\\
&-\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{w}_{11}(X_i)^T}{w_{31}(X_i)}\big[n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\big]dQ_i(t)+\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\frac{w_{21}(X_i,s)}{w_{31}(X_i)}d\hat{\pi}_n(s)dQ_i(t)\\
&+d\Big\{o_p\Big(n^{1/2}\big[\hat{m}_{\#}(t;\boldsymbol{\beta}_0)-m_*(t)\big]\Big)\Big\}. \qquad \text{(A.29)}
\end{aligned}$$

Define $d\{\mathbf{c}(t)\}=E[\frac{\mathbf{w}_{11}(X)^T}{w_{31}(X)}dQ(t)]$, and recall that $D_1(t,ds)=\frac{v\{m_*(s)\}}{B_2(s)}\times E[\frac{w_{21}(X,t)}{w_{31}(X)}dQ(s)]$. Multiplying both sides of (A.29) by $\frac{v\{m_*(t)\}}{B_2(t)}$, then replacing the empirical quantities with their limits, and finally replacing the arguments (t,s) with (s,u), (A.29) can be rewritten as

$$\begin{aligned}
&-n^{-1/2}\sum_{i=1}^{n}\frac{v\{m_*(s)\}}{B_2(s)}g\big\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}dM_i(s)\\
=&-d\hat{\pi}_n(s)-n^{-1/2}\sum_{i=1}^{n}\frac{v\{m_*(s)\}}{B_2(s)}\frac{1}{w_{31}(X_i)}\frac{1}{n}U_{L1}(m_*,\boldsymbol{\beta}_0,f_*,\dot{f}_*)(X_i)dQ_i(s)\\
&-\frac{v\{m_*(s)\}}{B_2(s)}d\big\{\mathbf{c}(s)\big\}n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+\int_0^{\tau}D_1(u,ds)d\hat{\pi}_n(u)\\
&+d\Big\{o_p\Big(n^{1/2}\big[\hat{m}_{\#}(s;\boldsymbol{\beta}_0)-m_*(s)\big]\Big)\Big\}.
\end{aligned}$$

Taking integration from t to τ with respect to s on both sides of the above expression yields

$$\begin{aligned}
&-n^{-1/2}\sum_{i=1}^{n}\int_t^{\tau}\frac{v\{m_*(s)\}}{B_2(s)}g\big\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}dM_i(s)\\
=&-\int_t^{\tau}d\hat{\pi}_n(s)-n^{-1/2}\sum_{i=1}^{n}\int_t^{\tau}\frac{v\{m_*(s)\}}{B_2(s)}\frac{1}{w_{31}(X_i)}\frac{1}{n}U_{L1}(m_*,\boldsymbol{\beta}_0,f_*,\dot{f}_*)(X_i)dQ_i(s)\\
&-\int_t^{\tau}\frac{v\{m_*(s)\}}{B_2(s)}d\big\{\mathbf{c}(s)\big\}n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+\int_t^{\tau}\int_0^{\tau}D_1(u,ds)d\hat{\pi}_n(u)+o_p(1). \qquad \text{(A.30)}
\end{aligned}$$

Similar to the arguments in Step A7 of the proof of Theorem 3.1, we have

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n}\int_t^{\tau}\frac{v\{m_*(s)\}}{B_2(s)}\frac{1}{w_{31}(X_i)}\frac{1}{n}U_{L1}(m_*,\boldsymbol{\beta}_0,f_*,\dot{f}_*)(X_i)dQ_i(s)\\
=&\,n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\tilde{\mu}_{\mathbf{Z}^*,i}(t)g\big\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}_i+f_*(X_i)\big\}dM_i(s)+o_p(1), \qquad \text{(A.31)}
\end{aligned}$$

where

$$\tilde{\mu}_{\mathbf{Z}^*,i}(t) = \int_0^\tau \frac{I(s > t)v\{m_*(s)\}}{B_2(s)} \frac{E[dQ(s)|X = X_i]}{E[\int_0^\tau dQ(u)|X = X_i]}.$$

Note that the last term on the right side of (A.30) can be expressed as

$$\int_t^\tau \int_0^\tau D_1(u, ds) d\hat{\pi}_n(u) = -\int_0^\tau a(t, u) d\hat{\pi}_n(u), \tag{A.32}$$

where

$$a(t, u) = -\int_t^\tau D_1(u, ds). \tag{A.33}$$

Using (A.30), (A.31) and (A.32), combined with (A.3), we have

$$\hat{\pi}_n(t) - \int_0^\tau a(t, u) d\hat{\pi}_n(u) = \Phi_n(t) = n^{-1/2} \sum_{i=1}^n \varphi_i(t),$$

where

$$\begin{aligned} \varphi_i(t) = -\int_t^\tau \frac{v\{m_*(s)\}}{B_2(s)} g\{m_*(s) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\} dM_i(s) - \int_t^\tau \frac{v\{m_*(s)\}}{B_2(s)} d\{\mathbf{c}(s)\} \\ \times \mathbf{A}^{-1} \int_0^\tau \{\mathbf{Z}_i - \boldsymbol{\mu}_{\mathbf{Z}}(t) - (\mathbf{Z}_i^* - \boldsymbol{\mu}_{\mathbf{Z}^*,i})\} g\{m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\} dM_i(t) \\ + \int_0^\tau \tilde{\mu}_{\mathbf{Z}^*,i}(t) g\{m_*(s) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\} dM_i(s), \end{aligned} \tag{A.34}$$

which are independent mean zero functions. Thus, $\Phi_n(t)$ converges weakly to a mean zero Gaussian process as $n \to \infty$. □

We continue to prove Theorem 3.2. Using integration by parts, we rewrite (A.28) as a Fredholm integral equation of the second kind with the kernel $\frac{\partial a(t,s)}{\partial s}$. We assume that equation (A.28) has a unique solution, which can be assured by the condition

$$\sup_{t \in [0,\tau]} \int_0^\tau \left| \frac{\partial a(t, s)}{\partial s} \right| ds < \infty. \tag{A.35}$$

Next, we construct a solution to equation (A.28) as

$$\hat{\pi}_n(t) = \Phi_n(t) + \int_0^\tau b(t, s) d\Phi_n(s), \tag{A.36}$$

where $b(t, s)$ is the solution to the following equation

$$b(t, s) - \int_0^\tau a(t, u) \frac{\partial b(u, s)}{\partial u} du = a(t, s), \qquad t, s \in [0, \tau). \tag{A.37}$$

Notably, (A.37) can be rewritten as a Fredholm integral equation of the second kind with the kernel $\frac{\partial a(t,s)}{\partial s}$. Thus, equation (A.37) has a unique solution under condition (A.35). Then, we can verify that $\hat{\pi}_n(t)$ defined in (A.36) is a solution to the integral equation (A.28).

Using the representation of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ in (A.3), we have

$$\begin{aligned} & n^{1/2}\big[V\big\{\hat{m}_\#(t;\hat{\boldsymbol{\beta}})\big\} - V\big\{\hat{m}_\#(t;\boldsymbol{\beta}_0)\big\}\big] \\ = & \int_t^\tau \frac{v\{m_*(s)\}E[\mathbf{Z}dQ(s)]}{B_2(s)}\mathbf{A}^{-1} \\ & \times n^{-1/2}\sum_{i=1}^n \int_0^\tau \big\{\mathbf{Z}_i - \boldsymbol{\mu}_{\mathbf{Z}}(t) - \big(\mathbf{Z}_i^* - \boldsymbol{\mu}_{\mathbf{Z}^*,i}\big)\big\} g\big\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\big\} dM_i(t) \\ & + o_p(1). \end{aligned}$$

Together with (A.36), we have

$$n^{1/2}\big[V\big\{\hat{m}_\#(t;\hat{\boldsymbol{\beta}})\big\} - V\big\{m_*(t)\big\}\big] = n^{-1/2}\sum_{i=1}^n \kappa_i(t) + o_p(1),$$

where

$$\begin{aligned} \kappa_i(t) = \varphi_i(t) + \int_0^\tau b(t,s)d\varphi_i(s) + \int_t^\tau \frac{v\{m_*(s)\}E[\mathbf{Z}dQ(s)]}{B_2(s)}\mathbf{A}^{-1} \\ \times \int_0^\tau \big\{\mathbf{Z}_i - \boldsymbol{\mu}_{\mathbf{Z}}(t) - \big(\mathbf{Z}_i^* - \boldsymbol{\mu}_{\mathbf{Z}^*,i}\big)\big\} g\big\{m_*(t) + \boldsymbol{\beta}_0^T\mathbf{Z}_i + f_*(X_i)\big\} dM_i(t) \end{aligned}$$

are independent mean zero functions. Based on the functional delta method, we have

$$n^{1/2}\big\{\hat{m}_0(t) - m_*(t)\big\} = n^{-1/2}\sum_{i=1}^n \frac{\kappa_i(t)}{v\{m_*(t)\}} + o_p(1)$$

on $[0,\tau)$. Thus, Theorem 3.2 is proved. Considering $\kappa_i(t)$ $(i = 1, \dots, n)$ are independent and identically distributed zero mean random variables for each t, the multivariate central limit theorem implies that $n^{1/2}\{\hat{m}_0(t) - m_*(t)\}$ converges in finite-dimensional distribution to a zero-mean Gaussian process. In view of the tightness of $n^{1/2}\sum_{i=1}^n \kappa_i(t)$, it follows from the functional central limit theorem that $n^{1/2}\{\hat{m}_0(t) - m_*(t)\}$ converges weakly to a mean zero Gaussian process. □

Proof of Theorem 3.3. Based on the assumptions $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = O_p(1)$, $n^{1/2}\{\hat{m}_0(t) - m_*(t)\} = O_p(1)$, and the regularity conditions of Theorem 3.3, it can be shown that

$$\|n^{-1}\mathbf{U}_L(\hat{m}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}_0, \hat{\gamma}_1)(x) - n^{-1}\mathbf{U}_L(m_*, \boldsymbol{\beta}_0, \hat{\gamma}_0, \hat{\gamma}_1)(x)\| = O_p\big(n^{-1/2}\big).$$

Since $\mathbf{U}_L(\hat{m}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}_0, \hat{\gamma}_1)(x) = 0$, we have

$$n^{-1}\mathbf{U}_L(m_*, \boldsymbol{\beta}_0, \hat{\gamma}_0, \hat{\gamma}_1)(x) = O_p\big(n^{-1/2}\big) = o_p(1/\sqrt{nh}).$$

For convenience, set $\tilde{\boldsymbol{\gamma}}(x) = (\gamma_0(x), h\gamma_1(x))^T$, $\hat{\tilde{\boldsymbol{\gamma}}}(x) = (\hat{\gamma}_0(x), h\hat{\gamma}_1(x))^T$, and $\tilde{\mathbf{f}}_*(x) = (f_*(x), h\dot{f}_*(x))^T$. We first show that $\hat{\tilde{\boldsymbol{\gamma}}}(x) \to \tilde{\mathbf{f}}_*(x)$ in probability as $n \to \infty$. Note that

$$
\begin{aligned}
&\frac{\partial}{\partial\tilde{\gamma}}\mathbf{U}_L(\hat{m}_0, \hat{\boldsymbol{\beta}}_0, \gamma_0, \gamma_1)(x) \\
={}& \sum_{i=1}^n \int_0^\tau \begin{bmatrix} 1 \\ (X_i - x)/h \end{bmatrix}^{\otimes 2} K_h(X_i - x) \\
&\times \dot{g}\big\{\hat{m}_0(t) + \hat{\boldsymbol{\beta}}^T\mathbf{Z}_i + \gamma_0(x) + \gamma_1(x)(X_i - x)\big\} dN_i(t) \\
&- \sum_{i=1}^n \int_0^\tau \begin{bmatrix} 1 \\ (X_i - x)/h \end{bmatrix}^{\otimes 2} K_h(X_i - x) Y_i(t) \\
&\times d\dot{g}\big\{\hat{m}_0(t) + \hat{\boldsymbol{\beta}}^T\mathbf{Z}_i + \gamma_0(x) + \gamma_1(x)(X_i - x)\big\}
\end{aligned}
$$

is non-degenerate. Moreover, by the uniform law of large numbers and standard nonparametric techniques, we can show that $n^{-1}\frac{\partial}{\partial\tilde{\boldsymbol{\gamma}}}\mathbf{U}_L(\hat{m}_0, \boldsymbol{\beta}_0, \gamma_0, \gamma_1)(x)$ converges to a non-degenerate deterministic matrix

$$
\begin{aligned}
&\dot{\mathbf{u}}_{\tilde{\boldsymbol{\gamma}}}(m_*, \boldsymbol{\beta}_0, \gamma_0)(x) \\
={}& r(x)E\big[(\Delta - 1)\dot{g}\big\{m_*(\tilde{T}) + \boldsymbol{\beta}_0^T\mathbf{Z} + \gamma_0(x)\big\} + \dot{g}\{m_*(0) + \boldsymbol{\beta}_0^T\mathbf{Z} + \gamma_0(x)|X = x\big] \\
&\times \begin{bmatrix} 1 & 0 \\ 0 & k_2 \end{bmatrix}.
\end{aligned}
$$

Again by the law of large numbers, and the consistency of $\hat{m}_0(\cdot)$ and $\hat{\boldsymbol{\beta}}$, we can show

$$
n^{-1}\mathbf{U}_L(\hat{m}_0, \hat{\boldsymbol{\beta}}, \gamma_0, \gamma_1)(x) \to \mathbf{u}(m_*, \boldsymbol{\beta}_0, \gamma_0)(x),
$$

where $\mathbf{u}(m_*, \boldsymbol{\beta}_0, \gamma_0)(x)$ equals

$$
\begin{bmatrix} r(x) \\ 0 \end{bmatrix} E\big[(\Delta - 1)g\big\{m_*(\tilde{T}) + \boldsymbol{\beta}_0^T\mathbf{Z} + \gamma_0(x)\big\} + g\big\{m_*(0) + \boldsymbol{\beta}_0^T\mathbf{Z} + \gamma_0(x)\big\} - \tilde{T}|X = x\big].
$$

Note that $\mathbf{u}(m_*, \boldsymbol{\beta}_0, f_*)(x) = 0$. Following the arguments of [13], the consistency of $\hat{\tilde{\boldsymbol{\gamma}}}(x)$ holds. Using the Taylor expansion, we obtain

$$
\begin{aligned}
&n^{-1}\mathbf{U}_L(m_*, \boldsymbol{\beta}_0, \hat{\tilde{\boldsymbol{\gamma}}})(x) - n^{-1}\mathbf{U}_L(m_*, \boldsymbol{\beta}_0, \tilde{\mathbf{f}}_*)(x) \\
={}& n^{-1}\left[\frac{\partial}{\partial\tilde{\boldsymbol{\gamma}}}\mathbf{U}_L(m_*, \boldsymbol{\beta}_0, \tilde{\boldsymbol{\gamma}})(x)\right]_{\tilde{\boldsymbol{\gamma}} = \tilde{\mathbf{f}}_*} \big\{\hat{\tilde{\boldsymbol{\gamma}}}(x) - \tilde{\mathbf{f}}_*(x)\big\} + o_p(1/\sqrt{nh}). \qquad \text{(A.38)}
\end{aligned}
$$

Note that we have shown $n^{-1}\mathbf{U}_L(m_*, \boldsymbol{\beta}_0, \hat{\tilde{\boldsymbol{\gamma}}})(x) = o_p(1/\sqrt{nh})$. By the strong law of large numbers, we have

$$
\boldsymbol{\Omega}_1(x) \triangleq \lim_{n\to\infty} n^{-1}\left[\frac{\partial}{\partial\tilde{\boldsymbol{\gamma}}}\mathbf{U}_L(m_*, \boldsymbol{\beta}_0, \tilde{\boldsymbol{\gamma}})(x)\right]_{\tilde{\boldsymbol{\gamma}} = \tilde{\mathbf{f}}_*} = \dot{\mathbf{u}}_{\tilde{\boldsymbol{\gamma}}}(m_*, \boldsymbol{\beta}_0, f_*)(x).
$$

It is easily verified that

$$\mathbf{U}_L(m_*, \boldsymbol{\beta}_0, \tilde{\mathbf{f}}_*)(x) = \mathbf{II}_1(x) + \mathbf{II}_2(x),$$

where

$$\mathbf{II}_1(x) = \sum_{i=1}^{n} \int_0^{\tau} \left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x) g\big\{m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\} dM_i(t),$$

and

$$\begin{aligned}
\mathbf{II}_2(x) = &\sum_{i=1}^{n} (\Delta_i - 1)\left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x) \\
&\times g\big\{m_*(\tilde{T}_i) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(x) + \dot{f}_*(x)(X_i - x)\big\} \\
&- \sum_{i=1}^{n} (\Delta_i - 1)\left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x) g\big\{m_*(\tilde{T}_i) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\} \\
&+ \sum_{i=1}^{n} \left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x) g\big\{m_*(0) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(x) + \dot{f}_*(x)(X_i - x)\big\} \\
&- \sum_{i=1}^{n} \left(1, \frac{X_i - x}{h}\right)^T K_h(X_i - x) g\big\{m_*(0) + \boldsymbol{\beta}_0^T \mathbf{Z}_i + f_*(X_i)\big\}.
\end{aligned}$$

Following the proof of Theorem 4 of [2], and by the martingale central limit theorem, it follows that

$$\big(n^{-1}h\big)^{1/2} \mathbf{II}_1(x) \to N\big(0, \boldsymbol{\Omega}_2(x)\big) \tag{A.39}$$

in distribution, where

$$\boldsymbol{\Omega}_2(x) = r(x) \begin{bmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{bmatrix} E\left[\int_0^{\tau} g^2\big\{m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z} + f_*(x)\big\} dN(t) | X = x\right],$$

with $\nu_0 = \int [K(x)]^2 dx$ and $\nu_2 = \int x^2 [K(x)]^2 dx$. Using the Taylor expansion around x, we obtain

$$n^{-1} \mathbf{II}_2(x) = -\boldsymbol{\Omega}_1(x) \mathbf{b}_n(x) + o_p\big(h^2\big),$$

where

$$\begin{aligned}
\mathbf{b}_n(x) = &\frac{h^2}{2} \ddot{f}_*(x) r(x) \boldsymbol{\Omega}_1(x)^{-1} \begin{bmatrix} k_2 \\ 0 \end{bmatrix} E\big[(\Delta - 1)\dot{g}\big\{m_*(\tilde{T}) + \boldsymbol{\beta}_0^T \mathbf{Z} + f_*(x)\big\} \\
&+ \dot{g}\big\{m_*(0) + \boldsymbol{\beta}_0^T \mathbf{Z} + f_*(x)\big\} | X = x\big].
\end{aligned}$$

Combining with (A.38), we have

$$(nh)^{1/2} \boldsymbol{\Omega}_1(x) \big\{\hat{\tilde{\boldsymbol{\gamma}}}(x) - \tilde{\mathbf{f}}_*(x) - \mathbf{b}_n(x) + o_p\big(h^2\big)\big\} = -\big(n^{-1}h\big)^{1/2} \mathbf{II}_1(x) + o_p(1).$$

Since nh^5 is bounded, and using (A.39), Theorem 3.3 is proved. □

Proof of Theorem 3.4. The proof of Theorem 3.4 is similar to the proof of Theorem 3.1, and hence is omitted here. □

Appendix B: One step estimator and its properties

To implement the proposed algorithm in Section 2, we need an initial estimator of the nonparametric component $f(\cdot)$. Following [1, 4], we propose to use the one-step estimator as the initial value. Specifically, we consider the following local estimating equation of $m(\cdot)$ for fixed $\boldsymbol{\beta}$ and $\gamma_1(x)$ and covariate value $X = x$ in the compact support of X

$$\begin{aligned}&\sum_{i=1}^{n}\int_0^{\tau} K_h(X_i - x) g\big\{m(t) + \boldsymbol{\beta}^T \mathbf{Z}_i + \gamma_1(x)(X_i - x)\big\} dN_i(t)\\&\quad - \sum_{i=1}^{n}\int_0^{\tau} K_h(X_i - x) Y_i(t) d\Big(g\big\{m(t) + \boldsymbol{\beta}^T \mathbf{Z}_i + \gamma_1(x)(X_i - x)\big\} + t\Big) = 0.\end{aligned} \tag{B.1}$$

Let $\tilde{m}_x(t;\boldsymbol{\beta},\gamma_1)$ be the estimator of $m(t)$ derived from (B.1), we propose the following estimating function for $\boldsymbol{\beta}$ and $\gamma_1(x)$ as

$$\begin{aligned}\tilde{\mathbf{U}}_x(\boldsymbol{\beta},\gamma_1) = \sum_{i=1}^{n}\int_0^{\tau} &\big(\mathbf{Z}_i^T, X_i - x\big)^T K_h(X_i - x)\big[g\big\{\tilde{m}_x(t;\boldsymbol{\beta},\gamma_1) + \boldsymbol{\beta}^T \mathbf{Z}_i\\&+ \gamma_1(x)(X_i - x)\big\} dN_i(t) - Y_i(t) d\Big(g\big\{\tilde{m}_x(t;\boldsymbol{\beta},\gamma_1) + \boldsymbol{\beta}^T \mathbf{Z}_i\\&+ \gamma_1(x)(X_i - x)\big\} + t\Big)\Big].\end{aligned} \tag{B.2}$$

Solving for $\tilde{\mathbf{U}}_x(\boldsymbol{\beta},\gamma_1) = 0$, we obtained the estimators $\tilde{\boldsymbol{\beta}}(x)$ and $\tilde{\gamma}_1(x)$ of $\boldsymbol{\beta}$ and $\gamma_1(x)$, respectively. $\tilde{m}_x(t) = \tilde{m}_x(t;\tilde{\boldsymbol{\beta}}(x),\tilde{\gamma}_1(x))$ is the estimator of $m(t)$. The estimator of $f(x)$ can be constructed as $\tilde{\gamma}_0(x) = \int_0^x \tilde{\gamma}_1(u)du$.

Note that the intercept term $\gamma_0(t)$ is absorbed into the function $m(t)$ because of the local nature of (B.1) and (B.2). As discussed in [1, 4], the final estimator based on full iteration of the estimating equations (2.3), (2.4) and (2.7) are at least as efficient as the one-step estimators. To establish the local consistency of the one-step estimators $\tilde{\boldsymbol{\beta}}(x)$, $\tilde{\gamma}_0(x)$ and $\tilde{\gamma}_1(x)$, we first define

$$\mathbf{A}_x = \begin{bmatrix}\mathbf{A}_{\boldsymbol{\beta},x} & \mathbf{0}\\ \mathbf{0}^T & A_{\gamma_1,x}\end{bmatrix},$$

where

$$A_{\gamma_1,x} = k_2 r(x) E\left[\int_0^{\tau} dQ_x(t) | X = x\right]$$

and

$$\mathbf{A}_{\boldsymbol{\beta},x} = r(x) E\left[\int_0^{\tau}\big\{\mathbf{Z} - \boldsymbol{\mu}_{1x}(t) - \boldsymbol{\mu}_{2x}(t)\big\} dQ_x(t) | X = x\right]$$

with $k_2 = \int x^2 K(x)dx$, $r(x)$ as the marginal density of X at x, and

$$dQ_x(t) = \dot{g}\big\{m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z} + f_*(x)\big\} dN(t) - Y(t) d\dot{g}\big\{m_*(t) + \boldsymbol{\beta}_0^T \mathbf{Z} + f_*(x)\big\},$$

$$B_x(t,s) = \exp\Big(-\int_s^t \frac{E[dQ_x(u)|X=x]}{E[Y(u)\dot{g}\{m_*(u)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(x)\}|X=x]}\Big),$$

$$\boldsymbol{\mu}_{1x}(t) = \frac{E[(\Delta-1)\mathbf{Z}\dot{g}\{m_*(\tilde{T})+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(x)\}B_x(t,\tilde{T})I(\tilde{T}<t)|X=x]}{E[Y(t)\dot{g}\{m_*(t)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(x)\}|X=x]},$$

$$\boldsymbol{\mu}_{2x}(t) = \frac{E[\mathbf{Z}\dot{g}\{m_*(0)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(x)\}B_x(t,0)|X=x]}{E[Y(t)\dot{g}\{m_*(t)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(x)\}|X=x]}.$$

Lemma B1. *Under the regularity conditions (C1)–(C6), and the condition that $\boldsymbol{A}_x$ is finite and nondegenerate for any x in the compact support of X, if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then the one-step estimators $\tilde{\boldsymbol{\beta}}(x)$, $\tilde{\gamma}_0(x)$ and $\tilde{\gamma}_1(x)$ are locally consistent.*

Proof of Lemma B. Following the arguments in the proof of Theorem 3.1 that shows the consistency of $\hat{m}_0(t;\boldsymbol{\beta}_0,f_*)$ to $m_*(t)$, we can show that $\tilde{m}_x(t;\boldsymbol{\beta}_0,\dot{f}_*)$ converges almost surely to $m_*(t)+f_*(x)$ on $[0,\tau]$. Plugging $\tilde{m}_x(t;\boldsymbol{\beta},\gamma_1)$ into the left side of (B.1), and taking derivatives with respect to $\boldsymbol{\beta}$ and γ_1, following the arguments as in showing (A.6), we have, for $t \in [0,\tau]$,

$$\frac{\partial \tilde{m}_x(t;\boldsymbol{\beta},\gamma_1)}{\partial \boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0,\gamma_1=\dot{f}_*} = -\int_t^\tau \frac{B_x(s,t)E[\mathbf{Z}dQ_x(s)|X=x]}{E[Y(s)\dot{g}\{m_*(s)+\boldsymbol{\beta}_0^T\mathbf{Z}+f_*(x)\}|X=x]} + o_p(1),$$

and

$$\frac{\partial \tilde{m}_x(t;\boldsymbol{\beta},\gamma_1)}{\partial \gamma_1}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0,\gamma_1=\dot{f}_*} = o_p(1).$$

It can be checked that $n^{-1}\tilde{\mathbf{U}}_x(\boldsymbol{\beta},\gamma_1)$ converges almost surely to a deterministic function $\tilde{\mathbf{u}}_x(\boldsymbol{\beta},\gamma_1)$ for $(\boldsymbol{\beta}^T,\gamma_1)^T$ in a small neighborhood of $(\boldsymbol{\beta}_0^T,\dot{f}_*)^T$, and $\tilde{\mathbf{u}}_x(\boldsymbol{\beta}_0,\dot{f}_*)=0$. Moreover, we can check that

$$\frac{1}{n}\frac{\partial \tilde{\mathbf{U}}_x(\boldsymbol{\beta},\gamma_1)}{\partial(\boldsymbol{\beta}^T,\gamma_1)}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0,\gamma_1=\dot{f}_*} = \mathbf{A}_x + o_p(1).$$

By the conditions of Lemma B, $\mathbf{A}_x$ is finite and nondegenerate. Thus, $\frac{1}{n}\frac{\partial \tilde{\mathbf{U}}_x(\boldsymbol{\beta},\gamma_1)}{\partial(\boldsymbol{\beta}^T,\gamma_1)}$ converges to $\mathbf{A}_x$ in an arbitrarily small neighborhood of $(\boldsymbol{\beta}_0,\dot{f}_*)$. Using the fact that $n^{-1}\tilde{\mathbf{U}}_x(\boldsymbol{\beta}_0,\dot{f}_*) \to 0$ as $n \to \infty$ and $\tilde{\mathbf{U}}_x(\tilde{\boldsymbol{\beta}}(x),\tilde{\gamma}_1(x)) = 0$, we have the local consistency of $\tilde{\boldsymbol{\beta}}(x)$ and $\tilde{\gamma}_1(x)$. The local consistency of $\tilde{\gamma}_0(x)$ follows from that of $\tilde{\gamma}_1(x)$. □

Funding

The research of Haijin He was supported in part by the National Natural Science Foundation of China (NSFC) (Grant No. 11701387) and the Science and Technology Planning Project of Shenzhen Municipality, PR China (Grant No. 20220810171450002). The research of Jingheng Cai was supported by NSFC (Grant No. 11771463). The research of Xinyuan Song was supported by the Research Grant Council of the HKSAR (No. 14302220).

References

[1] Cai, J., Fan, J., Jiang, J. and Zhou, H. (2007). Partially linear hazard regression for multivariate survival data. *Journal of the American Statistical Association* **102** 538–551. MR2370851

[2] Cai, J., Fan, J., Jiang, J. and Zhou, H. (2008). Partially linear hazard regression with varying coefficients for multivariate survival data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 141–158. MR2412635

[3] Cai, J., He, H., Song, X. and Sun, L. (2017). An additive-multiplicative mean residual life model for right-censored data. *Biometrical Journal* **59** 579–592. MR3648615

[4] Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association* **92** 477–489. MR1467842

[5] Carroll, R. J., Ruppert, D. and Welsh, A. H. (1998). Local estimating equations. *Journal of the American Statistical Association* **93** 214–227. MR1614624

[6] Chen, Y., Jewell, N., Lei, X. and Cheng, S. (2005). Semiparametric estimation of proportional mean residual life model in presence of censoring. *Biometrics* **61** 170–178. MR2135857

[7] Chen, Y. Q. (2007). Additive expectancy regression. *Journal of the American Statistical Association* **102** 153–166. MR2345536

[8] Chen, Y. Q. and Cheng, S. (2005). Semiparametric regression analysis of mean residual life with censored survival data. *Biometrika* **92** 19–29. MR2158607

[9] Chen, Y. Q. and Cheng, S. (2006). Linear life expectancy regression with censored data. *Biometrika* **93** 303–313. MR2278085

[10] Cortese, G., Holmboe, S. A. and Scheike, T. H. (2017). Regression models for the restricted residual mean life for right-censored and left-truncated data. *Statistics in Medicine* **36** 1803–1822. MR3648623

[11] Cuzick, J. (1992). Semiparametric additive regression. *Journal of the Royal Statistical Society: Series B* **54** 831–843. MR1185225

[12] Engle, R. F., Granger, C. W., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association* **81** 310–320.

[13] Fan, J., Gijbels, I. and King, M. (1997). Local likelihood and local partial likelihood in hazard regression. *The Annals of Statistics* **25** 1661–1690. `https://doi.org/10.1214/aos/1031594736` MR1463569

[14] Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association* **87** 942–951.

[15] Härdle, W., Liang, H. and Gao, J. (2004). *Partially Linear Models*. Springer-Verlag, Heidelberg. MR1787637

[16] He, H., Cai, J., Song, X. and Sun, L. (2017). Analysis of proportional mean residual life model with latent variables. *Statistics in Medicine* **36**

813–826. MR3597658

[17] He, H., Pan, D., Song, X. and Sun, L. (2019). Additive mean residual life model with latent variables under right censoring. *Statistica Sinica* **29** 47–66. MR3889356

[18] Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *The Annals of Statistics* **27** 1536–1563. MR1742499

[19] Jin, Z., Ying, Z. and Wei, L. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88** 381–390. MR1844838

[20] Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons. MR1924807

[21] Liang, H., Härdle, W. and Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics* **27** 1519–1535. MR1742498

[22] Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data. *Biometrika* **88** 1179–1185. MR1872228

[23] Lu, M. and McMahan, C. S. (2018). A partially linear proportional hazards model for current status data. *Biometrics* **74** 1240–1249. MR3908142

[24] Lu, W. and Zhang, H. H. (2007). Variable selection for proportional odds model. *Statistics in Medicine* **26** 3771–3781. MR2395832

[25] Lu, W. and Zhang, H. H. (2010). On estimation of partially linear transformation models. *Journal of the American Statistical Association* **105** 683–691. MR2724852

[26] Luk, A. O., So, W. Y., Ma, R. C., Kong, A. P., Ozaki, R., Ng, V. S., Yu, L. W., Lau, W. W., Yang, X., Chow, F. C., Chan, J. C. and Tong, P. C. (2008). Metabolic syndrome predicts new onset of chronic kidney disease in 5,829 patients with type 2 diabetes: A 5-year prospective analysis of the Hong Kong Diabetes Registry. *Diabetes Care* **31** 2357–2361.

[27] Ma, S. 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. MR2211086

[28] Maguluri, G. and Zhang, C. H. (1994). Estimation in the mean residual life regression model. *Journal of the Royal Statistical Society: Series B* **56** 477–489. MR1278221

[29] Mansourvar, Z., Martinussen, T. and Scheike, T. H. (2015). Semiparametric regression for restricted mean residual life under right censoring. *Journal of Applied Statistics* **42** 2597-2613. MR3428833

[30] Mansourvar, Z., Martinussen, T. and Scheike, T. H. (2016). An additive–multiplicative restricted mean residual life model. *Scandinavian Journal of Statistics* **43** 487–504. MR3503013

[31] Oakes, D. and Dasu, T. (1990). A note on residual life. *Biometrika* **77** 409–410. MR1064816

[32] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

[33] Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society: Series B* **50** 413–436. MR0970977

[34] Sun, L., Song, X. and Zhang, Z. (2012). Mean residual life models with

time-dependent coefficients under right censoring. *Biometrika* **99** 185–197. MR2899672

[35] Sun, L. and Zhang, Z. (2009). A class of transformed mean residual life models with censored survival data. *Journal of the American Statistical Association* **104** 803–815. MR2541596

[36] Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in Medicine* **16** 385–395.