Supplementary Appendix for
"On Estimation of Partially Linear Transformation Models"

Appendix A: One-step estimator and its properties

To implement the proposed computational algorithm in Section 2.2, we need an initial estimator $f^{(0)}(\cdot)$ of the nonparametric component $f(\cdot)$. Following the idea of Carroll et al. (1997) and Cai eta l. (2007), we propose to use the following *one-step* estimator as the initial value. To be specific, we consider the following local estimating equations of H, $\boldsymbol{\beta}$ and γ_1 for any given covariate value $X = x \in \mathcal{X}$:

$$\sum_{i=1}^{n} K_h(X_i - x)\Big[dN_i(t) - Y_i(t)d\Lambda\{H(t) + \boldsymbol{\beta}'\mathbf{Z}_i + \gamma_1(x)(X_i - x)\}\Big] = 0,\ t \geq 0, \quad \text{(A.1)}$$

$$\sum_{i=1}^{n} \int_0^{\tau} \begin{pmatrix} \mathbf{Z}_i \\ X_i - x \end{pmatrix} K_h(X_i - x)\Big[dN_i(t) - Y_i(t)d\Lambda\{H(t) + \boldsymbol{\beta}'\mathbf{Z}_i + \gamma_1(x)(X_i - x)\}\Big] = 0. \quad \text{(A.2)}$$

Note that the intercept parameter $\gamma_0(x)$ previously appearing in (6) and (7) is absorbed into the function $H(\cdot)$ because of the local nature of equation (A.1). Let $\tilde{H}_x(\cdot)$, $\tilde{\boldsymbol{\beta}}(x)$ and $\tilde{\gamma}_1(x)$ denote the resulting solutions of (A.1) and (A.2). Then the estimator of $f(x)$ can be constructed as $\tilde{\gamma}_0(x) = \int_0^x \tilde{\gamma}_1(u)du$.

As discussed by Carroll et al. (1997) and Cai et al. (2007), the final estimator based on *full iterations* of the estimating equations (4)-(7) are at least as efficient as the one-step estimators. In the next lemma, we establish the local consistency of the *one-step* estimators $\tilde{\beta}(x)$, $\tilde{\gamma}_0(x)$ and $\tilde{\gamma}_1(x)$.

Lemma 1. *Under the regularity conditions (C1)-(C5), if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, we have that the one-step estimators $\tilde{\boldsymbol{\beta}}(x)$, $\tilde{\gamma}_0(x)$ and $\tilde{\gamma}_1(x)$ are locally consistent provided that the matrix $\boldsymbol{A}_x$ is finite and nondegenerate for any $x \in \mathcal{X}$.*

The definition of $\boldsymbol{A}_x$ and the proof of Lemma 1 will be given in the Appendix B.

Appendix B: Proofs of Theorems

Throughout the Appendix B, the notion $||\cdot||$ denotes the L^2 norm of a vector. Define a ϵ-ball in R^p centered at β_0 as $D^1_\varepsilon = \{\boldsymbol{\beta} : ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \leq \varepsilon\}$. Moreover, define $D^2_\varepsilon = \{f : \sup_{x\in\mathcal{X}} |f(x) - f_0(x)| \leq \varepsilon\}$, where $\mathcal{X}$ is the bounded support of X.

Proof of Lemma 1. Given $\boldsymbol{\beta}$ and γ_1, the left-hand side of (A.1) is monotone in H. Thus, the solution to (A.1) is unique if there exists one. Let $\tilde{H}_x(\cdot;\boldsymbol{\beta},\gamma_1)$ denote the resulting solution. By using the similar arguments of Chen et al. (2002), it is easy to show that the solution $\tilde{H}_x(\cdot;\boldsymbol{\beta},\gamma_1)$ exists if $\boldsymbol{\beta}$ is in a small neighborhood of $\boldsymbol{\beta}_0$ and $\gamma_1(x)$ is bounded on $\mathcal{X}$. Moreover, we can show that $\tilde{H}_x(\cdot;\boldsymbol{\beta}_0,\gamma_1)$ converges almost surely to the function $H_0(t) + f_0(x)$ on $[0,\tau]$. Plugging $\tilde{H}_x(\cdot;\boldsymbol{\beta},\gamma_1)$ into (A.1) and taking the derivative of the resulting estimating function with respect to $\boldsymbol{\beta}$ and γ_1, we have, for $t \in [0,\tau]$,

$$\lim_{n\to\infty} \frac{\partial \tilde{H}_x(t;\boldsymbol{\beta},\gamma_1)}{\partial \gamma_1} |_{\boldsymbol{\beta}=\boldsymbol{\beta}_0,\gamma_1=\dot{f}_0} = 0 \text{ and } \lim_{n\to\infty} \frac{\partial \tilde{H}_x(t;\boldsymbol{\beta},\gamma_1)}{\partial \boldsymbol{\beta}} |_{\boldsymbol{\beta}=\boldsymbol{\beta}_0,\gamma_1=\dot{f}_0} = -\boldsymbol{a}_x(t),$$

where

$$\boldsymbol{a}_x(t) = \int_0^t \frac{\lambda^*_x\{H_0(s)\}E[\mathbf{Z}\dot{\lambda}\{H_0(s) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(x)\}Y(s)|X=x]}{\lambda^*_x\{H_0(t)\}B_{2x}(s)} dH_0(s),$$

$$B_{2x}(t) = E[\lambda\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(x)\}Y(t)|X=x],$$

$$B_x(t,s) = \exp\Big(\int_s^t \frac{E[\dot{\lambda}\{H_0(u) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(x)\}Y(u)|X=x]}{E[\lambda\{H_0(u) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(x)\}Y(u)|X=x]} dH_0(u)\Big),$$

and $\lambda^*_x\{H_0(t)\} = B_x(t,\zeta_{0x})$. The constant ζ_{0x} is defined similarly as ζ_0 given in Section 2.3. The above asymptotic representation of the derivatives can be derived using the similar empirical process techniques given in the appendix of Chen et al. (2002), which is omitted here. Note that $\tilde{H}_x(t;\boldsymbol{\beta},\gamma_1)$ is monotone increasing in t. Therefore, $\tilde{H}_x(t;\boldsymbol{\beta}_n,\gamma_{1n})$ converges uniformly to $H_0(t) + f_0(x)$ on $[0,\tau]$ provided $\boldsymbol{\beta}_n$ converges to $\boldsymbol{\beta}_0$ and γ_{1n} is bounded.

We plug $\tilde{H}_x(\cdot;\boldsymbol{\beta},\gamma_1)$ into (A.2), and denote n^{-1} multiplying the left-hand side of the resulting estimating function by $\tilde{\boldsymbol{U}}_x(\boldsymbol{\beta},\gamma_1)$. By the monotonicity of $\tilde{H}_x(\cdot;\boldsymbol{\beta},\gamma_1)$, the law of large numbers and some standard nonparametric techniques, we have that $\tilde{\boldsymbol{U}}_x(\boldsymbol{\beta}, \gamma_1)$ converges almost surely to a deterministic vector $\tilde{\boldsymbol{u}}_x(\boldsymbol{\beta}, \gamma_1)$ for β in a small neighborhood of $\boldsymbol{\beta}_0$ and $\gamma_1(x)$ in a small neighborhood of $\dot{f}_0(x)$. It is easy to show that $\tilde{\boldsymbol{u}}_x(\boldsymbol{\beta}_0, \dot{f}_0) = 0$. Moreover, we can show that

$$\boldsymbol{A}_x \equiv \lim_{n\to\infty} \frac{\partial \tilde{\boldsymbol{U}}_x(\boldsymbol{\beta}, \gamma_1)}{\partial(\boldsymbol{\beta}', \gamma_1)} \mid_{\boldsymbol{\beta}=\boldsymbol{\beta}_0, \gamma_1=\dot{f}_0} = - \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\beta},x} & \mathbf{0} \\ \mathbf{0}' & A_{\gamma_1,x} \end{pmatrix},$$

where $A_{\gamma_1,x} = k_2 g(x) E[\lambda\{H_0(\tilde{T}) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(X)\}|X = x]$ with $k_2 = \int x^2 K(x)dx$, $g(x)$ is the marginal density of X at x, and

$$\boldsymbol{A}_{\boldsymbol{\beta},x} = g(x)E\left[\lambda\{H_0(\tilde{T}) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(X)\}\mathbf{Z}\{\mathbf{Z} - \boldsymbol{a}_x(\tilde{T})\}'|X = x\right].$$

By the assumption, the limiting matrix $\boldsymbol{A}_x$ is finite and nondegenerate. Following the similar techniques of Chen et al. (2002), we have, as $n \to \infty$ and then ε_1, $\varepsilon_x \to 0$

$$\sup_{\boldsymbol{\beta}\in D^1_{\varepsilon_1}, \gamma_1 \in D^3_{\varepsilon_x}} || \frac{\partial \tilde{\boldsymbol{U}}_x(\boldsymbol{\beta}, \gamma_1)}{\partial(\boldsymbol{\beta}', \gamma_1)} - \boldsymbol{A}_x|| \to 0 \qquad \text{(A.3)}$$

in probability, where $D^3_{\varepsilon_x} = \{\gamma_1 : |\gamma_1 - \dot{f}_0(x)| \le \varepsilon_x\}$. Consider $\tilde{\boldsymbol{U}}_x(\boldsymbol{\beta}, \gamma_1)$ as a random mapping from an arbitrarily small but fixed ball $D_\varepsilon \equiv \{(\boldsymbol{\beta}, \gamma_1) : ||(\boldsymbol{\beta}, \gamma_1) - (\boldsymbol{\beta}_0, \dot{f}_0(x))|| \le \varepsilon\}$ to another open connected set in R^{p+1}. Since A_x is finite and nondegenerate, (A.3) implies that, with probability tending to 1, the mapping $\tilde{\boldsymbol{U}}_x(\boldsymbol{\beta}, \gamma_1)$ is a homeomorphism from D_ε to its image, denoted as B_n. The convergence of $\tilde{\boldsymbol{U}}_x(\boldsymbol{\beta}_0, \dot{f}_0)$ to zero indicates that B_n contains $\mathbf{0} \in R^{p+1}$ with probability tending to 1. This proves that $\tilde{\boldsymbol{\beta}}(x)$ and $\tilde{\gamma}_1(x)$ are locally consistent since $\tilde{\boldsymbol{U}}_x(\tilde{\boldsymbol{\beta}}, \tilde{\gamma}_1) = 0$ and D_ε is centered at $(\boldsymbol{\beta}_0, \dot{f}_0(x))$ and can be arbitrarily small. The local consistency of $\tilde{\gamma}_1(x)$ on $\mathcal{X}$ further implies the local consistency of $\tilde{\gamma}_0(x)$. □

Proof of Theorem 1. To establish the asymptotic results in Theorem 1, we first need to prove the consistency of $\hat{\boldsymbol{\beta}}$, $\hat{f}(\cdot)$ and $\hat{H}(\cdot)$ obtained from the estimating equations (4)-(7). The global consistency of the estimators is extremely hard to derive due to the complex features of the proposed global and local estimating equations.

Here instead, we prove the local consistency of the estimators. In other words, we consider a small neighborhood of the true parameters $\boldsymbol{\beta}_0$ and f_0, and show that the corresponding estimating equations of $\boldsymbol{\beta}$ and f have unique solutions in this neighborhood. Moreover, for $\boldsymbol{\beta}$ and f being in this neighborhood, we will show that the estimating equations of H can produce an estimate close to the true H_0.

We have established the local consistency of the *one-step* estimators in Lemma A.1. As we discussed before, we can use the *one-step* estimators as the initial estimators for the *fully-iterated* estimators. Following the discussion of Carroll et al. (1997), it is expected that the *fully-iterated* estimators $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}_0(x)$ and $\hat{\gamma}_1(x)$ are also locally consistent.

We also note that, for fixed $\boldsymbol{\beta}$ and f, the estimating equation (4) is strictly monotone in H. In the following, we show that (4) has a unique solution. To see this, by the strong law of large numbers, there exist $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ such that for large n, $\boldsymbol{\beta} \in D^1_{\varepsilon_1}$, $f \in D^2_{\varepsilon_2}$, and $t \in [0, \tau]$,

$$\frac{1}{n}\sum_{i=1}^{n}\Big(N_i(t) - \Lambda[H_0\{\min(t, \tilde{T}_i)\} + a + \boldsymbol{\beta}' Z_i + f(X_i)]\Big) < 0, \tag{A.4}$$

$$\frac{1}{n}\sum_{i=1}^{n}\Big(N_i(t) - \Lambda[H_0\{\min(t, \tilde{T}_i)\} - a + \boldsymbol{\beta}' Z_i + f(X_i)]\Big) > 0, \tag{A.5}$$

for sufficiently large a. Thus, there exists a unique solution, denoted by $\hat{H}(t; \boldsymbol{\beta}, f)$, solving the following equation

$$\frac{1}{n}\sum_{i=1}^{n}\Big(N_i(t) - \Lambda[\hat{H}\{\min(t, \tilde{T}_i); \boldsymbol{\beta}, f\} + \boldsymbol{\beta}'\mathbf{Z}_i + f(X_i)]\Big) = 0.$$

Take the derivative with respect to t, we have

$$\frac{1}{n}\sum_{i=1}^{n}[dN_i(t) - Y_i(t)d\Lambda\{\hat{H}(t; \boldsymbol{\beta}, f) + \boldsymbol{\beta}'\mathbf{Z}_i + f(X_i)\}] = 0,$$

which implies equation (4) has a unique solution $\hat{H}(\cdot; \boldsymbol{\beta}, f)$ when $\boldsymbol{\beta} \in D^1_{\varepsilon_1}$ and $f \in D^2_{\varepsilon_2}$. Furthermore, since (A.4) and (A.5) hold for any $a > 0$ if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and

$f(\cdot) = f_0(\cdot)$, we have that $\hat{H}(t;\boldsymbol{\beta}_0, f_0)$ converges to $H_0(t)$ almost surly for $t \in [0,\tau]$. Due to the monotonicity of $\hat{H}(t;\boldsymbol{\beta}_0, f_0)$ and $H_0(t)$, the point-wise convergence can be further strengthened to the uniform convergence by applying the Glivenko-Cantelli Theorem (Shorack and Wellner, 1996).

Following the similar derivation in the proof of Lemma A.1. and applying some empirical process techniques, we can show that $\frac{\partial \hat{H}(t;\boldsymbol{\beta},\gamma_0)}{\partial \boldsymbol{\beta}}$ and $\frac{\partial \hat{H}(t;\boldsymbol{\beta},\gamma_0)}{\partial \gamma_0}$ are uniformly bounded on $[0,\tau]$ for $\boldsymbol{\beta} \in D^1_{\varepsilon_1}$ and $\gamma_0(\cdot) \in D^2_{\varepsilon_2}$. Note that $\hat{H}(t;\boldsymbol{\beta},\gamma_0)$ is monotone increasing in t. Therefore, the consistency of $\hat{H}(\cdot) \equiv \hat{H}(\cdot;\hat{\boldsymbol{\beta}},\hat{\gamma}_0)$ holds provided that $\hat{\boldsymbol{\beta}}$ and $\hat{\gamma}_0(\cdot)$ are respectively consistent estimators of $\boldsymbol{\beta}_0$ and $f_0(\cdot)$.

Given the consistency of the estimators $\hat{\boldsymbol{\beta}}$, $\hat{H}$, $\hat{\gamma}_0$ and $\hat{\gamma}_1$, we now establish the following asymptotic representation of $\hat{\boldsymbol{\beta}}$:

$$\begin{aligned} & n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ = \quad & (\boldsymbol{A}_1 - \boldsymbol{A}_2)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} [\{\mathbf{Z}_i - \boldsymbol{m}_{\mathbf{Z}}(t)\} - (\mathbf{Z}_i^* - \boldsymbol{m}_{\mathbf{Z}^*,i})] dM_i(t) + o_p(1), \end{aligned} \tag{A.6}$$

where $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\mathbf{Z}_i^*$ and $\boldsymbol{m}_{\mathbf{Z}^*,i}$ are defined in Section 2.3. Assuming (A.6) holds, then by the martingale central limit theorem and the regularity conditions given in Theorem 2.1, $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to a normal random vector with the mean zero and the variance-covariance matrix $\boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})'$ as n goes to infinity. In the following, we prove (A.6) in seven steps.

Step 1. From (3) and (4), following Step A2 of the proof in Chen et al. (2002), we have

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} dM_i(t) &= \frac{1}{n}\sum_{i=1}^{n} dN_i(t) - \frac{1}{n}\sum_{i=1}^{n} Y_i(t) d\Lambda\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\} \\ = \quad & \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i(t)\lambda\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\}}{\lambda^*\{H_0(t)\}} d\left[\Lambda^*\{\hat{H}(t;\boldsymbol{\beta}_0, f_0)\} - \Lambda^*\{H_0(t)\}\right] + \\ & \frac{1}{n}\sum_{i=1}^{n} Y_i(t)[\Lambda^*\{\hat{H}(t;\boldsymbol{\beta}_0, f_0)\} - \Lambda^*\{H_0(t)\}] d\left[\frac{\lambda\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\}}{\lambda^*\{H_0(t)\}}\right] \\ & + d\{o_p(\hat{H}(t;\boldsymbol{\beta}_0, f_0) - H_0(t))\} \end{aligned} \tag{A.7}$$

By simple algebra and the empirical process approximation theory, it is easy to show that

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} Y_i(t) d\left[\frac{\lambda\{H_0(t)+\boldsymbol{\beta}_0'\mathbf{Z}_i+f_0(X_i)\}}{\lambda^*\{H_0(t)\}}\right] \\ = & \frac{\lambda^*\{H_0(t)\}dB_1(t)-B_2(t)d\lambda^*\{H_0(t)\}}{[\lambda^*\{H_0(t)\}]^2}+o_p(n^{-1/2})=o_p(n^{-1/2}), \qquad (A.8) \end{aligned}$$

where the second last "=" is due to the fact $\lambda^*\{H_0(t)\}dB_1(t)-B_2(t)d\lambda^*\{H_0(t)\}=0$. Therefore, the equation (A.7) becomes

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} dM_i(t) &= \frac{B_2(t)}{\lambda^*\{H_0(t)\}} d\Big[\Lambda^*\{\hat{H}(t;\boldsymbol{\beta}_0,f_0)\}-\Lambda^*\{H_0(t)\}\Big] \\ &\quad +o_p(n^{-1/2})+d\{o_p(\hat{H}(t;\boldsymbol{\beta}_0,f_0)-H_0(t))\}, \end{aligned}$$

which leads to

$$\begin{aligned} & \Lambda^*\{\hat{H}(t;\boldsymbol{\beta}_0,f_0)\}-\Lambda^*\{H_0(t)\} \\ = & \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)} dM_i(s)+o_p(n^{-1/2})-\int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)} d\{o_p(\hat{H}(s;\boldsymbol{\beta}_0,f_0)-H_0(s))\} \\ = & \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)} dM_i(s)+o_p(n^{-1/2}). \qquad (A.9) \end{aligned}$$

The $o_p(n^{-1/2})$ term in the last equality can be obtained using the mean value theorem for integration based on the $\sqrt{n}$-consistency of the estimator $\hat{H}(\cdot;\beta_0,f_0)$, which can be established using the empirical process theory for Z-estimators (van der Vaart and Wellner, 1996).

Step 2. Note that

$$\sum_{i=1}^{n}\Big[dN_i(t)-Y_i(t)d\Lambda\{\hat{H}(t;\boldsymbol{\beta},f)+\boldsymbol{\beta}'Z_i+f(X_i)\}\Big]=0.$$

Take derivative with respect to $\boldsymbol{\beta}$, we have that

$$\sum_{i=1}^{n} Y_i(t) d\Big[\lambda\{\hat{H}(t;\boldsymbol{\beta},f)+\boldsymbol{\beta}'Z_i+f(X_i)\}\{Z_i+\frac{\partial}{\partial\boldsymbol{\beta}}\hat{H}(t,\boldsymbol{\beta},f)\}\Big]=0.$$

Following the similar technique used in *Step 1*, we can show that

$$\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\beta}} \hat{H}(t; \boldsymbol{\beta}, f)|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0, f=f_0} \\
= & -\int_0^t \frac{\lambda^*\{H_0(s)\} E[\mathbf{Z}\dot{\lambda}\{H_0(s) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(X)\}Y(s)]}{\lambda^*\{H_0(t)\}B_2(s)} dH_0(s) \\
& + o_p(1) \equiv -\boldsymbol{a}(t) + o_p(1). \qquad (A.10)
\end{aligned}$$

Step 3. Define

$$\boldsymbol{U}_1(\boldsymbol{\beta}, f) = \frac{1}{n}\sum_{i=1}^n \mathbf{Z}_i \Big[\delta_i - \Lambda\{\hat{H}(\tilde{T}_i; \boldsymbol{\beta}, f) + \boldsymbol{\beta}'\mathbf{Z}_i + f(X_i)\}\Big]. \qquad (A.11)$$

Then by the law of large numbers and (A.10), we have

$$\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{U}_1(\boldsymbol{\beta}, f)|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0, f=f_0} \\
= & -\int_0^\tau E\left[\{\mathbf{Z} - \boldsymbol{\mu}_{\mathbf{Z}}(t)\}\mathbf{Z}'\dot{\lambda}\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(X)\}Y(t)\right] dH_0(t) + o_p(1) \\
= & -\boldsymbol{A}_1 + o_p(1). \qquad (A.12)
\end{aligned}$$

Step 4. For $0 \le t \le \tau$, define $\hat{H}_0(t; \boldsymbol{\beta}) = \hat{H}(t; \boldsymbol{\beta}, \hat{\gamma}_0)$, where $(\hat{\gamma}_0, \hat{\gamma}_1)$ are the solutions of equations (6) and (7) at convergence. For any x, define

$$\begin{aligned}
& n\boldsymbol{U}_2(\gamma_0, \gamma_1, H, \boldsymbol{\beta})(x) \\
\equiv & \begin{pmatrix} \sum_{i=1}^n \int_0^\tau K_h(X_i - x)\Big[dN_i(t) - Y_i(t)d\Lambda\{H(t) + \boldsymbol{\beta}'\mathbf{Z}_i + \gamma_0(x) + \gamma_1(x)(X_i - x)\}\Big] \\ \sum_{i=1}^n \int_0^\tau \frac{X_i - x}{h} K_h(X_i - x)\Big[dN_i(t) - Y_i(t)d\Lambda\{H(t) + \boldsymbol{\beta}'\mathbf{Z}_i + \gamma_0(x) + \gamma_1(x)(X_i - x)\}\Big] \end{pmatrix}.
\end{aligned}$$

Then $\boldsymbol{U}_2\{\hat{\gamma}_0, \hat{\gamma}_1, \hat{H}_0(\cdot; \hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}\}(x) = 0$ for any x, where $\hat{\boldsymbol{\beta}}$ and $\hat{H}_0(\cdot; \hat{\boldsymbol{\beta}})$ are the solutions of equations (4) and (5) at convergence. By the Taylor expansion and the law of large numbers, we have

$$\begin{aligned}
& \boldsymbol{U}_2\{\hat{\gamma}_0, \hat{\gamma}_1, \hat{H}_0(\cdot; \hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}\}(x) \\
= & \boldsymbol{U}_2\{\hat{\gamma}_0, \hat{\gamma}_1, \hat{H}_0(\cdot; \boldsymbol{\beta}_0), \boldsymbol{\beta}_0\}(x) - \frac{1}{n}\sum_{i=1}^n K_h(X_i - x)\begin{pmatrix} 1 \\ \frac{X_i - x}{h} \end{pmatrix} \times \\
& \Big[\Lambda\{\hat{H}_0(\tilde{T}_i; \hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\} \\
& \quad -\Lambda\{\hat{H}_0(\tilde{T}_i; \boldsymbol{\beta}_0) + \beta_0'\mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\}\Big] \\
= & \boldsymbol{U}_2\{\hat{\gamma}_0, \hat{\gamma}_1, \hat{H}_0(\cdot; \boldsymbol{\beta}_0), \boldsymbol{\beta}_0\}(x) - \boldsymbol{E}_1(x) + o_p(n^{-1/2}),
\end{aligned}$$

where

$$\boldsymbol{E}_1(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x)\begin{pmatrix} 1 \\ \frac{X_i - x}{h}\end{pmatrix}\lambda\{\hat{H}_0(\tilde{T}_i;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0'\mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\}$$

$$\times\{\mathbf{Z}_i + \frac{\partial}{\partial\boldsymbol{\beta}}\hat{H}_0(\tilde{T}_i;\boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0). \quad \text{(A.13)}$$

Following the steps similar to the derivation of (A.2), we can show that

$$\frac{\partial}{\partial\boldsymbol{\beta}}\hat{H}_0(t;\boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = -\int_0^t \frac{\lambda^*\{H_0(s)\}E[\mathbf{Z}\dot{\lambda}\{H_0(s) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(X)\}Y(s)]}{\lambda^*\{H_0(t)\}B_2(s)}dH_0(s) + o_p(1).$$

In addition, we have

$$\begin{aligned}
&\boldsymbol{U}_2\{\hat{\gamma}_0,\hat{\gamma}_1,\hat{H}_0(\cdot;\boldsymbol{\beta}_0),\boldsymbol{\beta}_0\}(x)\\
=&\boldsymbol{U}_2(\hat{\gamma}_0,\hat{\gamma}_1,H_0,\boldsymbol{\beta}_0)(x) - \frac{1}{n}\sum_{i=1}^{n}K_h(X_i - x)\begin{pmatrix} 1 \\ \frac{X_i - x}{h}\end{pmatrix}\times\\
&\Big[\Lambda\{\hat{H}_0(\tilde{T}_i;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0'\mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\}\\
&\quad - \Lambda\{H_0(\tilde{T}_i) + \boldsymbol{\beta}_0'\mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\Big]\\
=&\boldsymbol{U}_2(\hat{\gamma}_0,\hat{\gamma}_1,H_0,\boldsymbol{\beta}_0)(x) - \boldsymbol{E}_2(x) + o_p(n^{-1/2}),
\end{aligned}$$

where

$$\boldsymbol{E}_2(x) = \frac{1}{n}\sum_{i=1}^{n}K_h(X_i - x)\begin{pmatrix} 1 \\ \frac{X_i - x}{h}\end{pmatrix}\frac{\lambda\{H_0(\tilde{T}_i) + \boldsymbol{\beta}_0'\mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\}}{\lambda^*\{H_0(\tilde{T}_i\}}$$

$$\times\left[\Lambda^*\{\hat{H}_0(\tilde{T}_i;\boldsymbol{\beta}_0)\} - \Lambda^*\{H_0(\tilde{T}_i)\}\right], \quad \text{(A.14)}$$

and

$$\boldsymbol{U}_2(\hat{\gamma}_0,\hat{\gamma}_1,H_0,\boldsymbol{\beta}_0)(x) = \boldsymbol{U}_2(f_0,\dot{f}_0,H_0,\boldsymbol{\beta}_0)(x) - \boldsymbol{E}_3(x)\begin{pmatrix}\hat{\gamma}_0(x) - f_0(x)\\ h\{\hat{\gamma}_1(x) - \dot{f}_0(x)\}\end{pmatrix} + o_p(n^{-1/2}),$$

where

$$\boldsymbol{E}_3(x) = \frac{1}{n}\sum_{i=1}^{n}K_h(X_i - x)\begin{pmatrix} 1 \\ \frac{X_i - x}{h}\end{pmatrix}\begin{pmatrix} 1 & \frac{X_i - x}{h}\end{pmatrix}\lambda_0\{H_0(\tilde{T}_i) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(x) + \dot{f}_0(x)(X_i - x)\}.$$

(A.15)

Combine all the equations above, we have

$$\boldsymbol{E}_3(x)\begin{pmatrix}\hat{\gamma}_0(x)-f_0(x)\\ h\{\hat{\gamma}_1(x)-\dot{f}_0(x)\}\end{pmatrix}=\boldsymbol{U}_2(f_0,\dot{f}_0,H_0,\boldsymbol{\beta}_0)(x)-\boldsymbol{E}_1(x)-\boldsymbol{E}_2(x)+o_p(n^{-1/2}).$$

By standard nonparametric techniques and the law of large numbers, we can show that $\boldsymbol{E}_3(x)$ converges to a deterministic function

$$\boldsymbol{e}_3(x)=g(x)E[\lambda\{H_0(\tilde{T})+\boldsymbol{\beta}_0'\mathbf{Z}+f_0(x)\}|X=x]\begin{pmatrix}1&0\\0&k_2\end{pmatrix},$$

Furthermore, we have

$$\begin{aligned}\boldsymbol{E}_1(x)&=g(x)\begin{pmatrix}E[\{\mathbf{Z}-\boldsymbol{a}(\tilde{T})\}'\lambda\{H_0(\tilde{T})+\boldsymbol{\beta}_0'\mathbf{Z}+f_0(x)\}|X=x]\\ 0'\end{pmatrix}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\&\quad+o_p(n^{-1/2})\equiv\begin{pmatrix}\boldsymbol{e}_1'(x)\\0'\end{pmatrix}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+o_p(n^{-1/2}),\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{E}_2(x)&=\int_0^\tau g(x)\begin{pmatrix}E[\frac{\lambda\{H_0(\tilde{T})+\boldsymbol{\beta}_0'\mathbf{Z}+f_0(x)\}}{\lambda^*\{H_0(\tilde{T})\}}Y(t)|X=x]\\0\end{pmatrix}d\left[\Lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\}-\Lambda^*\{H_0(t)\}\right]\\&\quad+o_p(n^{-1/2})\equiv\int_0^\tau\begin{pmatrix}e_2(x,t)\\0\end{pmatrix}d\left[\Lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\}-\Lambda^*\{H_0(t)\}\right]+o_p(n^{-1/2}).\end{aligned}$$

Therefore,

$$\begin{aligned}\begin{pmatrix}\hat{\gamma}_0(x)-f_0(x)\\ h\{\hat{\gamma}_1(x)-\dot{f}_0(x)\}\end{pmatrix}&=\boldsymbol{e}_3^{-1}(x)U_2(f_0,\dot{f}_0,H_0,\boldsymbol{\beta}_0)(x)-\boldsymbol{e}_3^{-1}(x)\begin{pmatrix}\boldsymbol{e}_1'(x)\\0'\end{pmatrix}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\&-\int_0^\tau\boldsymbol{e}_3^{-1}(x)\begin{pmatrix}e_2(x,t)\\0\end{pmatrix}d\left[\Lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\}-\Lambda^*\{H_0(t)\}\right]+o_p(n^{-1/2}),\qquad\text{(A.16)}\end{aligned}$$

where

$$\boldsymbol{e}_3^{-1}(x)=\frac{1}{g(x)E[\lambda\{H_0(\tilde{T})+\boldsymbol{\beta}_0'\mathbf{Z}+f_0(x)\}|X=x]}\begin{pmatrix}1&0\\0&\frac{1}{k_2}\end{pmatrix}\equiv\begin{pmatrix}\frac{1}{e_{31}(x)}&0\\0&\frac{1}{k_2e_{31}(x)}\end{pmatrix}.$$

In particular, for any given x, we have

$$\hat{\gamma}_0(x) - f_0(x) = \frac{1}{e_{31}(x)} U_{21}(f_0, \dot{f}_0, H_0, \boldsymbol{\beta}_0)(x) - \frac{\boldsymbol{e}_1'(x)}{e_{31}(x)}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$
$$- \int_0^\tau \frac{e_2(x,t)}{e_{31}(x)} d\left[\Lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\} - \Lambda^*\{H_0(t)\}\right] + o_p(n^{-1/2}), \qquad \text{(A.17)}$$

and

$$h\{\hat{\gamma}_1(x) - \dot{f}_0(x)\} = \frac{1}{k_2 e_{31}(x)} U_{22}(f_0, \dot{f}_0, H_0, \boldsymbol{\beta}_0)(x) + o_p(n^{-1/2}), \qquad \text{(A.18)}$$

where $U_2(f_0, \dot{f}_0, H_0, \boldsymbol{\beta}_0)(x) = \{U_{21}(f_0, \dot{f}_0, H_0, \boldsymbol{\beta}_0)(x), U_{22}(f_0, \dot{f}_0, H_0, \boldsymbol{\beta}_0)(x)\}'$.

Step 5. Note that $\hat{H}_0(\cdot;\boldsymbol{\beta})$ is the solution of the following equation for a fixed $\boldsymbol{\beta}$,

$$\sum_{i=1}^n \Big[dN_i(t) - Y_i(t) d\Lambda\{\hat{H}_0(t;\boldsymbol{\beta}) + \boldsymbol{\beta}'\mathbf{Z}_i + \hat{\gamma}_0(X_i)\}\Big] = 0, \qquad \text{(A.19)}$$

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

$$\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i \Big[dN_i(t) - Y_i(t) d\Lambda\{\hat{H}_0(t;\boldsymbol{\beta}) + \boldsymbol{\beta}' Z_i + \hat{\gamma}_0(X_i)\}\Big] = 0. \qquad \text{(A.20)}$$

Based on the equation (A.19) and following the derivation in *Step 1*, we can show that

$$\frac{1}{n}\sum_{i=1}^n dM_i(t) = \frac{B_2(t)}{\lambda^*\{H_0(t)\}} d\Big[\Lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\} - \Lambda^*\{H_0(t)\}\Big]$$
$$+ \quad \frac{1}{n}\sum_{i=1}^n Y_i(t)\{\hat{\gamma}_0(X_i) - f_0(X_i)\} d\lambda\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\}$$
$$+ \quad d\{o_p(\hat{H}_0(t;\boldsymbol{\beta}_0) - H_0(t))\}. \qquad \text{(A.21)}$$

Define

$$\boldsymbol{U}_1(\boldsymbol{\beta}, \hat{H}_0, \hat{\gamma}_0) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i \Big[dN_i(t) - Y_i(t) d\Lambda\{\hat{H}_0(t;\boldsymbol{\beta}) + \boldsymbol{\beta}'\mathbf{Z}_i + \hat{\gamma}_0(X_i)\}\Big].$$

Then by the Taylor expansion and some simple calculations, we have

$$0 = \boldsymbol{U}_1(\hat{\boldsymbol{\beta}}, \hat{H}_0, \hat{\gamma}_0) = \boldsymbol{U}_1(\boldsymbol{\beta}_0, \hat{H}_0, f_0) - \boldsymbol{S}_2 - \boldsymbol{S}_1 + o_p(n^{-1/2}),$$

where

$$\boldsymbol{S}_1 = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i Y_i(t)\{\hat{\gamma}_0(X_i) - f_0(X_i)\} d\lambda\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\},$$

$$\boldsymbol{S}_2 = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i Y_i(t) d\lambda\{\hat{H}_0(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\}\{\mathbf{Z}_i + \frac{\partial \hat{H}_0(t;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0).$$

In addition,

$$\begin{aligned} \boldsymbol{U}_1(\boldsymbol{\beta}_0, \hat{H}_0, f_0) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i\Big[dN_i(t) - Y_i(t)d\Lambda\{\hat{H}_0(t;\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\}\Big] \\ = & \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i dM_i(t) - \frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i\frac{\lambda\{H_0(\tilde{T}_i) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\}}{\lambda^*\{H_0(\tilde{T}_i)\}} \times \\ & \Big[\Lambda^*\{\hat{H}_0(\tilde{T}_i;\boldsymbol{\beta}_0) - \Lambda^*\{H_0(\tilde{T}_i)\}\Big] + o_p(n^{-1/2}). \end{aligned}$$

Following the technique used in *Step 3*, we have $\boldsymbol{S}_2 = \boldsymbol{A}_1(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(n^{-1/2})$. Thus,

$$\begin{aligned} \frac{1}{n}\int_0^\tau \mathbf{Z}_i dM_i(t) &= \boldsymbol{A}_1(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i Y_i(t)\{\hat{\gamma}_0(X_i) - f_0(X_i)\} d\lambda\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\} \\ + & \frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i\frac{\lambda\{H_0(\tilde{T}_i) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\}}{\lambda^*\{H_0(\tilde{T}_i)\}}\Big[\Lambda^*\{\hat{H}_0(\tilde{T}_i;\boldsymbol{\beta}_0) - \Lambda^*\{H_0(\tilde{T}_i)\}\Big] + o_p(n^{-1/2}). \end{aligned} \tag{A.22}$$

Step 6. We now plug the representation of $\hat{\gamma}_0(x) - f_0(x)$ given in (A.17) into (A.21) and (A.22) respectively. For (A.21), we have

$$\begin{aligned} & \frac{B_2(t)}{\lambda^*\{H_0(t)\}} d\Big[\Lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\} - \Lambda^*\{H_0(t)\}\Big] \\ & -d\{\boldsymbol{c}_1(t)\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - \int_0^\tau c_2(t,s) d\Big[\Lambda^*\{\hat{H}_0(s;\boldsymbol{\beta}_0)\} - \Lambda^*\{H_0(s)\}\Big] dH_0(t) \\ = & \frac{1}{n}\sum_{i=1}^{n} dM_i(t) - \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\frac{U_{21}(f_0, \dot{f}_0, H_0, \boldsymbol{\beta}_0)(X_i)}{e_{31}(X_i)} d\lambda\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\} \\ + & d\{o_p(\hat{H}_0(t;\boldsymbol{\beta}_0) - H_0(t))\}, \end{aligned} \tag{A.23}$$

where

$$d\{\boldsymbol{c}_1(t)\} = E\left[\frac{\boldsymbol{e}_1'(X)}{e_{31}(X)}\dot{\lambda}\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(X)\}Y(t)\right] dH_0(t),$$

$$c_2(t,s) = E\left[\frac{e_2(X,s)}{e_{31}(X)}\dot{\lambda}\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(X)\}Y(t)\right].$$

Multiplying both sides of the equation (A.23) by $\boldsymbol{\alpha}(t)\frac{\lambda^*\{H_0(t)\}}{B_2(t)}$ and integrating t from 0 to τ, we have

$$\begin{aligned}
&\int_0^\tau \left[\boldsymbol{\alpha}(t) - \int_0^\tau \boldsymbol{\alpha}(s)\frac{\lambda^*\{H_0(s)\}}{B_2(s)}c_2(s,t)dH_0(s)\right]d\left[\Lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\} - \Lambda^*\{H_0(t)\}\right] \\
&\quad -\boldsymbol{A}_{22}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
= &\ \frac{1}{n}\sum_{i=1}^n \int_0^\tau \boldsymbol{\alpha}(t)\frac{\lambda^*\{H_0(t)\}}{B_2(t)}dM_i(t) - \\
&\ \frac{1}{n}\sum_{i=1}^n \int_0^\tau \boldsymbol{\alpha}(t)\frac{\lambda^*\{H_0(t)\}}{B_2(t)}Y_i(t)\frac{U_{21}(f_0,\dot{f}_0,H_0,\boldsymbol{\beta}_0)(X_i)}{e_{31}(X_i)}d\lambda\{H_0(t)+\boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\} \\
&\ +o_p(n^{-1/2}), \qquad \text{(A.24)}
\end{aligned}$$

where $\boldsymbol{A}_{22} = \int_0^\tau \boldsymbol{\alpha}(t)\frac{\lambda^*\{H_0(t)\}}{B_2(t)}d\{\boldsymbol{c}_1(t)\}$. For (A.22), we have

$$\begin{aligned}
&(\boldsymbol{A}_1 - \boldsymbol{A}_{21})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \int_0^\tau \{\boldsymbol{c}_3(t) - \boldsymbol{c}_4(t)\}d\left[\Lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\} - \Lambda^*\{H_0(t)\}\right] \\
&\quad -\frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i dM_i(t) \\
= &\ -\frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i Y_i(t)\frac{U_{21}(f_0,\dot{f}_0,H_0,\boldsymbol{\beta}_0)(X_i)}{e_{31}(X_i)}d\lambda\{H_0(t)+\boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\} + o_p(n^{-1/2}),
\end{aligned}$$

(A.25)

where

$$\begin{aligned}
\boldsymbol{A}_{21} &= \int_0^\tau E\left[\frac{\mathbf{Z}\boldsymbol{e}_1'(X)}{e_{31}(X)}\dot{\lambda}\{H_0(t)+\boldsymbol{\beta}_0'\mathbf{Z}+f_0(X)\}Y(t)\right]dH_0(t), \\
\boldsymbol{c}_3(t) &= E\left[\frac{\mathbf{Z}\lambda\{H_0(\tilde{T})+\boldsymbol{\beta}_0'\mathbf{Z}+f_0(X)\}Y(t)}{\lambda^*(\tilde{T})}\right], \\
\boldsymbol{c}_4(t) &= \int_0^\tau E\left[\frac{e_2(X,t)}{e_{31}(X)}\mathbf{Z}\dot{\lambda}\{H_0(s)+\boldsymbol{\beta}_0'\mathbf{Z}+f_0(X)\}Y(s)\right]dH_0(s).
\end{aligned}$$

Recall that, based on (8), $\boldsymbol{\alpha}(t) - \int_0^\tau D_1(s,t)\boldsymbol{\alpha}(s)dH_0(s) = \boldsymbol{D}_2(t)$, where

$$D_1(s,t) \equiv \frac{\lambda^*\{H_0(s)\}}{B_2(s)}c_2(s,t) \text{ and } \boldsymbol{D}_2(t) \equiv \boldsymbol{c}_3(t) - \boldsymbol{c}_4(t). \qquad \text{(A.26)}$$

Thus, (A.24) - (A.25) gives

$$(\boldsymbol{A}_1 - \boldsymbol{A}_2)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i - \boldsymbol{m}_{\mathbf{Z}}(t)\}dM_i(t) - (\boldsymbol{G}_1 - \boldsymbol{G}_2) + o_p(n^{-1/2}),$$

(A.27)

where $\boldsymbol{A}_2 = \boldsymbol{A}_{21} - \boldsymbol{A}_{22}$ and

$$\boldsymbol{\rho}(x) = \boldsymbol{e}_1(x)/e_{31}(x), \tag{A.28}$$

$$\boldsymbol{G}_1 = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i Y_i(t)\frac{U_{21}(f_0,\dot{f}_0,H_0,\boldsymbol{\beta}_0)(X_i)}{e_{31}(X_i)}d\lambda\{H_0(t)+\boldsymbol{\beta}_0'\mathbf{Z}_i+f_0(X_i)\},$$

$$\boldsymbol{G}_2 = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \boldsymbol{\alpha}(t)\frac{\lambda^*\{H_0(t)\}}{B_2(t)}Y_i(t)\frac{U_{21}(f_0,\dot{f}_0,H_0,\boldsymbol{\beta}_0)(X_i)}{e_{31}(X_i)}d\lambda\{H_0(t)+\boldsymbol{\beta}_0'\mathbf{Z}_i+f_0(X_i)\}.$$

Step 7. By the Taylor expansion and standard nonparametric techniques, we have

$$\begin{aligned}\boldsymbol{G}_1 &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \frac{\mathbf{Z}_i Y_i(t)}{e_{31}(X_i)}\frac{1}{n}\sum_{j=1}^{n}\int_0^\tau K_h(X_j-X_i)\Big[dN_j(t) - \\ &\quad Y_j(t)d\Lambda\{H_0(t)+\boldsymbol{\beta}_0'\mathbf{Z}_j+f_0(X_i)+\dot{f}_0(X_i)(X_j-X_i)\}\Big]d\lambda\{H_0(t)+\boldsymbol{\beta}_0'\mathbf{Z}_i+f_0(X_i)\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i^* dM_i(t) + o_p(n^{-1/2}), \end{aligned}\tag{A.29}$$

and similarly

$$\boldsymbol{G}_2 = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \boldsymbol{m}_{\mathbf{Z}^*,i}dM_i(t) + o_p(n^{-1/2}). \tag{A.30}$$

Therefore, from (A.27), (A.29) and (A.30) we have (A.6). □

Proof of Theorem 2. We now provide a constructive argument for establishing the asymptotic representation of $\sqrt{n}\{\hat{H}(t)-H_0(t)\}$. Recall that $\hat{H}_0(t;\boldsymbol{\beta}) \equiv \hat{H}(t;\boldsymbol{\beta},\hat{\gamma}_0)$, where $(\hat{\gamma}_0,\hat{\gamma}_1)$ are the solutions of equations (4) and (5) at convergence. Our estimator of H is $\hat{H}(t) = \hat{H}_0(t;\hat{\boldsymbol{\beta}})$. As in Chen et al. (2002) for the linear transformation model, we will first derive the asymptotic representation of $\sqrt{n}[\Lambda^*\{\hat{H}_0(t;\hat{\boldsymbol{\beta}})\} - \Lambda^*\{H_0(t)\}]$ for any $t \in (0,\tau]$. Note that

$$\begin{aligned} &\Lambda^*\{\hat{H}_0(t;\hat{\boldsymbol{\beta}})\} \\ =\;& \Lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\} + \lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\}\left(\frac{\partial \hat{H}_0(t;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + o_p(|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0|) \\ =\;& \Lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\} - \int_0^t \frac{E[\mathbf{Z}'\dot{\lambda}\{H_0(s)+\boldsymbol{\beta}_0'\mathbf{Z}+f_0(X)\}Y(s)]}{B_2(s)}d\Lambda^*\{H_0(s)\}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \\ &+ o_p(|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0|), \end{aligned}$$

where the last equality above is established in the proof of Theorem 1 (Step 4). In the proof of Theorem 1, we already established the asymptotic representation of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, that is,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \boldsymbol{A}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} \{\mathbf{Z}_i - \boldsymbol{m}_{\mathbf{Z}}(t) - (\mathbf{Z}_i^* - \boldsymbol{m}_{\mathbf{Z}^*,i})\} dM_i(t) + o_p(1).$$

Thus, we only need to establish the asymptotic representation of $\sqrt{n}[\Lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\} - \Lambda^*\{H_0(t)\}]$ for any $t \in (0,\tau]$. Define $\hat{\mu}_n(t) = \sqrt{n}[\Lambda^*\{\hat{H}_0(t;\boldsymbol{\beta}_0)\} - \Lambda^*\{H_0(t)\}]$. We have the following lemma.

Lemma 2. *Under the regularity conditions (C1)-(C7), if $nh^2/\{\log(1/h)\} \to \infty$ and $nh^4 \to 0$, we have that $\hat{\mu}_n(t)$ asymptotically satisfies the following integral equation:*

$$\hat{\mu}_n(t) - \int_0^{\tau} a(t,s) d\hat{\mu}_n(s) = W_n(t), \quad t \in (0,\tau], \tag{A.31}$$

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

For now, we assume Lemma 2 holds and its proof will be given later. Applying integration by part, we can rewrite equation (A.31) as a Fredholm integral equation of the second kind with the kernel $\frac{\partial a(t,s)}{\partial s}$. Here, we assume that equation (A.31) has a unique solution, which can be assured if

$$\sup_{t\in[0,\tau]} \int_0^{\tau} |\frac{\partial a(t,s)}{\partial s}| ds < \infty. \tag{A.32}$$

Next, we construct a solution to equation (A.31) as

$$\hat{\mu}_n(t) = W_n(t) + \int_0^{\tau} b(t,s) dW_n(s), \tag{A.33}$$

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

$$b(t,s) - \int_0^{\tau} a(t,u) \frac{\partial b(u,s)}{\partial u} du = a(t,s), \quad t,s \in [0,\tau]. \tag{A.34}$$

We note that equation (A.34) can be rewritten as a Fredholm integral equation of the second kind with the same kernel $\frac{\partial a(t,s)}{\partial s}$. Therefore, under condition (A.32), the above equation has a unique solution. Then it is easy to show that $\hat{\mu}_n(t)$ defined in (A.33) is a solution to the integral equation (A.31). Based on the asymptotic representation (A.33) of $\hat{\mu}_n(t)$, we have $\sqrt{n}[\Lambda^*\{\hat{H}_0(t;\hat{\boldsymbol{\beta}})\}-\Lambda^*\{H_0(t)\}] = \frac{1}{\sqrt{n}}\sum_{i=1}^n \kappa_i(t)+o_p(1)$, where

$$\begin{aligned}\kappa_i(t) &= w_i(t) + \int_0^\tau b(t,s)dw_i(s) - \int_0^t \frac{E[\mathbf{Z}'\dot{\lambda}\{H_0(s)+\boldsymbol{\beta}_0'\mathbf{Z}+f_0(X)\}Y(s)]}{B_2(s)}d\Lambda^*\{H_0(s)\} \\ &\quad \boldsymbol{A}^{-1}\int_0^\tau \{\mathbf{Z}_i - \boldsymbol{m}_{\mathbf{Z}}(t) - (\mathbf{Z}_i^* - \boldsymbol{m}_{\mathbf{Z}^*,i})\}dM_i(t),\end{aligned}$$

are independent mean zero functions. Then based on the functional delta method, we have $\sqrt{n}\{\hat{H}(t) - H_0(t)\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\kappa_i(t)}{\lambda^*\{H_0(t)\}} + o_p(1)$, which can then be shown to converge weakly to a mean zero Gaussian process using the functional central limit theorem (Pollard 1990, Theorem 10.6). The root-n consistency of $\hat{H}(t)$ is then established. □

Proof of Lemma 2. Based on equation (A.23), we have

$$\begin{aligned}&\frac{B_2(t)}{\lambda^*\{H_0(t)\}}d\{\hat{\mu}_n(t)\} - d\{\boldsymbol{c}_1(t)\}\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) - \int_0^\tau c_2(t,s)d\{\hat{\mu}_n(s)\}dH_0(t) \\ =\ & \frac{1}{\sqrt{n}}\sum_{i=1}^n dM_i(t) - \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i(t)\frac{U_{21}(f_0,\dot{f}_0,H_0,\boldsymbol{\beta}_0)(X_i)}{e_{31}(X_i)}d\lambda\{H_0(t)+\boldsymbol{\beta}_0'\mathbf{Z}_i+f_0(X_i)\} \\ &+ d\{o_p(\hat{H}_0(t;\boldsymbol{\beta}_0) - H_0(t))\}.\end{aligned}$$

According to the proof of Theorem 1, $\int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)}d\{\boldsymbol{c}_1(s)\}\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ can be represented asymptotically as the summation of independent mean zero functions multiplying by $n^{-1/2}$. Moreover, following Step 7 of the proof of Theorem 1, we have

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)}Y_i(s)\frac{U_{21}(f_0,\dot{f}_0,H_0,\boldsymbol{\beta}_0)(X_i)}{e_{31}(X_i)}d\lambda\{H_0(s)+\boldsymbol{\beta}_0'\mathbf{Z}_i+f_0(X_i)\} \\ =\ & \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^t \tilde{m}_{\mathbf{Z}^*,i}(s)dM_i(s) + o_p(1),\end{aligned}$$

where

$$\tilde{m}_{\mathbf{Z}^*,i}(t) = \int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)}\frac{E[\dot{\lambda}\{H_0(s)+\boldsymbol{\beta}_0'\mathbf{Z}+f_0(X)\}Y(s)|X=X_i]}{E[\lambda\{H_0(s)+\boldsymbol{\beta}_0'\mathbf{Z}+f_0(X)\}|X=X_i]}dH_0(s).$$

Therefore, we have asymptotically,

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

where $a(t,s) = \int_0^t \frac{\lambda^*\{H_0(u)\}}{B_2(u)} c_2(u,s)dH_0(u)$ and

$$\begin{aligned} w_i(t) &= \int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)} dM_i(s) - \int_0^t \tilde{m}_{\mathbf{Z}^*,i}(s)dM_i(s) \\ &\quad + \int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)} d\{\boldsymbol{c}_1(s)\} \int_0^\tau \{\mathbf{Z}_i - \boldsymbol{m}_{\mathbf{Z}}(t) - (\mathbf{Z}_i^* - \boldsymbol{m}_{\mathbf{Z}^*,i})\}dM_i(t), \end{aligned}$$

which are independent mean zero functions. Thus, $W_n(t)$ converges weakly to mean zero Gaussian process as $n \to \infty$, which completes the proof. □

PROOF OF THEOREM 3. Note that $\boldsymbol{U}_2(\hat{\gamma}_0, \hat{\gamma}_1, \hat{H}, \hat{\boldsymbol{\beta}})(x) = 0$. Based on the assumptions $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = O_p(1)$, $\sqrt{n}|\hat{H}(t) - H_0(t)| = O_p(1)$ for $t \in (0,\tau]$, and the conditions of Theorem 3, it is easy to show

$$\begin{aligned} \sup_{t\in(0,\tau]} ||\frac{1}{n}\sum_{i=1}^n K_h(X_i - x) \begin{pmatrix} 1 \\ \frac{X_i - x}{h} \end{pmatrix} \Big[\lambda\{\hat{H}(t) + \hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\} \\ -\lambda\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z}_i + \hat{\gamma}_0(x) + \hat{\gamma}_1(x)(X_i - x)\}\Big]|| = O_p(n^{-1/2}). \quad \text{(A.35)} \end{aligned}$$

Thus, $\boldsymbol{U}_2(\hat{\gamma}_0, \hat{\gamma}_1, H_0, \boldsymbol{\beta}_0)(x) = O_p(n^{-1/2}) = o_p(1/\sqrt{nh})$. Let $\tilde{\boldsymbol{\gamma}}(x) = (\gamma_0(x), h\gamma_1(x))'$, $\hat{\tilde{\boldsymbol{\gamma}}}(x) = (\hat{\gamma}_0(x), h\hat{\gamma}_1(x))'$ and $\tilde{\boldsymbol{f}}(x) = (f_0(x), h\dot{f}_0(x))'$.

We first show that $\hat{\tilde{\boldsymbol{\gamma}}}(x) \to \tilde{\boldsymbol{f}}(x)$ in probability as $n \to \infty$. Note that

$$\begin{aligned} \frac{\partial}{\partial\tilde{\boldsymbol{\gamma}}}\boldsymbol{U}_2(\gamma_0, \gamma_1, \hat{H}, \hat{\boldsymbol{\beta}}) &= -\frac{1}{n}\sum_{i=1}^n K_h(X_i - x)\begin{pmatrix} 1 \\ \frac{X_i - x}{h} \end{pmatrix} \times \\ &\quad \Big(1, \frac{X_i - x}{h}\Big)\lambda\{\hat{H}(\tilde{T}_i) + \hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \gamma_0(x) + \gamma_1(x)(X_i - x)\}, \end{aligned}$$

which is negative definite. In addition, by the strong law of large numbers and standard nonparametric techniques, it can be shown that the above derivative also converges to a deterministic negative definite matrix, denoted by $-\dot{\boldsymbol{u}}_{\tilde{\boldsymbol{\gamma}}}(\gamma_0, H_0, \boldsymbol{\beta}_0)$,

where

$$\dot{\boldsymbol{u}}_{\tilde{\boldsymbol{\gamma}}}(\gamma_0, H_0, \boldsymbol{\beta}_0) = g(x)E[\lambda\{H_0(\tilde{T}) + \boldsymbol{\beta}_0'\mathbf{Z} + \gamma_0(x)\}|X = x] \begin{pmatrix} 1 & 0 \\ 0 & k_2 \end{pmatrix}.$$

Furthermore, by the strong law of large number and the consistency of $\hat{H}$ and $\hat{\boldsymbol{\beta}}$, we have $\boldsymbol{U}_2(\gamma_0, \gamma_1, \hat{H}, \hat{\boldsymbol{\beta}}) \to \boldsymbol{u}(\gamma_0, H_0, \boldsymbol{\beta}_0)$ in probability, where

$$\boldsymbol{u}(\gamma_0, H_0, \boldsymbol{\beta}_0)(x) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} g(x)E[\delta - \Lambda\{H_0(\tilde{T}) + \boldsymbol{\beta}_0'\mathbf{Z} + \gamma_0(x)\}|X = x].$$

Note that $\boldsymbol{u}(f_0, H_0, \boldsymbol{\beta}_0)(x) = 0$. Following the techniques of Fan, Gijbels and King (1997), the consistency of $\hat{\tilde{\boldsymbol{\gamma}}}(x)$ holds. Using the Taylor expansion, we have

$$\begin{aligned} & \boldsymbol{U}_2(\hat{\tilde{\boldsymbol{\gamma}}}, H_0, \boldsymbol{\beta}_0)(x) \\ = \; & \boldsymbol{U}_2(\tilde{\boldsymbol{f}}, H_0, \boldsymbol{\beta}_0)(x) + \frac{\partial}{\partial \tilde{\boldsymbol{\gamma}}} \boldsymbol{U}_2(\tilde{\boldsymbol{\gamma}}^*, H_0, \boldsymbol{\beta}_0)(x)\{\hat{\tilde{\boldsymbol{\gamma}}}(x) - \tilde{\boldsymbol{f}}(x)\} = o_p(1/\sqrt{nh}), \end{aligned} \tag{A.36}$$

where $\tilde{\boldsymbol{\gamma}}^*(x)$ lies between $\hat{\tilde{\boldsymbol{\gamma}}}(x)$ and $\tilde{\boldsymbol{f}}(x)$, and thus $\tilde{\boldsymbol{\gamma}}^*(x) \to \tilde{\boldsymbol{f}}(x)$ in probability.

By the strong law of large numbers, we obtain

$$\boldsymbol{V}_1(x) \equiv - \lim_{n\to\infty} \frac{\partial}{\partial \tilde{\boldsymbol{\gamma}}} \boldsymbol{U}_2(\tilde{\boldsymbol{f}}, H_0, \boldsymbol{\beta}_0)(x) = \dot{\boldsymbol{u}}_{\tilde{\boldsymbol{\gamma}}}(f_0, H_0, \boldsymbol{\beta}_0). \tag{A.37}$$

Also, we have

$$\begin{aligned} & \boldsymbol{U}_2(\tilde{f}, H_0, \boldsymbol{\beta}_0)(x) \equiv \boldsymbol{U}_{2,1}(\tilde{f}, H_0, \boldsymbol{\beta}_0)(x) + \boldsymbol{U}_{2,2}(\tilde{f}, H_0, \boldsymbol{\beta}_0)(x) \\ = \; & \frac{1}{n}\sum_{i=1}^{n} \int_0^{\tau} K_h(X_i - x) \begin{pmatrix} 1 \\ \frac{X_i - x}{h} \end{pmatrix} dM_i(t) + \frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x) \begin{pmatrix} 1 \\ \frac{X_i - x}{h} \end{pmatrix} \times \\ & \Big[\Lambda\{H_0(\tilde{T}_i) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(X_i)\} - \Lambda\{H_0(\tilde{T}_i) + \boldsymbol{\beta}_0'\mathbf{Z}_i + f_0(x) + \dot{f}_0(x)(X_i - x)\}\Big]. \end{aligned}$$

Then following the proof of Theorem 4 given in the Appendix B of Cai et al. (2007) and the martingale central limit theorem, we have

$$(nh)^{1/2}\boldsymbol{U}_{2,1}(\tilde{f}, H_0, \boldsymbol{\beta}_0)(x) \to N\{0, \boldsymbol{V}_2(x)\}$$

as n goes to infinity, where

$$\boldsymbol{V}_2(x) = \begin{pmatrix} 1 & 0 \\ 0 & k_2 \end{pmatrix} g(x) \int_0^\tau E[Y(t)\lambda\{H_0(t) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(x)\}|X = x]dH_0(t). \quad \text{(A.38)}$$

Define $\tilde{\Lambda}(X_i) = \Lambda\{H_0(\tilde{T}_i)+\boldsymbol{\beta}_0'\mathbf{Z}_i+f_0(X_i)\}-\Lambda\{H_0(\tilde{T}_i)+\boldsymbol{\beta}_0'\mathbf{Z}_i+f_0(x)+\dot{f}_0(x)(X_i-x)\}$. Using the Taylor expansion of $\tilde{\Lambda}$ around x, we have $\boldsymbol{U}_{2,2}(x) = \boldsymbol{V}_1(x)\boldsymbol{b}_n(x) + o_p(h^2)$, where

$$\boldsymbol{b}_n(x) = \frac{h^2}{2}\ddot{f}_0(x)\boldsymbol{V}_1^{-1}(x)E[\lambda\{H_0(\tilde{T}) + \boldsymbol{\beta}_0'\mathbf{Z} + f_0(x)\}|X = x]\begin{pmatrix} k_2 \\ 0 \end{pmatrix}. \quad \text{(A.39)}$$

Combining with (A.35), we have

$$\boldsymbol{V}_1(x)(nh)^{1/2}[\{\hat{\tilde{\gamma}}(x) - \tilde{\boldsymbol{f}}(x)\} - \boldsymbol{b}_n(x) + o_p(h^2)] = (nh)^{1/2}\boldsymbol{U}_{2,1}(\tilde{f}, H_0, \boldsymbol{\beta}_0)(x) + o_p(1).$$

Since nh^5 is bounded, the results of Theorem 3 hold. □

Proof of Theorem 4. The proof of Theorem 4 can be similarly derived as that for Theorem 1 and hence is omitted here. □

Appendix C: Simulation studies on the stability and efficiency of the proposed algorithm

C.1. Numerical results for using different initial values

To solve equations (4) and (5), as given in Step 0 of our computational algorithm, we need to first choose an initial value for the nonlinear covariate effect, i.e $\hat{f}^{(0)}(\cdot)$. After $\hat{f}^{(0)}(\cdot)$ is fixed, (4) and (5) become the equations of Chen et al. (2002) for the linear transformation model. In our simulations and the real data application, we chose $\hat{f}^{(0)}(\cdot) \equiv 0$, and then solved (4) and (5) using the method of Chen et al. (2002) iteratively until some stopping rule was met. The algorithm worked well in all the examples and it converged within 30 iterations in almost all simulations.

Here, we did additional simulation studies to examine the effects of using different initial values for the nonlinear covariate on the solution. We considered design I

($f(x) = 8(x - x^3)$) subject to a covariate-independent censoring with 20% censoring rate. Besides the constant initial value (i.e. $\hat{f}^{(0)}(\cdot) \equiv 0$), we also tried the linear form ($f(x) = \theta_{11}x$) and the quadratic form ($f(x) = \theta_{21}x + \theta_{22}x^2$) as the initial value of $f(\cdot)$. Then we solved the corresponding linear transformation model using Chen et al. (2002) to obtain the estimates $\hat{\theta}_{11}$, $\hat{\theta}_{21}$ and $\hat{\theta}_{22}$ to obtain the initial value $\hat{f}^{(0)}(\cdot)$. We ran 500 simulations and the results were summarized in Table 1. Based on these results, we observed that different initial values of the nonlinear function produced almost the same estimates, with little differences in biases and almost identical sample standard deviations. From these limited simulation studies, it seemed that the algorithm worked quite stably with respect to different choices of initial values.

C.2. Numerical results for efficiency comparison

Similar to many other estimating equation methods, the proposed martingale-based estimating equations are not guaranteed to be semiparametric efficient. Following the suggestion of one referee, we conducted additional simulations to study the efficiency loss of our estimator if the true covariate effect is linear while the proposed method is used. To be specific, we considered a linear transformation model with three covariates: Z_1, Z_2 and X as in our old simulations. However, the true covariate effect of X is linear with the coefficient of 1.0 instead of nonlinear. All other settings are the same as before. We examined the scenario with 20% covariate-independent censoring and compared our method with Chen, Jin and Ying (2002)'s method (denoted as CJY) for the linear transformation model. We did 500 runs and summarized the estimated coefficients for Z_1 and Z_2 in Table 2. Based on these results, we observed that if all the covariate effects are in fact linear, CJY's method showed less biases and smaller variances than our method, but the differences are very small. This indicated that our method may loose some efficiency due to the kernel estimation of the nonlinear covariate effect under the linear transformation model, but the efficiency lost is not much.

Additional references:

Pollard, D. (1990). *Empirical Processes: Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics Volume **2**, Hayward.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics,* New York: John Wiley & Sons.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics,* New York: Springer.

Table 1: Simulation results for different initial estimates of nonlinear covariates.

	$\beta_{01} = -1.0$		$\beta_{02} = 1.0$	
Initial Est.	Bias	SD	Bias	SD
	$\zeta = 0$			
Constant	0.004	0.246	0.036	0.373
Linear	0.004	0.246	0.036	0.373
Quadratic	0.005	0.246	0.026	0.372
	$\zeta = 1$			
Constant	-0.033	0.420	0.004	0.679
Linear	-0.038	0.421	0.004	0.678
Quadratic	-0.074	0.422	0.008	0.678
	$\zeta = 0.5$			
Constant	-0.021	0.331	0.004	0.537
Linear	-0.021	0.331	0.004	0.537
Quadratic	-0.013	0.331	0.006	0.535

Constant, the initial estimate was chosen as $\hat{f}^{(0)}(x) \equiv 0$; Linear, $\hat{f}^{(0)}(x) = \hat{\theta}_{11}x$; Quadratic, $\hat{f}^{(0)}(x) = \hat{\theta}_{21}x + \hat{\theta}_{22}x^2$; Here in the second and third methods, the estimates $\hat{\theta}_{11}$, $\hat{\theta}_{21}$ and $\hat{\theta}_{22}$ were obtained by solving the corresponding linear transformation model using the method of Chen et al. (2002).

Table 2: Efficiency comparisons of our method and CYJ's method under linear transformation model.

method	$\beta_{01} = -1.0$			$\beta_{02} = 1.0$		
	Bias	SD	MSE	Bias	SD	MSE
			$\zeta = 0$			
Our	-0.068	0.254	0.069	0.089	0.408	0.174
CJY	-0.055	0.245	0.063	-0.008	0.401	0.161
			$\zeta = 1$			
Our	-0.038	0.405	0.165	0.031	0.704	0.497
CJY	-0.030	0.397	0.158	-0.029	0.686	0.471
			$\zeta = 0.5$			
Our	-0.039	0.328	0.109	0.035	0.563	0.318
CYZ	-0.029	0.320	0.103	-0.036	0.545	0.298

MSE, mean squared error; Our, our proposed method for the partial linear transformation model, where the third covariate was assumed nonlinear; CYZ, Chen, Jin and Ying (2002)'s method for the linear transformation model.