A Second Order Semiparametric Method for Survival Analysis, with Application to an AIDS Clinical Trial Study

Fei Jiang
Department of Statistics, University of Hongkong
Yanyuan Ma
Department of Statistics, Penn State University
J. Jack Lee
Department of Biostatistics, University of Texas MD Anderson Cancer Center

July 17, 2016

Summary

Motivated from a recent AIDS clinical trial study A5175, we propose a semiparametric framework to describe time to event data, where only the dependence of the mean and variance of the time on the covariates are specified through a restricted moment model. We use a second-order semiparametric efficient score combined with a nonparametric imputation device for estimation. Compared with an imputed weighted least square method, the proposed approach improves the efficiency of the parameter estimation whenever the third moment of the error distribution is nonzero. We compare the method with a parametric survival regression method in the A5175 study data analysis. In the data analysis, the proposed method shows better fit to the data with smaller mean squared residuals. In summary, this work provides a semiparametric framework in modeling and estimation of the survival data. The framework has wide applications in data analysis.

Keywords: Censoring, Efficiency, Imputation, Kernel, Nonparametric, Restricted moments, Semiparametrics, Two stage.

1 Introduction

A new AIDS Clinical Trials Group study, A5175, was recently conducted to evaluate several antiretroviral regimens in diverse populations. One primary goal of the study is to investigate the safety of these regimens so as to maximize the efficiency of the antiretroviral delivery in various areas (Campbell et al. 2012). The primary safety endpoint of the study is a patient's time to one of the following three early adverse reactions: onset of a grade ≥ 3 severity sign, a grade ≥ 3 laboratory abnormality and a change of the initial treatment due to toxicity of the treatment. A patient's event was considered to be censored if he/she did not meet the primary endpoint criteria at the end of the study or at the final medication dose. In addition, the study also collected patients' CD4 counts at the baseline and then at the weeks 8, 24, 72 and 96. Compared with the primary safety endpoint, the CD4 counts information was obtained relatively easily in a shorter period of time. Although the CD4 counts information is primarily used in inferring the treatment efficacy (Campbell et al. 2012), it is also related to the safety of the antiretroviral regimens. For example, Hirsch (2008) showed that using the same antiretroviral regimen at a higher CD4 counts level would lower the risk of toxicities. Thus, it is natural to expect that an analysis on the primary safety endpoint would be more efficient if the short term information on CD4 counts can be included. This motivates us to develop methods to analyze the relation between CD4 counts and the primary safety endpoint, with the goal of ameliorating the existing post-trial data analysis procedures. In addition, we also explore the usage of the proposed methods in the clinical trial design stages so as to improve trial efficiency.

In the A5175 study, safety of a treatment is described by time to adverse events, and all the subsequent decisions are made based on the inference on the event time. This motivates us to model the time to the primary safety endpoint directly as a function of the covariates. In contrast, traditional time to event models such as Cox proportional hazard model focus on evaluating the covariate effect on the disease risk and do not provide direct inference on the event time. Our preliminary analysis (Section 3.3) on the A5175 study data shows that both the mean and variance of the primary safety endpoint depend on the short term CD4 counts. To capture this relation while remain flexible, we use a semiparametric second order restricted moment (RMM2) model to specify the mean and variance structures of the primary safety endpoint while leaving all other aspects of the model unspecified. The model has the characteristics of capturing the central structure while remaining flexible in non-crucial parts of the model. By modeling the variance in addition to the mean, the RMM2 model enriches the structure of the classical restricted moment model.

To obtain accurate parameter estimation and to perform proper inference on time to the primary safety endpoint, we devise a semiparametric estimation procedure for the RMM2 model used in fitting the A5175 data. To our best knowledge, such modeling and estimation approaches have not been considered in survival models. In classical regression models, parameter estimation is often performed using the ordinary least square (OLS) method, which is efficient when the errors are normally distributed (Gallant 2009). However, the additional variance structures in the A5175 study data implies that the OLS estimators may not be optimal. Under the complete data settings, Wang & Leblanc (2008) proposed a second order least square method when the error variances are constant. The method was later generalized to covariate dependent error variances and shown to minimise the variances of the estimators (Kim & Ma 2012).

The A5175 study data is further subject to censoring. This prevents the direct application of the methods described above because without fully specifying the event time distribution, the score functions of the censored subjects are difficult to obtain. In a completely different context, Wang et al. (2012) proposed a nonparametric score imputation method to cope with censoring when covariates are discrete. The nonparametric score imputation method often performs competitively compared to the optimal augmented inverse probability weighting method in terms of estimation variability in finite samples (Wang et al. 2012), while the former has more intuitive form and is more interpretable. This inspires us to examine the nonparametric imputation strategy and extend the method to incorporate continuous covariates (CD4 counts in the A5175 study data). We then generalize the semiparametric estimation method of Kim & Ma (2012) to handle survival data. We develop an imputation based semiparametric efficient estimator for the RMM2 model (RMM2-ISE), which combines the nonparametric score imputation with the second order least square score function introduced in Kim & Ma (2012). We derive its asymptotic estimation variance and establish its root-n consistency and asymptotic normality. We evaluate the finite sample properties of the RMM2-ISE estimator. We further compare the RMM2-ISE estimation procedure with a simpler method, which we name the imputed weighted least square (IWLS) method through simulation studies. We developed IWLS here to combine nonparametric score imputation and weighted least square score functions. Similar idea was used in Lipsitz et al. (1999) to handle missing covariates. Moreover, we apply the RMM2-ISE method for analyzing the A5175 study data. The RMM2-ISE method also shows better data fitting compared with the method combining the accelerated failure time Weibull model and maximized likelihood estimation (AFT-Weibull-ML). Throughout the paper, we choose the Weibull survival time model to fit the data for comparisons because it is sufficiently flexible to accommodate the increasing, decreasing and constant hazard rates (Klein & Moeschberger 2010).

The rest of the paper is structured as follows. In Section 2, we describe the RMM2 model and introduce a second-order semiparametric efficient estimator. We also describe the nonparametric imputation method for treating censored observations, and study its properties. In Section 3, we analyze the A5175 study data using our modeling and estimation methods, after examining them via simulation studies. We conclude the paper with a discussion in Section 4, and relegate all the technical proofs to Appendix in the supplementary document.

2 Modeling and methodological development

2.1 RMM2 model in complete data

We first introduce the RMM2 model under the general complete data settings, we then define the specific model for the A5175 study data. Let Y_i, W_i denote the i.i.d. response random variables and covariates, respectively. In our paper, Y_i is the survival time on the logarithmic scale. Let $\boldsymbol{\beta}, \boldsymbol{\gamma}$ denote the parameters associated with the mean and the variance, respectively. A general RMM2 model has the form

$$g(Y_i) = m(W_i, \boldsymbol{\beta}) + \xi_i, \tag{1}$$

where $g(\cdot)$ is a known link function, $E(\xi_i|W_i) = 0$ and $E(\xi_i^2|W_i) = \sigma^2(W_i, \boldsymbol{\gamma})$. Here $m(\cdot)$ is a generic function known up to the parameter $\boldsymbol{\beta}$ and $\sigma^2(\cdot)$ is a generic positive function known up to the parameter $\boldsymbol{\gamma}$. Note that different from the usual regression models, the error variation is also specified as a function of W_i.

Based on Kim & Ma (2012), the semiparametric efficient estimator can be obtain by solving estimating equations formed by the sum of the following efficient score functions

$$\begin{aligned} \mathbf{S}_{\boldsymbol{\beta},\mathrm{eff}}(W_i, Y_i) &= \frac{\partial m(W_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \left\{ \frac{\xi_i}{\sigma^2(W_i, \boldsymbol{\gamma})} - \frac{E(\xi_i^3 | W_i) D_i}{\sigma^2(W_i, \boldsymbol{\gamma}) E(D_i^2 | W_i)} \right\} \\ \mathbf{S}_{\boldsymbol{\gamma},\mathrm{eff}}(W_i, Y_i) &= \frac{D_i}{E(D_i^2 | W_i)} \frac{\partial \sigma^2(W_i, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}, \end{aligned} \quad (2)$$

where

$$D_i = \xi_i^2 - \sigma^2(W_i, \boldsymbol{\gamma}) - E(\xi_i^3 \mid W_i) \xi_i / \sigma^2(W_i, \boldsymbol{\gamma}).$$

Note that when the third moment $E(\xi_i^3 | W_i) = 0$, the score function for $\boldsymbol{\beta}$ is the same as that for the OLS estimator. This fact shows that in estimating $\boldsymbol{\beta}$, the resulting estimator is at least as efficient, and is often more efficient compared with the OLS estimator. Further, if $E(\xi_i^3 | W_i) \neq 0$, the resulting estimator gains efficiency by making use of the additional variance structure. We point out that although the true third and fourth moments of ξ_i conditional on W_i are needed in the expression of (2), in practice, their parametrically or nonparametrically estimated versions can be plugged in and the resulting estimation efficiency of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ will not be affected (Kim & Ma 2012). In Section 3, we provide specific estimators of $E(\xi_i^3 \mid W_i)$ and $E(\xi_i^4 \mid W_i)$ both parametrically and nonparametrically using no additional data.

The above efficient score functions are derived under the complete data setting. In the next section, we modify the efficient score functions and introduce the estimating equations for censored survival data. We further derive the statistical properties of the resulting estimators.

2.2 The imputation estimator

The A5175 study data is complicated by censoring. More specifically, let T_i, C_i be the primary safety endpoint and the censoring time for the ith subject on the logarithmic scale. We observe only $X_i = \min(T_i, C_i)$ and the censoring indicator $\Delta_i = I(T_i \leq C_i)$, for $i = 1, \ldots, n$. A widely accepted method for handling censoring is the likelihood-based approach, such as that used for the AFT-Weibull model. Because of the full parameterization of the survival time distribution in AFT-Weibull models, the probability that an event happens after a certain time can be expressed as a function of a finite dimensional parameter. The parameter estimation can then be performed through maximizing the likelihood of the observed data. Although this method has long been known, its application is limited due to its nonrobustness, in that as soon as the true population distribution deviates from the AFT-Weibull model, the method leads to misleading results. In this paper, we introduce a nonparametric score imputation method to deal with the censored primary safety endpoints, which makes much less assumptions and is more robust. The method extends Wang et al. (2012)'s approach under the discrete setting by including the CD4 counts as a continuous covariate. Combined with the RMM2 model and the semiparametric efficient score equations, the method yields consistent estimators as long as the first two moment assumptions are satisfied.

Throughout the text, we use capital letters to denote the random variable and small letters to denote the corresponding realizations. For identifiability and simplicity, we assume the censoring distribution is independent of the survival time and the covariates. We consider the efficient score function $\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i, t_i) = (\mathbf{S}_{\boldsymbol{\beta},\mathrm{eff}}(w_i, t_i)^{\mathrm{T}}, \mathbf{S}_{\boldsymbol{\gamma},\mathrm{eff}}(w_i, t_i)^{\mathrm{T}})^{\mathrm{T}}$ for the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$, where w_i, t_i are the values of the CD4 counts, and the primary safety endpoint, respectively. We define the RMM2-ISE estimating equation under the survival settings as

$$\sum_{i=1}^{n} \delta_i \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i, t_i) + (1 - \delta_i) E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i, T_i) \mid T_i > X_i, W_i = w_i, X_i = x_i\}, \tag{3}$$

where δ_i is the realization of Δ_i. Thus, if a subject has an observed primary safety endpoint, we use the original efficient score function. However, if a subject is censored, we use the expected value of the score function conditional on the CD4 counts, given that no adverse reaction has happened before the censoring time.

Without specifying the population distribution of the primary safety endpoint, we evaluate the conditional expectation in model (3) nonparametrically via kernel method, which has good asymptotic properties with properly chosen bandwidth (Devroye 1981). We define

$$\mathbf{Q}_{\boldsymbol{\theta},i}(w_i, x_i) = E\left\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i, T_i) \mid T_i > X_i, W_i = w_i, X_i = x_i\right\}$$

$$
\begin{aligned}
&= E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,T_i) \mid T_i > x_i, W_i = w_i, C_i = x_i\} \\
&= \frac{E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,T_i) I(T_i > x_i) \mid W_i = w_i, C_i = x_i\}}{E\{I(T_i > x_i) \mid W_i = w_i, C_i = x_i\}} \\
&= \frac{E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,T_i) I(T_i > x_i) \mid W_i = w_i\}}{E\{I(T_i > x_i) \mid W_i = w_i\}},
\end{aligned}
$$

where the last equality is because C_i and T_i are independent given W_i. If T_i's are observed, we would simply use the nonparametric kernel regressions to approximate the two conditional expectations above. However, because T_i's are only observed when $\Delta_i = 1$, we need to further modify the two averages with the inverse probability weighted averages, where the weights are the probability of censoring time after event time, i.e. the survival function of the censoring process $G(\cdot \mid W) = G(\cdot)$ under the assumption that censoring is independent of the covariate. The kernel estimator of $\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}$ is thus written as

$$
\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i) = \frac{\sum_{j=1}^n \delta_j \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_j,x_j) I(x_j > x_i) K_h(w_j - w_i)/\widehat{G}(x_j)}{\sum_{j=1}^n \delta_j I(x_j > x_i) K_h(w_j - w_i)/\widehat{G}(x_j)}, \tag{4}
$$

where

$$
\widehat{G}(t_j) \equiv \prod_{x_i \le t_j} \left\{1 - \frac{(1-\Delta_i)}{\sum_{k=1}^n I(x_k \ge x_i)}\right\}.
$$

is the Kaplan-Meier estimator for the survival function of the censoring distribution $G(\cdot)$, and $K_h(\cdot) \equiv K(\cdot/h)/h$, where K is a kernel function and h is a bandwidth. When $h \to 0$, the imputed score functions reduce to the ones introduced in Wang et al. (2012) in the discrete covariate settings.

Specifically, to obtain $\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i)$, we use the product limit estimator to estimate G. We choose the Gaussian kernel with bandwidth $h = n^{-2/15} h_s$, where $h_s = 1.06\sigma n^{-1/5}$ is Silverman's rule-of-thumb bandwidth (page 45, Silverman (1986)), and σ is the standard deviation of W_i. Because h_s has the order of $n^{-1/5}$, the proposed bandwidth, h, satisfies $nh^4 \to 0$ and $nh^2 \to \infty$ when $n \to \infty$. Note that because of the indicators δ_j and $I(x_j > x_i)$, only the uncensored data from the individuals who have not met the safety event criteria at x_i contribute to the summations in $\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i)$. After computing $\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_j,t_j)$ and $\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i)$ for the uncensored and censored observations respectively, we obtain the RMM2-ISE estimators $\widehat{\boldsymbol{\theta}}$ through solving the estimating equation

$$
\sum_{i=1}^n \delta_i \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,t_i) + (1-\delta_i)\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i) = 0. \tag{5}
$$

Under the Assumptions A1–A8 listed in Appendix A.1, we rigorously establish the consistency and asymptotic properties of the estimator, i.e. we obtain $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_p(1)$, and $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to N\{0, A^{-1}\boldsymbol{\Omega}(A^{-1})^{\mathrm{T}}\}$ in distribution, where $A, \boldsymbol{\Omega}$ are defined in Theorem 2 of the Appendix. We elaborate the consistency and asymptotic normality of the RMM2-ISE in Theorems 1 and 2 followed by their detailed proofs in Appendix in the supplementary document.

3 Analysis of the A5175 Study Data

We are now ready to analyze the A5175 study data using the RMM2-ISE method. Before the analysis, we first perform a numerical evaluation of the estimation procedure on simulated samples and compare the estimation results with the IWLS method introduced in Section 1. The IWLS estimator is obtained by solving (5), but with $\mathbf{S}_{\boldsymbol{\theta},\text{eff}}$ in it replaced by $\sigma^{-2}(W_i)\xi_i\partial m(W_i, \boldsymbol{\beta})/\partial\boldsymbol{\beta}$, which is the score function associated with weighted least square method. Here $\sigma^2(W_i)$ is the conditional variance of ξ_i given W_i, which can be replaced by its consistent estimator. The consistent estimator can be obtained by using the non-censored observations, because our score function is first constructed for the fully observed samples which only relies on the $\sigma^2(W_i)$ for the non-censored observation. We discuss several different ways of estimating $\sigma^2(W_i)$ later in this section. Note that the same replacement of $\mathbf{S}_{\boldsymbol{\theta},\text{eff}}$ is needed in calculating $\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i, x_i)$ in (5). The asymptotic variance of the IWLS estimator can be shown to be the same as $\boldsymbol{\Omega}$ in Theorem 2, except that $\mathbf{S}_{\boldsymbol{\theta},\text{eff}}$ needs to be replaced by $\sigma^{-2}(W_i)\xi_i\partial m(W_i, \boldsymbol{\beta})/\partial\boldsymbol{\beta}$ and $\mathbf{Q}_{\boldsymbol{\theta}_0,i}$ is also adapted correspondingly. It is readily seen that the asymptotic estimation variances of the RMM2-ISE and the IWLS methods have the same structure except for the different forms of $\mathbf{S}_{\boldsymbol{\theta},\text{eff}}$. This suggests that, intuitively the RMM2-ISE method would have better asymptotic efficiency, because the score function for RMM2-ISE is more efficient than that for IWLS (Wang & Leblanc 2008, Kim & Ma 2012) in the complete data settings, and the kernel imputation induces the same type of asymptotic variance inflation for both methods when the data is subject to censoring. We explore the required sample sizes and censoring rates for implementing the RMM2-ISE procedure, and show that the procedure yields accurate estimators under reasonable uncensored sample sizes. Moreover, we show via simulation that the RMM2-ISE method gains efficiency compare with the simpler IWLS method when the third moment of the error distribution does not vanish. These conclusions are crucial, because they support the applications of the RMM2-ISE method to the A5175 study data.

3.1 Evaluation of methods

We illustrate the relative performance of the RMM2-ISE estimator and the IWLS estimator through demonstrating that the former is more efficient than the latter. Note that we use the same imputation method in both estimation procedures.

In the complete data setting, the RMM2-ISE estimator is shown to be more efficient than the IWLS estimator when the conditional third moment of the error distribution is nonzero (Wang & Leblanc 2008, Kim & Ma 2012). To illustrate this point as well as the consistency of the estimators under the setting with censoring, we generate the data as the following. The covariate W_i is the logarithm of a random variable generated from the Uniform (0, 5) distribution. The error term $\xi_i = \chi^2(k_i) - k_i$, where $\chi^2(k_i)$ is generated from the chi-squared distribution with the degree of freedom $k_i = (\gamma_0 + \gamma_1 W_i^2)/2$. Note that the variance of ξ_i is $\sigma^2(W_i, \boldsymbol{\gamma}) = 2k_i$, which depends on the covariate, and $E(\xi_i^3|W_i)$ does not vanish. We generate the time to safety endpoint T_i from the exponential model

$$\log T_i = \beta_0 \exp(\beta_1 W_i) + \xi_i. \tag{6}$$

We further generate the censoring time from exponential distributions. We vary the exponential rate parameters to obtain various censoring rates. We assess the performances of the RMM2-ISE estimator and the IWLS estimator at the different censoring rates.

Following model (2), we obtain the semiparametric efficient score functions for the above model as

$$\begin{aligned}
\mathbf{S}_{\boldsymbol{\beta},\mathrm{eff}}(W,T) &= \{\exp(\beta_1 W), \beta_0 \exp(\beta_1 W) W\}^{\mathrm{T}} \left\{ \frac{\xi}{\sigma^2(W,\boldsymbol{\gamma})} - \frac{E(\xi^3 \mid W) D}{\sigma^2(W,\boldsymbol{\gamma}) E(D^2 \mid W)} \right\} \\
\mathbf{S}_{\boldsymbol{\gamma},\mathrm{eff}}(W,T) &= (1, W^2)^{\mathrm{T}} \frac{D}{E(D^2 \mid W)}.
\end{aligned}$$

We then impute the above score functions as described in Section 2 to estimate the parameters.

We use the true $E(\xi_i^3|W_i), E(\xi_i^4|W_i)$ to obtain the RMM2-ISE estimator, and use the true $E(\xi_i^2|W_i)$ to form the optimal weights $1/\sigma^2(W_i, \boldsymbol{\gamma}_0)$ to obtain the IWLS estimator. This guarantees that both estimators achieve their optimal performance in the complete data setting. In other words, we avoid the hidden efficiency loss due to possible misspecification of moment functions in both estimators to keep the comparison fair. We compare the biases and variances of the resulting RMM2-ISE and IWLS estimators in all the numerical experiments.

3.2 Numerical results for the estimation procedures

We use a sample size of $n = 400$ and generate 1000 data sets from model (6), with $\beta_0 = 1$, $\boldsymbol{\gamma} = (1, 0.1)$. In Table 1, we present the performance of the RMM2-ISE estimator and the IWLS estimator under different specifications of β_1 and censoring rates. Here $E(\xi_i^3|W_i)$ is estimated through fitting a linear model between $\widehat{\xi_i^3}$ and the covariates, and $E(\xi_i^4|W_i)$ is estimated through fitting a quadratic model between $\widehat{\xi_i^4}$ and the covariates. Here $\widehat{\xi_i}$'s are the residuals after fitting a linear regression for the non-censored observations. The linear model is simple and the most common regression model in practice, while the quadratic model ensures the nonnegativeness of the regression function. We first fit the working model based on the non-censored residuals and covariates, then use the fitted model to impute the additional censored moments. Note that neither the linear nor the quadratic model is the true model of these conditional moments. However, for the IWLS method, we used $E(\xi_i^2|W_i)$ under the true model. This means that we compared a sub-optimal RMM2-ISE method with the optimal IWLS method. Hence theoretically there is no guarantee that the RMM2-ISE estimator should outperform the IWLS estimator. We used this particularly harsh setting for the RMM2-ISE estimator to test its performance stability and robustness to the working models. As we can see, if no observation is censored (the censoring rate is 0%), both estimates are close to the true values. This illustrates the consistency of the estimators when no observation is censored. Further, the RMM2-ISE estimator has smaller biases and variances compared with the IWLS estimator, which illustrates the better accuracy and efficiency of the RMM2-ISE estimator compared to the IWLS estimator. When the censoring rate is greater than 0, the RMM2-ISE estimator continues to perform well. In fact, even when the censoring rate is moderately large (25%), the RMM2-ISE estimation is still close to the truth (with less than 0.1 absolute biases). Because censoring reduces the information contained in the sample for inferring the population distributions, both the RMM2-ISE and IWLS estimators start to deteriorate when the censoring rates further increase. However, the RMM2-ISE estimator has smaller deterioration compared with the IWLS estimator under all situations. For example, the IWLS estimator for β_0 shows more than 0.1 absolute biases when the censoring rate is 15%, while the corresponding RMM2-ISE estimator keeps the absolute biases within 0.1 until the censoring rate reaches 50%. Compared to the estimation of β_0, the RMM2-ISE and IWLS methods perform better in estimating the parameter of clinical interest β_1. Nevertheless, the IWLS estimator has biases greater than 0.1 when the censoring rate is 50%, while this occurs for the RMM2-ISE estimator only when the censoring rate reaches 75%. Overall, compared with the IWLS estimator, the RMM2-ISE estimator generally has smaller biases in estimating $\boldsymbol{\beta}$. The standard deviations of the RMM2-ISE estimator are smaller than those of the IWLS estimator on average. In conclusion, the RMM2-ISE method performs better than the IWLS method in terms of smaller biases and variations of the resulting estimation. In the simulation studies, we see that the bias increases when the censoring rate increases. Compared with $\widehat{\boldsymbol{\beta}}$, $\widehat{\gamma}$ has larger bias and variance. However, this does not indicate that the estimator is inconsistent. In fact, when we further increase the sample size, we observe a clear reduction in the biases. Thus, the relatively large bias at high censoring rate we observe here is a finite sample phenomenon.

In Table 2, we compare the estimated asymptotic standard deviation derived in Theorem 2 with the empirical estimation standard deviation summarized from the simulated samples. The results show that when the censoring rate is small ($\leq 25\%$), the asymptotic standard deviation estimators are close to the empirical ones, while their performance deteriorates when the censoring rate increases. In the latter case, it may be preferable to use the bootstrap method to assess the estimation variability, as suggested in Ma & Yin (2010) and Wang et al. (2012). For example, we performed additional bootstrap method for the 50% censoring rate case in Table 2. The resulting bootstrap standard deviation is (0.046 0.079 0.122 0.065), which is much closer to the empirical standard deviation (0.062, 0.071, 0.137, 0.071) than the estimated asymptotic standard deviation (0.028, 0.028, 0.102, 0.075).

In the above evaluations, we demonstrate that the RMM2-ISE method can accurately estimate the covariate effect when the sample size is more than 400 and censoring rate is less than 50%. Further, the RMM2-ISE estimator has better efficiency and smaller mean squared errors than the IWLS estimator. This encourages us to use the RMM2-ISE method to analyze the A5175 study data, as we demonstrate in the next section. Moreover, we show that the asymptotic standard deviations are close to the true ones when the observed sample size is sufficient. Finally, we show that the misspecification of $E(\xi_i^3|W_i)$ and $E(\xi_i^4|W_i)$ does not affect the estimations for the parameters β_0, β_1. Thus, in practice, we can estimate the conditional moments roughly by constructing simple models between W and the power functions of the residuals, such as the linear models. This is also justified in Wang et al. (2008), which shows that the estimation procedures using the true and the estimated moment functions have similar performance.

Finally, we also perform the simulation studies when $E(\xi_i^2|W_i)$ in the IWLS method, and $E(\xi_i^3|W_i)$, $E(\xi_i^4|W_i)$ in the RMM2-ISE method are estimated using the nonparametric Nadaraya–Watson kernel method for sample size 800. Compared with IWLS, RMM2-ISE gives less biased result and has smaller variation for estimating the covariate effect. In general, the estimators in Table 3 show larger biases and variations compared to the results in Table 1.

3.3 Analysis of the A5175 study data

We apply the RMM2-ISE method to the A5175 study data, which aims to evaluate the safety of the antiretroviral regimens. We find that the RMM2-ISE gives a better fit to the A5175 study data compared with the commonly used AFT-Weibull-ML method.

We use a total of 1008 patients who have been assigned to the open-label antiretroviral therapy with efavirenz plus lamivudine-zidovudine (EFV+3TC-ZDV) and atazanavir plus didanosine-EC plus emtricitabine (ATV+DDI+FTC) treatment arms. A total of 460 patients have their safety events censored, resulting in a censoring rate of 46%. For each patient, we compute the mean of the CD4 counts before his/her safety event occurs. To stabilize the numerical computations, we standardize the event times and mean CD4 counts by their sample standard deviations, which are approximately 40 and 160, respectively. The transformation is monotone so that it does not affect the following inference.

We denote the standardized event time as T_i, the logarithm of the standardized mean CD4 counts as W_i. We first fit the complete data with the linear model

$$\log T_i = \beta_0 + \beta_1 W_i + \xi_i,$$

such that $E(\xi_i|W_i) = 0$. Note that, here we only use the non-censored cases to do the initial analysis because our score functions in (2) is only constructed for the non-censored cases. Further, the data set contains 548 observed survival times, it is sufficient to reveal the general pattern of the error distribution. We plot the residuals $\hat{\xi}_i = \log T_i - \widehat{\beta}_0 - \widehat{\beta}_1 W_i$ versus the covariate in Figure 1(A), where $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are the least square estimators of β_0, β_1, respectively. The residuals are centered at zero which suggests the model is adequate to capture the mean structure. Further, the error variation becomes larger when the covariate value increases, which implies a dependency of the error variance on the covariate. To explore this dependency, we plot the residual squares $\hat{\xi}_i^2$ versus the covariates in Figure 1(B). The plot shows that the variation has a nonlinear relation with W_i. We therefore enrich the linear mean model by further modeling the variance $\sigma^2(W_i, \boldsymbol{\gamma})$. We considered various nonlinear forms of $\sigma^2(W_i, \boldsymbol{\gamma})$ and found the form $\sigma^2(W_i, \boldsymbol{\gamma}) = (\gamma_0 + \gamma_1 W_i)^2$ both adequate and parsimonious, in that it captures the variability pattern well, and it is simple and yields the smallest estimation variability for $\widehat{\boldsymbol{\beta}}$, and this $\widehat{\boldsymbol{\beta}}$ is closest to the one from the IWLS method among all the nonlinear models we experimented. Because the misspecification of σ^2 may lead to inconsistent estimators, in practice, we suggest to first use proper variance modeling tools, such as graphical tools, to determine suitable functional forms for $\sigma^2(W_i, \boldsymbol{\gamma})$. After that, we can select the resulting $\widehat{\boldsymbol{\beta}}$ from RMM2-ISE which are reasonably close to the one from IWLS, because IWLS is always a consistent method regardless of whether the variance form is correctly specified. Finally, we can refine our choices by comparing the variances of $\widehat{\boldsymbol{\beta}}$ among the possible candidate variance models.

We implement the RMM2-ISE estimation on this specific model, and obtained the estimates $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\gamma}_0, \widehat{\gamma}_1) = (-0.75, 1.00, 1.25, -0.047)$, with associated standard errors $\{\text{sd}(\widehat{\beta}_0), \text{sd}(\widehat{\beta}_1), \text{sd}(\widehat{\gamma}_0), \text{sd}(\widehat{\gamma}_1)\} = (0.047, 0.056, 0.021, 0.049)$. The 95% confidence intervals for the parameters $(\beta_0, \beta_1, \gamma_0, \gamma_1)$ are $\{(-0.84, -0.66),\ (0.89, 1.11),\ (1.20, 1.29),\ (-0.14, 0.049)\}$, which show significant effect of the CD4 cell counts on the primary safety endpoint. The covariate effect, γ_1, is not significant, which coincides with the local regression line we added in Figure 1(B). The local regression technique was proposed by Cleveland (1979). It uses local segments of data to build a function nonparametrically to describe the relation between the response and the covariate. It can be seen that, the local regression line is nearly flat, which suggests there is no statistically significant effect from the covariate. We also perform IWLS estimation and obtain $(\widehat{\beta}_0, \widehat{\beta}_1) = (-0.74, 0.99)$, with associated standard errors $\{\mathrm{sd}(\widehat{\beta}_0), \mathrm{sd}(\widehat{\beta}_1)\} = (0.052, 0.056)$. Note that to obtain the second moment $\sigma^2(W_i)$ as the weight, we first form the regression residuals. Then we propose a working model for $\sigma^2(W_i)$ the same as the second moment model used in the RMM-ISE method, i.e., let $\sigma^2(W_i) = (\gamma_0 + \gamma_1 W_i)^2$, and then perform the usual regression analysis to estimate the parameters in the model and hence obtain the second moment. The results show that the RMM2-ISE estimation is as efficient as the IWLS method. The similar efficiency is not unexpected, because as shown in Figure 1(C), the estimated conditional third moments of the error terms, i.e. $E(\widehat{\xi}_i^3 | W_i)$, are nearly 0. In fact, when we regress $\widehat{\xi}_i^3$ on W_i, the resulting intercept is 0.0036 with confidence interval (-0.02, 0.020), and the resulting covariate effect is 0.0004 with the confidence interval (-0.0044, 0.011). However, from another aspect, the analysis does demonstrate that the RMM2-ISE method is at least as efficient as the IWLS method. Therefore we employ the RMM2-ISE method for the subsequent analyses which ensures the estimators have variances no greater than those resulting from the IWLS method.

To compare the performance of the RMM2-ISE method with that of the commonly used AFT-Weibull-ML method for the Weibull model, we calculated the mean squared residuals on the logarithmic scale based on the 548 fully observed samples, and obtained the values 1.93 for the RMM2-ISE and 4.69 for the AFT-Weibull-ML method respectively. The comparison based on the observed samples was justified and suggested by Little (1992) in the missing at random framework, which is the setting that the non–informative censoring belongs. To avoid overfitting, we performed an additional 2-fold cross validation. The cross validation errors (mean squared predictive error) for the proposed method and AFT-Weibull-ML method are 1.89 and 4.23 respectively, indicating that the proposed method outperforms the AFT-Weibull-ML method. The RMM2-ISE method provides a much better fit to the data than the AFT-Weibull-ML method, which also implies that the survival time distribution deviates from Weibull.

After demonstrating the better performance of RMM2-ISE in fitting the A5175 study data, we continue to explore the relation between the CD4 counts and the time to primary safety endpoint in subgroups. We further divided the sample by gender and analyze the CD4 counts effects for 479 females and 529 males separately. The estimated $\boldsymbol{\beta}$ in the female group is $(\widehat{\beta}_0, \widehat{\beta}_1) = (0.14, 0.16)$, the standard errors $\{\mathrm{sd}(\widehat{\beta}_0), \mathrm{sd}(\widehat{\beta}_1))\} = (0.14, 0.21)$, which gives the confidence intervals $\{(-0.13, 0.41), (-0.25, 0.57)\}$. The estimated $\boldsymbol{\beta}$ in the male group is $(\widehat{\beta}_0, \widehat{\beta}_1) = (-0.36, 0.31)$, the standard errors $\{\mathrm{sd}(\widehat{\beta}_0), \mathrm{sd}(\widehat{\beta}_1))\} = (0.12, 0.14)$, which gives the confidence intervals $\{(-0.60, -0.12), (0.04, 0.58)\}$. In the female group, the CD4 counts do not have a significant positive effect on the primary safety endpoints, while the effect is significant in the male group. Further, the CD4 counts effect is higher in the male patients than in the female patients. It is worth mentioning that when the AFT-Weibull model is used, no difference between female and male patients can be discovered. In this case, the estimator are $(0.91, 0.32)$, $(0.96, 0.31)$, the standard deviations are $(0.100, 0.09)$, $(0.109, 0.112)$ and the 95% confident intervals are $\{(0.69, 1.13), (0.11, 0.55)\}$, $\{(0.76, 1.16), (0.13, 0.49)\}$ for the females and males, respectively. In practice, because the CD4 counts are positively related to time to adverse events, we suggest giving the antiretroviral regimens at higher CD4 counts level to prevent severe side-effects from the drugs. Further, because the CD4 counts effects are different in the two genders, we suggest differentiating the drug scheduling for men and women.

Using the RMM2-ISE method, we develop a strategy to personalize the drug scheduling based on the A5175 data, where the patients are all enrolled at the beginning of the trial and continuously monitored in the trial, as we now describe. We first define a safety cut-off value regarding the primary safety endpoint. The drug usage is considered to be safe for a patient if the patient's estimated primary safety endpoint is later than the cut-off value. A patient's CD4 counts are taken at the beginning of the trial (baseline), week 8, week 48, week 72, etc. At a measurement time, say at week 48, we collect the CD4 counts information on each patient, and collect his/her primary safety event time or his/her censoring time if either has happened. For the patient who has not experienced primary safety time and who has not been censored, we use the measurement time as censoring time. We then use the average observed CD4 counts and the event/censoring time to obtain the estimator for the coefficients, i.e., $\boldsymbol{\beta}$. Then for any patient who has not experienced the primary safety event at the 48th week, we use the estimate $\widehat{\boldsymbol{\beta}}$ and his/her average observed CD4 counts to predict his/her primary safety event time. If the predicted primary safety event time is to the right of the safety cut-off value, the treatment is considered safe for the patient. This patient is eliminated from the current trial and move to the next treatment phase. We perform this estimation and prediction procedure at weeks 8, 48, 72 and make corresponding decisions at each measurement time based on the remaining patients in the trial.

We use the 75% sample quantile of the standardized primary safety endpoints, 2 (corresponding to 79.14 in the original data), as a sample cut-off value. In practice, different and possibly more meaningful cut-off values can be chosen based on existing medical knowledge. We choose to start to treat a patient with the antiretroviral regimens when the lower bound of the estimated confidence interval for the mean of $\log T_i$, i.e., $\widehat{\beta}_0 + \widehat{\beta}_1 W_i - 1.96 \left\{(1, W_i)^{\mathrm{T}} \widehat{\Sigma} (1, W_i)\right\}^{1/2}$ is greater $\log(2)$, where $\widehat{\Sigma}$ is the estimated variance-covariance matrix for $\widehat{\boldsymbol{\beta}}$. We perform the analysis in the following three groups of patients. Group 1 contains patients who only have baseline CD4 counts recorded. Groups 2 contains patients who have the CD4 counts measured at and before the 48th week. Group 3 contains patients who have the CD4 counts measured at and before the 96th week. The results show that in Group 1, 95 out of 188 (50.5%) patients have the lower bound of the estimated confidence interval smaller than log(2). Further in Group 2 and Group 3, the ratios are 132 out of 201 (66%) and 94 out of 131 (72%), respectively. Therefore, in these three groups, we can start to treat 50.5%, 66%, 72% of the patients at the baseline randomization time, 48th week, or the 96th week, respectively. Since the CD4 counts are obtained prior to the primary safety endpoints, the strategy allows the patients to be treated earlier when the evidences of the treatment safety are sufficient, and thus improves the efficiency of delivering the safe treatments to the patients.

In conclusion, we compared the RMM2-ISE method with the AFT-Weibull-ML method.

The RMM2-ISE outperforms the AFT-Weibull-ML method in giving smaller fitted mean of the squared residuals. Further, we discovered that the positive CD4 counts effect in men are higher than that in women on average, while this pattern is not captured by the AFT-Weibull-ML method. Finally, we propose a strategy for personalizing drug scheduling based on the mean of the repeatedly measured CD4 counts. The strategy allows early treatment delivery to the patients based on their CD4 counts information, and ultimately enhance the treat efficiency.

4 Discussion

This work is motivated by the A5175 study (Campbell et al. 2012). We intend to use the short term CD4 counts to infer the primary safety endpoints. The complex data configuration motivates us to construct the RMM2 model which models the additional variance structures observed from the data. We propose the nonparametric imputed version of the semiparametric efficient method for parameter estimations to handle censoring. The theoretical derivations show that the resulting estimators are consistent and asymptotically normally distributed. The efficiency of the RMM2-ISE estimators is demonstrated to be better than that of the IWLS estimators. When fitting the A5175 study data, the RMM2-ISE method outperforms the AFT-Weibull-ML method in terms of having smaller mean squared residuals.

In the A5175 data analysis, due to the limitation of the univariate kernel specification, we did not include multiple covariates in the regression function. The method can be extended to include multiple covariates through utilizing multivariate kernels. Such extension will enhance the applicability of the model in more general situations.

In conclusion, to analyze the A5175 data, the RMM2 model avoids the model assumptions on the full likelihood and is more flexible than the parametric models. Further, in terms of parameter estimation, the RMM2-ISE method takes advantage of the additional information in the variance structure and has better efficiency than the imputed weighted least squares method. In general, the RMM2-ISE approach provides a more robust and efficient way in analyzing post-trial data.

We have assumed that the censoring process is independent of the covariates and the survival process for simplicity. This assumption can be relaxed to allow the censoring time to depend on the covariates w_j. In this case, we can use a nonparametric kernel based Kaplan-Meier estimator

$$\widehat{G}(t_j \mid W_j = w_j) = \prod_{x_i \leq t_j} \left\{ 1 - \frac{(1-\Delta_i) K_h(w_i - w_j)}{\sum_{k=1}^{n} I(x_k \geq x_i) K_h(w_k - w_j)} \right\}$$

in (4). However, the subsequent development will also need to be adapted to reflect the covariate-dependent nature of the censoring process and the analysis will be more complex.

References

Campbell, T. B., Smeaton, L. M., Kumarasamy, N., Flanigan, T., Klingman, K. L., Firnhaber, C., Grinsztejn, B., Hosseinipour, M. C., Kumwenda, J., Lalloo, U., Riviere, C., Sanchez, J., Melo, M., Supparatpinyo, K., Tripathy, S., Martinez, A. I., Nair, A., Walawander, A., Moran, L., Chen, Y., Snowden, W., Rooney, J. F., Uy, J., Schooley, R. T., De Gruttola, V., Hakim, J. G. & study team of the ACTG, P. (2012), 'Efficacy and safety of three antiretroviral regimens for initial treatment of hiv-1: a randomized clinical trial in diverse multinational settings', *PLoS Med.* **9**(8), e1001290.

Cleveland, W. S. (1979), 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American statistical association* **74**(368), 829–836.

Devroye, L. (1981), 'On the almost everywhere convergence of nonparametric regression function estimates.', *Ann. Statist.* **9**(6), 1310.

Fleming, T. R. & Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, Wiley series in probability and mathematical statistics: Applied probability and statistics, New York, N.Y. : Wiley, c1991.

Gallant, A. R. (2009), *Nonlinear Statistical Models*, Vol. 310, Wiley. com.

Gill, R. (1980), *Censoring and Stochastic Integrals*, Vol. 124, Amsterdam: Mathematisch Centrum, Netherlands.

Hirsch, M. S. (2008), 'Initiating therapy: when to start, what to use', *J. Infect. Dis.* **197**(Supplement 3), S252–S260.

Kim, M. & Ma, Y. (2012), 'The efficiency of the second-order nonlinear least squares estimator and its extension', *Annals of the Institute of Statistical Mathematics* **64**, 751–764.

Klein, J. P. & Moeschberger, M. L. (2010), *Survival Analysis: Techniques for Censored and Truncated Data*, Statistics for Biology and Health, New York : Springer, c2003.

Lipsitz, S. R., Ibrahim, J. G. & Zhao, L. P. (1999), 'A weighted estimating equation for missing covariate data with properties similar to maximum likelihood', *Journal of the American Statistical Association* **94**(448), 1147–1160.

Little, R. J. (1992), 'Regression with missing x's: a review', *Journal of the American Statistical Association* **87**(420), 1227–1237.

Ma, Y. & Yin, G. (2010), 'Semiparametric median residual life model and inference.', *Can. J. Statist.* **38**(4), 665 – 679.

Robins, J. M. & Rotnitzky, A. (1992), Recovery of information and adjustment for dependent censoring using surrogate markers, *in* 'AIDS Epidemiol.', Springer, New York, pp. 297–331.

Silverman, B. W. (1986), *Density estimation for statistics and data analysis / B.W. Silverman.*, Monographs on Statistics and Applied Probability: 26, London ; New York : Chapman and Hall, 1986.

Wang, L. & Leblanc, A. (2008), 'Second-order nonlinear least squares estimation.', *Ann. Int. Statist. Math.* **60**(4), 883 – 900.

Wang, S., Joshi, S., Mboudjeka, I., Liu, F., Ling, T., Goguen, J. D. & Lu, S. (2008), 'Relative immunogenicity and protection potential of candidate yersinia pestis antigens against lethal mucosal plague challenge in balb/c mice.', *Vaccine* **26**(13), 1664–1674.

Wang, Y., Garcia, T. P. & Ma, Y. (2012), ‘Nonparametric estimation for censored mixture data with application to the cooperative huntington’s observational research trial.’, *J. Am. Statist. Assoc.* **107**(500), 1324 – 1338.

Table 1: Comparisons of the optimal imputed weighted least squares (IWLS) estimator and the imputation-based semiparametric efficient (RMM2-ISE) estimator. Sample size, $n = 400$, $\beta_0 = 1$, $\boldsymbol{\gamma} = (1, 0.1)^{\mathrm{T}}$. SD represents the sample empirical standard deviation based on 1000 simulations.

Truth		IWLS				RMM2-ISE					
β_0	β_1	$\widehat{\beta}_0$	$\widehat{\beta}_1$	SD($\widehat{\beta}_0$)	SD($\widehat{\beta}_1$)	$\widehat{\beta}_0$	$\widehat{\beta}_1$	SD($\widehat{\beta}_0$)	SD($\widehat{\beta}_1$)	$\widehat{\gamma}_0$	$\widehat{\gamma}_1$
0% censoring rate											
1.0	-0.2	0.954	-0.195	0.053	0.041	0.972	-0.196	0.048	0.036	0.802	0.071
1.0	-0.4	0.958	-0.393	0.056	0.033	0.974	-0.399	0.046	0.030	0.794	0.078
1.0	-0.6	0.966	-0.593	0.060	0.034	0.980	-0.598	0.049	0.030	0.797	0.091
1.0	-0.8	0.970	-0.790	0.070	0.040	0.988	-0.793	0.054	0.037	0.840	0.097
15% censoring rate											
1.0	-0.2	0.895	-0.187	0.053	0.045	0.951	-0.190	0.042	0.041	0.728	0.068
1.0	-0.4	0.894	-0.391	0.057	0.043	0.951	-0.392	0.045	0.037	0.730	0.078
1.0	-0.6	0.897	-0.595	0.067	0.047	0.956	-0.588	0.048	0.041	0.742	0.084
1.0	-0.8	0.894	-0.803	0.074	0.061	0.957	-0.786	0.055	0.048	0.751	0.093
25% censoring rate											
1.0	-0.2	0.851	-0.184	0.057	0.049	0.925	-0.180	0.046	0.049	0.744	0.063
1.0	-0.4	0.850	-0.396	0.059	0.047	0.926	-0.385	0.047	0.044	0.742	0.067
1.0	-0.6	0.849	-0.607	0.066	0.057	0.929	-0.587	0.055	0.053	0.755	0.068
1.0	-0.8	0.842	-0.823	0.073	0.073	0.927	-0.787	0.059	0.062	0.768	0.078
50% censoring rate											
1.0	-0.2	0.736	-0.193	0.057	0.053	0.848	-0.185	0.054	0.054	0.794	0.042
1.0	-0.4	0.729	-0.425	0.065	0.057	0.855	-0.386	0.055	0.059	0.776	0.059
1.0	-0.6	0.722	-0.665	0.075	0.082	0.853	-0.605	0.063	0.073	0.787	0.067
1.0	-0.8	0.709	-0.912	0.088	0.122	0.848	-0.814	0.071	0.097	0.792	0.070
75% censoring rate											
1.0	-0.2	0.581	-0.248	0.055	0.074	0.739	-0.218	0.061	0.073	0.843	0.034
1.0	-0.4	0.577	-0.558	0.063	0.123	0.741	-0.465	0.058	0.092	0.838	0.033
1.0	-0.6	0.567	-0.889	0.072	0.200	0.736	-0.711	0.072	0.138	0.818	0.048
1.0	-0.8	0.556	-1.245	0.080	0.309	0.724	-0.953	0.076	0.191	0.795	0.069

Table 2: Estimation variations when $\beta_0 = 1, \beta_1 = -0.6, \gamma_0 = 1, \gamma_1 = 0.1$: SD represents the empirical standard deviation from the 1000 simulation runs. $\widehat{\mathrm{SD}}$ represents the theoretic asymptotic standard derivation.

Censoring rate	SD($\widehat{\beta}_0$)	SD($\widehat{\beta}_1$)	$\widehat{\mathrm{SD}}$($\widehat{\beta}_0$)	$\widehat{\mathrm{SD}}$($\widehat{\beta}_1$)	SD($\widehat{\gamma}_0$)	SD($\widehat{\gamma}_1$)	$\widehat{\mathrm{SD}}$($\widehat{\gamma}_0$)	$\widehat{\mathrm{SD}}$($\widehat{\gamma}_1$)
0%	0.049	0.030	0.045	0.031	0.171	0.095	0.254	0.118
15%	0.048	0.041	0.054	0.039	0.128	0.100	0.174	0.124
25%	0.055	0.053	0.045	0.036	0.127	0.088	0.161	0.123
50%	0.063	0.073	0.028	0.028	0.137	0.071	0.102	0.075

Table 3: Estimation results from 1000 simulation runs when $n = 800, \beta_0 = 1, \beta_1 = -0.6, \gamma_0 = 1, \gamma_1 = 0.1$. $E(\xi^2|W), E(\xi^3|W), E(\xi^4|W)$ are estimated by the nonparametric kernel regression method.

	IWLS				RMM2-ISE					
Censoring rate	$\widehat{\beta}_0$	$\widehat{\beta}_1$	SD($\widehat{\beta}_0$)	SD($\widehat{\beta}_1$)	$\widehat{\beta}_0$	$\widehat{\beta}_1$	SD($\widehat{\beta}_0$)	SD($\widehat{\beta}_1$)	$\widehat{\gamma}_0$	$\widehat{\gamma}_1$
0%	0.979	-0.638	0.089	0.094	0.978	-0.595	0.067	0.054	0.896	0.070
15%	0.908	-0.654	0.075	0.100	0.948	-0.598	0.072	0.061	0.598	0.066
25%	0.848	-0.671	0.084	0.129	0.902	-0.616	0.067	0.054	0.418	0.073
50%	0.754	-0.710	0.072	0.101	0.806	-0.651	0.071	0.061	0.226	0.062

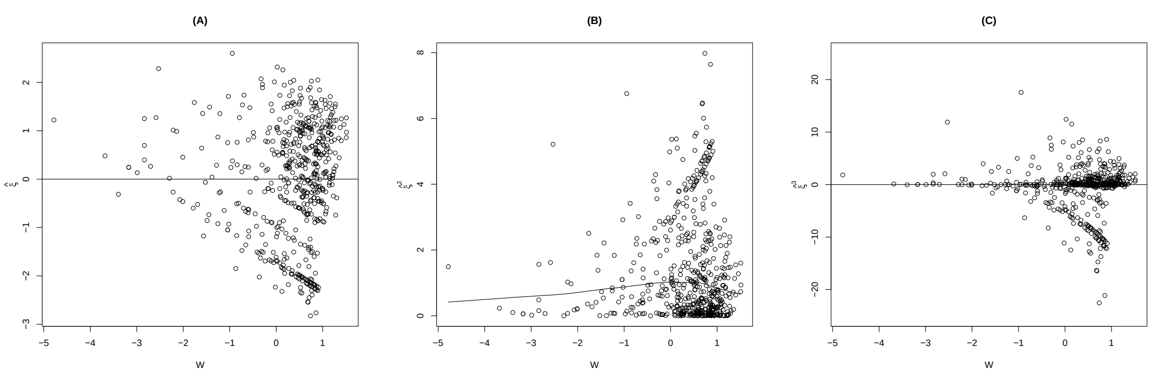

Figure 1: The preliminary analysis results for the A5175 study data. (A) the residual versus the covariate, (B) the residual squared versus the covariate and a local regression line describing the relation between ξ^2 and the covariate, (C) the scatter plot of the covariate–residual cubed and the estimated third moment of the error distribution as a function of the covariate.

Appendix

A.1 Assumptions

We first state the regularity conditions under which the RMM2-ISE estimator has good asymptotic properties.

A1: The kernel function $K(\cdot)$ is nonnegative, has compact support, and satisfies $\int K(s)ds = 1$, $\int K(s)sds = 0$ and $\int K(s)s^2ds < \infty$, $\int K^2(s)ds < \infty$.

A2: The bandwidth for the kernel function satisfies $nh^4 \to 0$, $nh^2 \to \infty$ as $n \to \infty$.

A3: The cumulative hazard function for the censoring time $\Lambda^c(t) < \infty$, for all t.

A4: Let $B(h,u) = E\{hI(T \geq u)\}/S(u-)$. Then $B\left[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,T_i)\{1-G(T_i)\},u\right]^2 < \infty$, $B\{\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,C_i)I(T_i > C_i),u\}^2 < \infty$ where $S(u-) = Pr(T > u)$.

A5: $\tau \equiv \inf\{t : G(t) = 0\} < \infty$.

A6: Let $\mathbf{a}^{\otimes 2} \equiv \mathbf{a}\mathbf{a}^{\mathrm{T}}$ throughout the text, then $E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,T_i)^{\otimes 2}\} < \infty$ and $E\{\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,X_i)^{\otimes 2}\} < \infty$, for all $\boldsymbol{\theta}$.

A7: There exists an open set Θ that contains the true parameter $\boldsymbol{\theta}_0$. In Θ, $E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,t_i)\} = 0$ has a unique solution at $\boldsymbol{\theta}_0$.

A8: Let

$$
\begin{aligned}
U_0(\boldsymbol{\theta}) &= E[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,T_i) - E\{\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,C_i)I(C_i < T_i)|W_i,X_i\}] \\
&\quad + E\{(1-\Delta_i)\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,X_i)\}.
\end{aligned}
$$

U_0 is continuous in $\boldsymbol{\theta} \in \Theta$, and has derivative bounded away from 0 and ∞.

Assumptions (A1) and (A2) ensure the consistency of the kernel estimator $\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i)$. Assumptions (A3) and (A4) guarantee the consistency of $\widehat{G}$ in approximating G, the censoring time distribution function. Assumption (A5) ensures that at any finite time, there is a positive chance that safety endpoint can be observed. Finally, Assumption (A6) ensures the boundedness of the asymptotic estimation variances. Assumption (A7) is the usual condition for identifiability of parameters. Assumption (A8) is usually to ensure the score function is continuous and differentiable.

A.2 Notations

We accept that assumptions A1-A5 hold throughout the text. Here, we define the following notations used in the proofs.

$f(w_i) \equiv$ density of W at w_i,

$R_i(t) \equiv I(X_i \geq t)$,

$R(t) = \sum_i R_i(t)$,

$T, t_j \equiv$ overall survival time,

$C, c_j \equiv$ censoring time,

$X, x_j \equiv \min(T, C)$ and $\min(t_j, c_j)$, respectively

U_n: $U-$ statistics,

$S(u) = Pr(T > u)$,

$B(h, u) = \frac{E\{hI(T \geq u)\}}{S(u-)}$,

$\mathbf{v}_i = (w_i, t_i, \delta_i)^{\mathrm{T}}$.

We first list several equalities that are used during the derivation:

$$\begin{aligned} R(t) &= n\widehat{G}(t-)\widehat{S}(t-) \\ \frac{\widehat{G}(t) - G(t)}{G(t)} &= -\int_0^t \frac{\widehat{G}(u-)}{G(u)} \frac{dM^c(u)}{R(u)} \\ \frac{\delta_i}{G(x_i)} &= 1 - \int \frac{dM_i^c(u)}{G(u)} \end{aligned} \tag{A.1}$$

These equations are given on page 37 in Gill (1980), page 313 in Robins & Rotnitzky (1992) and in Ma & Yin (2010), hence we do not give the detailed derivations here.

A.3 Lemmas

Lemma 1 *Letting*

$$\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i) = \frac{\sum_{j=1}^n \delta_j \mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_j,t_j) I(x_j > x_i) K_h(w_j - w_i)/\widehat{G}(t_j)}{\sum_{j=1}^n \delta_j I(x_j > x_i) K_h(w_j - w_i)/\widehat{G}(t_j)}$$

and

$$\begin{aligned}\mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i) &= E\left\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_i,T_i) \mid T_i > X_i, W_i = w_i, X_i\right\} \\ &= \frac{E\left\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_i,T_i) I(T_i > x_i) \mid W_i = w_i, x_i\right\}}{E\left\{I(T_i > x_i) \mid W_i = w_i, x_i\right\}},\end{aligned}$$

we have

$$\begin{aligned}&\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i) - \mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i) \\ =& \frac{f^{-1}(w_i)\frac{1}{n}\sum_{j=1}^n \delta_j \mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_j,t_j) I(x_j > x_i) K_h(w_j - w_i)/\widehat{G}(t_j)}{f^{-1}(w_i)\frac{1}{n}\sum_{j=1}^n \delta_j I(x_j > x_i) K_h(w_j - w_i)/\widehat{G}(t_j)} \\ &-\frac{E\left\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_i,T_i) I(T_i > x_i) \mid w_i, x_i\right\}}{E\left\{I(T_i > x_i) \mid w_i, x_i\right\}} \\ =& [f^{-1}(w_i)\frac{1}{n}\sum_{j=1}^n \delta_j \mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_j,t_j) I(x_j > x_i) K_h(w_j - w_i)/\widehat{G}(t_j) \\ &-E\left\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_i,T_i) I(T_i > x_i) \mid w_i, x_i\right\}] E\left\{I(T_i > x_i) \mid w_i, x_i\right\}^{-1} \\ &-\mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i)[f^{-1}(w_i)\frac{1}{n}\sum_{j=1}^n \delta_j I(x_j > x_i) K_h(w_j - w_i)/\widehat{G}(t_j) \\ &-E\left\{I(T_i > x_i) \mid w_i, x_i\right\}] E\left\{I(T_i > x_i) \mid w_i, x_i\right\}^{-1} + o_p(n^{-1/2}).\end{aligned}$$

Proof: Letting

$$\begin{aligned}\widehat{A} &= f^{-1}(w_i)\frac{1}{n}\sum_{j=1}^n \delta_j \mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_j,t_j) I(x_j > x_i) K_h(w_j - w_i)/\widehat{G}(t_j), \\ \widehat{B} &= f^{-1}(w_i)\frac{1}{n}\sum_{j=1}^n \delta_j I(x_j > x_i) K_h(w_j - w_i)/\widehat{G}(t_j), \\ A &= E\left\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_i,T_i) I(T_i > x_i) \mid w_i, x_i\right\}, \\ B &= E\left\{I(T_i > x_i) \mid w_i, x_i\right\},\end{aligned}$$

then by Taylor expansion,

$$\frac{\widehat{A}}{\widehat{B}} - \frac{A}{B} = \frac{1}{B}(\widehat{A} - A) - \frac{A}{B^2}(\widehat{B} - B) + A^*(\widehat{B} - B)^2/(B^{*3}) - (\widehat{A} - A)(\widehat{B} - B)/(B^{*2}),$$

where $(A^{*\mathrm{T}}, B^*)^{\mathrm{T}}$ is a point on the line connecting $(\widehat{A}^{\mathrm{T}}, \widehat{B})^{\mathrm{T}}$ and $(A^{\mathrm{T}}, B)^{\mathrm{T}}$. Note that $\widehat{A}$ and $\widehat{B}$ are the kernel regression estimators of A and B respectively, hence $\widehat{A}-A$ and $\widehat{B}-B$ are both of order $O_p\{h^2+(nh)^{-1/2}\}$. Thus, the last two terms of the above display are of order $O_p\{h^4+(nh)^{-1}\} = o_p(n^{-1/2})$ under the assumption that $nh^4 \to 0$ and $nh^2 \to \infty$. This proves the results.

<u>Lemma</u> 2

$$n^{-1/2}\sum_{i=1}^{n}(1-\delta_i)(\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i)-\mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i)) = n^{-1/2}\sum_{i=1}^{n}\rho_i(\boldsymbol{\theta})+o_p(1),$$

where

$$\begin{aligned}
\rho_i(\boldsymbol{\theta}) &= \frac{\delta_i \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,x_i)}{G(t_i)}\{1-G(t_i)\} \\
&- \frac{\delta_i}{G(t_i)} E\{I(t_i>C_j)\mathbf{Q}_{\boldsymbol{\theta},j}(w_i,C_j) \mid \mathbf{v}_i\} \\
&+ \int \frac{B[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_j,T_j)\{1-G(T_j)\},u]}{G(u)} dM_i^c(u) \\
&- \int \frac{B\{\mathbf{Q}_{\boldsymbol{\theta},j}(W_j,C_j)I(T_j>C_j),u\}}{G(u)} dM_i^c(u).
\end{aligned}$$

Proof: We first derive the asymptotic expansion of $\widehat{A}$ and $\widehat{B}$. Since $\widehat{A}$ and $\widehat{B}$ have a common form, in the following, we first derive the general asymptotic expansion

$$\sum_{j=1}^{n}\frac{\delta_j K_h(w_j-w_i)\widetilde{f}(w_j,x_j,x_i)}{\widehat{G}(t_j)}$$

for an arbitrary $\widetilde{f}(w_j,x_j,x_i)$ function.

$$\begin{aligned}
&\sum_{j=1}^{n}\frac{\delta_j K_h(w_j-w_i)}{\widehat{G}(t_j)}\widetilde{f}(w_j,x_j,x_i) \\
=& \sum_{j=1}^{n}\frac{\delta_j K_h(w_j-w_i)\widetilde{f}(w_j,x_j,x_i)}{G(t_j)} + \sum_{j=1}^{n}\frac{\delta_j K_h(w_j-w_i)\widetilde{f}(w_j,x_j,x_i)}{\widehat{G}(t_j)}\left\{1-\frac{\widehat{G}(t_j)}{G(t_j)}\right\} \\
=& \sum_{j=1}^{n}\frac{\delta_j K_h(w_j-w_i)\widetilde{f}(w_j,x_j,x_i)}{G(t_j)} + \sum_{j=1}^{n}\frac{\delta_j K_h(w_j-w_i)\widetilde{f}(w_j,x_j,x_i)}{\widehat{G}(t_j)}\int_0^{t_j}\frac{\widehat{G}(u-)dM^c(u)}{G(u)R(u)} \\
=& \sum_{j=1}^{n}\frac{\delta_j K_h(w_j-w_i)\widetilde{f}(w_j,x_j,x_i)}{G(t_j)} + \frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_h(w_j-w_i)\widetilde{f}(w_j,x_j,x_i)}{\widehat{G}(t_j)}
\end{aligned}$$

$$
\begin{aligned}
&\times \int_0^{t_j} \frac{n\widehat{S}(u-)\widehat{G}(u-)dM^c(u)}{G(u)\widehat{S}(u-)R(u)} \\
= &\sum_{j=1}^n \frac{\delta_j K_h(w_j - w_i)\widetilde{f}(w_j, x_j, x_i)}{G(t_j)} + \frac{1}{n}\sum_{j=1}^n \frac{\delta_j K_h(w_j - w_i)\widetilde{f}(w_j, x_j, x_i)}{\widehat{G}(t_j)} \\
&\times \int_0^{t_j} \frac{R(u)dM^c(u)}{G(u)\widehat{S}(u-)R(u)} \\
= &\sum_{j=1}^n \frac{\delta_j K_h(w_j - w_i)\widetilde{f}(w_j, x_j, x_i)}{G(t_j)} + \frac{1}{n}\sum_{j=1}^n \frac{\delta_j K_h(w_j - w_i)\widetilde{f}(w_j, x_j, x_i)}{\widehat{G}(t_j)} \int_0^\infty \frac{R_j(u)dM^c(u)}{G(u)\widehat{S}(u-)} \\
= &\sum_{j=1}^n \frac{\delta_j K_h(w_j - w_i)\widetilde{f}(w_j, x_j, x_i)}{G(t_j)} + \frac{1}{n}\int_0^\infty \sum_{j=1}^n \frac{\delta_j K_h(w_j - w_i)\widetilde{f}(w_j, x_j, x_i)R_j(u)dM^c(u)}{\widehat{G}(t_j)G(u)\widehat{S}(u-)} \\
= &\sum_{j=1}^n \frac{\delta_j K_h(w_j - w_i)\widetilde{f}(w_j, x_j, x_i)}{G(t_j)} + \frac{nf(w_i)}{n}\int_0^\infty \sum_{j=1}^n \frac{\delta_j K_h(w_j - w_i)\widetilde{f}(w_j, x_j, x_i)R_j(u)}{nf(w_i)\widehat{G}(t_j)G(u)\widehat{S}(u-)} \\
&\times dM^c(u) \\
= &\sum_{j=1}^n \frac{\delta_j K_h(w_j - w_i)\widetilde{f}(w_j, x_j, x_i)}{G(t_j)} + f(w_i)\int_0^\infty \frac{E\left\{\widetilde{f}(w_i, T_i, x_i)I(T_i \geq u) \mid w_i, x_i\right\}}{G(u)S(u-)} dM^c(u) \\
&+ f(w_i)\sum_{i=1}^n \int_0^\infty \psi_n(u)dM_i^c(u) \\
= &\sum_{j=1}^n \frac{\delta_j K_h(w_j - w_i)\widetilde{f}(w_j, x_j, x_i)}{G(t_j)} + f(w_i)\int_0^\infty \frac{E\left\{\widetilde{f}(w_i, T_i, x_i)I(T_i \geq u) \mid w_i, x_i\right\}}{G(u)S(u-)} dM^c(u) \\
&+ o_p(\sqrt{n}),
\end{aligned}
$$

where

$$
\psi_n(t) = -\frac{E\left\{\widetilde{f}(w_i, T_i, x_i)I(T_i \geq u) \mid w_i, x_i\right\}}{G(u)S^{*2}(u-)}\{\widehat{S}(u-) - S(u-)\} + o_p(1),
$$

and S^* is a point on the line connecting $\widehat{S}(u-)$ and $S(u-)$. Note that the residual term in the above equation is $o_p(1)$ because the kernel estimator and the estimator $\widehat{G}(t_j)$ are consistent. Further $\{\widehat{S}(u-) - S(u-)\} = o_p(1)$ and $\frac{E\left\{\widetilde{f}(w_i, T_i)I(T_i \geq u) | w_i, x_i\right\}}{G(u)S^{*2}(u-)} = O(1)$, and thus $\psi_n(t) = o_p(1)$. Also, $\widehat{S}(u-) - S(u-)$ is $\mathcal{F}_t$-adapted and the residual term does not depend on the u in the integrand. Hence, $\psi_n(t)$ are predictable processes. Thus, the martingale central limit theorem gives us the results that $\sum_{i=1}^n \int_0^\infty \psi_n(u)dM_i^c(u)$ is of $o_p(n^{1/2})$.

Letting $\widetilde{f}(w_j, x_j, x_i) = \mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_j, t_j)I(x_j > x_i)$ in $\widehat{A}$ and $= I(x_j > x_i)$ in $\widehat{B}$, we have

$$
\widehat{A} - A
$$

$$
\begin{aligned}
&= f^{-1}(w_i)\frac{1}{n}\sum_{j=1}^{n}\delta_j\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_j,x_j)I(x_j>x_i)K_h(w_j-w_i)/G(t_j)\\
&\quad -E\left\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_i,T_i)I(T_i>x_i)\mid w_i,x_i\right\}\\
&\quad +\frac{1}{n}\int\frac{E\left\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_i,T_i)I(T_i>x_i)I(T_i\geq u)\mid w_i,x_i\right\}}{G(u)S(u-)}dM^c(u)+o_p(n^{-1/2}),
\end{aligned}
$$

and

$$
\begin{aligned}
&\widehat{B}-B\\
&= f^{-1}(w_i)\frac{1}{n}\sum_{j=1}^{n}\delta_jI(x_j>x_i)K_h(w_j-w_i)/G(t_j)-E\left\{I(T_i>x_i)\mid w_i,x_i\right\}\\
&\quad +\frac{1}{n}\int\frac{E\left\{I(T_i>x_i)I(T_i\geq u)\mid w_i,x_i\right\}}{G(u)S(u-)}dM^c(u)+o_p(n^{-1/2}).
\end{aligned}
$$

Plugging in the two equations, we have

$$
\begin{aligned}
&\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i)-\mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i)\\
&= \frac{f^{-1}(w_i)\frac{1}{n}\sum_{j=1}^{n}\delta_j\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_j,t_j)I(x_j>x_i)K_h(w_j-w_i)/G(t_j)}{E\left\{I(T_i>x_i)\mid w_i,x_i\right\}} && \text{(A.2)}\\
&\quad -\frac{f^{-1}(w_i)\frac{1}{n}\sum_{j=1}^{n}\delta_j\mathbf{Q}_{\boldsymbol{\theta},i}(x_i)I(x_j>x_i)K_h(w_j-w_i)/G(t_j)}{E\left\{I(T_i>x_i)\mid w_i,x_i\right\}} && \text{(A.3)}\\
&\quad +\frac{1}{nE\left\{I(t_i>x_i)\mid w_i,x_i\right\}}\int\frac{E\left\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_i,T_i)I(T_i>x_i)I(T_i\geq u)\mid w_i,x_i\right\}}{G(u)S(u-)}dM^c(u) && \text{(A.4)}\\
&\quad -\frac{\mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i)}{nE\left\{I(T_i>x_i)\mid w_i,x_i\right\}}\int\frac{E\left\{I(T_i>x_i)I(T_i\geq u)\mid w_i,x_i\right\}}{G(u)S(u-)}dM^c(u) && \text{(A.5)}\\
&\quad +o_p(n^{-1/2}).
\end{aligned}
$$

We then have to obtain the asymptotic properties for

$$
n^{-1/2}\sum_{i=1}^{n}(1-\delta_i)\left\{\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i)-\mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i)\right\}.
$$

We conduct separate computations for the assumptions (A.2) to (A.5).

For (A.2): We let

$$
\Pi_j = \frac{\delta_j\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_j,x_j)I(x_j>x_i)}{G(t_j)},
$$

$$\begin{aligned}
\mathbf{v}_i &= (\delta_i, x_i, w_i), \\
\mathbf{V}_i &= (\Delta_i, X_i, W_i), \\
\widehat{g}(v_i) &= \frac{1}{n}\sum_{j=1}^{n} \Pi_j K_h(w_j - w_i), \\
r(w_i) &= E(\Pi_i \mid W = w_i), \\
g(w_i) &= r(w_i) f(w_i), \\
H(\mathbf{v}_i) &= \frac{f^{-1}(w_i)(1-\delta_i)}{E\left\{I(T_i > x_i) \mid w_i, x_i\right\}}.
\end{aligned}$$

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} H(\mathbf{v}_i)\widehat{g}(\mathbf{v}_i) \\
= & \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{f^{-1}(w_i)(1-\delta_i)}{E\left\{I(T_i > x_i) \mid w_i, x_i\right\}}\widehat{g}(\mathbf{v}_i) \\
= & \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{f^{-1}(w_i)(1-\delta_i)}{E\left\{I(T_i > x_i) \mid w_i, x_i\right\}}\frac{1}{n}\sum_{j=1}^{n} \Pi_j K_h(w_j - w_i) \\
= & (n-1)/n\sqrt{n}\frac{1}{\binom{n}{2}}\sum_{i<j}\left\{\frac{f^{-1}(w_i)(1-\delta_i)\delta_j \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_j, x_j) I(x_j > x_i) K_h(w_j - w_i)}{2E\left\{I(T_i > x_i) \mid w_i, x_i\right\} G(t_j)}\right. \\
& \left. + \frac{f^{-1}(w_j)(1-\delta_j)\delta_i \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i, x_i) I(x_i > x_j) K_h(w_i - w_j)}{2E\left\{I(T_j > x_j) \mid w_j, x_j\right\} G(t_i)}\right\}.
\end{aligned}$$

We note that the remaining terms in the above equation are equal to 0 since $\delta_i(1-\delta_i) = 0$.

Letting

$$\begin{aligned}
u_{h1}(\mathbf{v}_i, \mathbf{v}_j) &= \frac{f^{-1}(w_i)(1-\delta_i)\delta_j \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_j, x_j) I(x_j > x_i) K_h(w_j - w_i)}{E\left\{I(T_i > x_i) \mid w_i, x_i\right\} G(t_j)} \\
u_{h2}(\mathbf{v}_i, \mathbf{v}_j) &= u_{h1}(\mathbf{v}_j, \mathbf{v}_i),
\end{aligned}$$

then $u_h(\mathbf{v}_i, \mathbf{v}_j) = \left\{u_{h1}(\mathbf{v}_i, \mathbf{v}_j) + u_{h2}(\mathbf{v}_i, \mathbf{v}_j)\right\}/2$ is the kernel of the $U-$statistic,

$$U_n = \frac{1}{\binom{n}{2}}\sum_{i<j} u_h(\mathbf{v}_i, \mathbf{v}_j).$$

We then compute

$$E\left\{u_h(\mathbf{V}_i, \mathbf{v}_j) \mid \mathbf{v}_j\right\} = [E\left\{u_{h1}(\mathbf{V}_i, \mathbf{v}_j) \mid \mathbf{v}_j\right\} + E\left\{u_{h2}(\mathbf{V}_i, \mathbf{v}_j) \mid \mathbf{v}_j\right\}]/2.$$

$$E\left\{u_{h1}(\mathbf{V}_i, \mathbf{v}_j) \mid \mathbf{v}_j\right\}$$

$$
\begin{aligned}
&= \frac{\delta_j \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_j, x_j)}{G(t_j)} E\left[\frac{f^{-1}(W_i)(1-\Delta_i)I(x_j > X_i)K_h(w_j - W_i)}{E\{I(T_i > X_i) \mid W_i, X_i\}} \mid \mathbf{v}_j\right] \\
&= \frac{\delta_j \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_j, x_j)}{G(t_j)} E\left[f^{-1}(W_i)K_h(w_j - W_i)E\{(1-\Delta_i)I(x_j > X_i) \mid T_i > X_i, W_i, X_i, \mathbf{v}_j\} \mid \mathbf{v}_j\right] \\
&= \frac{\delta_j \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_j, x_j)}{G(t_j)} E\left[f^{-1}(W_i)K_h(w_j - W_i)I(x_j > C_i) \mid \mathbf{v}_j\right] \\
&= \frac{\delta_j \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_j, x_j)}{G(t_j)} E\left[f^{-1}(W_i)K_h(w_j - W_i)E\{I(x_j > C_i) \mid W_i, \mathbf{v}_j\} \mid \mathbf{v}_j\right] \\
&= \frac{\delta_j \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_j, x_j)}{G(t_j)} E\left[f^{-1}(W_i)K_h(w_j - W_i)\{1 - G(x_j)\} \mid \mathbf{v}_j\right] \\
&= \frac{\delta_j \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_j, x_j)}{G(t_j)} \{1 - G(t_j)\} + O_p(h^2) \\
&= \widetilde{u}_{h1}(\mathbf{v}_j) + O_p(h^2).
\end{aligned}
$$

We note that in the above derivation, we assume the censoring time distribution is continuous.

In addition,

$$
\begin{aligned}
& E\{u_{h2}(\mathbf{V}_i, \mathbf{v}_j) \mid \mathbf{v}_j\} \\
&= \frac{f^{-1}(w_j)(1-\delta_j)}{E\{I(T_j > x_j) \mid w_j, x_j\}} E\left\{\frac{\Delta_i \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i, T_i)}{G(T_i)} I(X_i > x_j)K_h(w_j - W_i) \mid \mathbf{v}_j\right\} \\
&= \frac{f^{-1}(w_j)(1-\delta_j)}{E\{I(T_j > x_j) \mid w_j, x_j\}} E\left[K_h(w_j - W_i)E\left\{\frac{\Delta_i \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i, T_i)I(X_i > x_j)}{G(T_i)} \mid w_i, \mathbf{v}_j\right\} \mid \mathbf{v}_j\right] \\
&= \frac{f^{-1}(w_j)(1-\delta_j)}{E\{I(T_j > x_j) \mid w_j, x_j\}} E\left[K_h(w_j - W_i)E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i, T_i)I(T_i > x_j) \mid w_i, \mathbf{v}_j\} \mid \mathbf{v}_j\right] \\
&= \frac{f^{-1}(w_j)(1-\delta_j)}{E\{I(T_j > x_j) \mid w_j, x_j\}} E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_j, T_i)I(T_i > x_j) \mid W_i = w_j, \mathbf{v}_j\} f(w_j) + O_p(h^2) \\
&= \frac{(1-\delta_j)}{E\{I(T_j > x_j) \mid w_j, x_j\}} E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_j, T_i)I(T_i > x_j) \mid \mathbf{v}_j\} + O_p(h^2) \\
&= \widetilde{u}_{h2}(\mathbf{v}_j) + O_p(h^2).
\end{aligned}
$$

So, we have

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} H(\mathbf{v}_i)\widehat{g}(\mathbf{v}_i) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \{\widetilde{u}_{h1}(\mathbf{v}_j) + \widetilde{u}_{h2}(\mathbf{v}_j)\} - \sqrt{n} E\{u_{h1}(\mathbf{V}_i, \mathbf{V}_j)\} + o_p(1),
$$

where

$$
E\{u_{h1}(\mathbf{V}_i, \mathbf{V}_j)\} = E\left[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_j, T_j)I(T_j > C_i)\right] = E\left[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_j, T_j)\{1 - G(T_j)\}\right].
$$

For (A.3): We let

$$\begin{aligned}\Pi_j &= \frac{\delta_j I(x_j > x_i)}{G(t_j)} \\ H(\mathbf{v}_i) &= \frac{f^{-1}(w_i)\mathbf{Q}_{\boldsymbol{\theta},i}(x_i)(1-\delta_i)}{E\left\{I(T_i > x_i) \mid w_i, x_i\right\}}\end{aligned}$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} H(\mathbf{v}_i)\widehat{g}(\mathbf{v}_i) = \sqrt{n}\frac{1}{\binom{n}{2}}\sum_{i<j} 1/2\left\{u_{h1}(\mathbf{v}_i,\mathbf{v}_j) + u_{h2}(\mathbf{v}_i,\mathbf{v}_j)\right\},$$

where

$$u_{h1}(\mathbf{v}_i,\mathbf{v}_j) = u_{h2}(\mathbf{v}_j,\mathbf{v}_i) = \frac{f^{-1}(w_i)(1-\delta_i)\mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i)\delta_j I(x_j > x_i)}{E\left\{I(T_i > x_i) \mid w_i, x_i\right\} G(t_j)} K_h(w_j - w_i).$$

$$\begin{aligned}
& E\left\{u_{h1}(\mathbf{V}_i,\mathbf{v}_j) \mid \mathbf{v}_j\right\} \\
=\ & \frac{\delta_j}{G(t_j)} E\left[E\left\{I(x_j > C_i)\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,C_i) \mid W_i = w_j, x_j\right\} \mid \mathbf{v}_j\right] + O_p(h^2) \\
=\ & \frac{\delta_j}{G(t_j)} E\left[E\left\{I(t_j > C_i)\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,C_i) \mid W_i = w_j, t_j\right\} \mid \mathbf{v}_j\right] + O_p(h^2) \\
=\ & \frac{\delta_j}{G(t_j)} E\left\{I(t_j > C_i)\mathbf{Q}_{\boldsymbol{\theta},i}(W_i = w_j, C_i) \mid \mathbf{v}_j\right\} + O_p(h^2) \\
=\ & \frac{\delta_j}{G(t_j)} E\left\{I(t_j > C_i)\mathbf{Q}_{\boldsymbol{\theta},i}(w_j, C_i) \mid \mathbf{v}_j\right\} + O_p(h^2) \\
=\ & \widetilde{u}_{h1}(\mathbf{v}_j) + O_p(h^2),
\end{aligned}$$

and

$$\begin{aligned}
& E\left\{u_{h2}(\mathbf{V}_i,\mathbf{v}_j) \mid \mathbf{v}_j\right\} \\
=\ & \frac{f^{-1}(w_j)(1-\delta_j)\mathbf{Q}_{\boldsymbol{\theta},j}(w_j,x_j)}{E\left\{I(T_j > x_j) \mid w_j, x_j\right\}} E\left\{\frac{\Delta_i I(X_i > x_j) K_h(W_i - w_j)}{G(T_i)} \mid \mathbf{v}_j\right\} \\
=\ & \frac{f^{-1}(w_j)(1-\delta_j)\mathbf{Q}_{\boldsymbol{\theta},j}(w_j,x_j)}{E\left\{I(T_j > x_j) \mid w_j, x_j\right\}} E\left\{I(T_i > x_j) K_h(W_i - w_j) \mid \mathbf{v}_j\right\} \\
=\ & \frac{f^{-1}(w_j)(1-\delta_j)\mathbf{Q}_{\boldsymbol{\theta},j}(w_j,x_j)}{E\left\{I(T_j > x_j) \mid w_j, x_j\right\}} E\left\{I(T_i > x_j) \mid W_i = w_j, \mathbf{v}_j\right\} f(w_j) + O_p(h^2) \\
=\ & \frac{(1-\delta_j)\mathbf{Q}_{\boldsymbol{\theta},j}(w_j,x_j)}{E\left\{I(T_j > x_j) \mid w_j, x_j\right\}} E\left\{I(T_i > x_j) \mid W_i = w_j, \mathbf{v}_j\right\} + O_p(h^2) \\
=\ & \widetilde{u}_{h2}(\mathbf{v}_j) + O_p(h^2).
\end{aligned}$$

Further, we have

$$
\begin{aligned}
& E\left\{u_{h1}(\mathbf{V}_i,\mathbf{V}_j)\right\} \\
= \; & E\left(\frac{\delta_j}{G(T_j)}E\left[E\left\{I(T_j>C_i)\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,C_i)\mid W_i=W_j,C_i,T_j\right\}\mid \mathbf{V}_j\right]\right)+O_p(h^2) \\
= \; & E\left(E\left[E\left\{I(T_j>C_i)\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,C_i)\mid W_i=W_j,C_i,T_j\right\}\mid \mathbf{V}_j\right]\right)+O_p(h^2) \\
= \; & E\left\{I(T_j>C_i)\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,C_i)\right\}+O_p(h^2).
\end{aligned}
$$

The last equation holds because W_i are $i.i.d.$ Therefore, the same as before, we have

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^n H(\mathbf{v}_i)\widehat{g}(\mathbf{v}_i)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\{\widetilde{u}_{h1}(\mathbf{v}_j)+\widetilde{u}_{h2}(\mathbf{v}_j)\}-\sqrt{n}E\left\{u_{h1}(\mathbf{V}_i,\mathbf{V}_j)\right\}+o_p(1).
$$

For (A.4):

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^n(1-\delta_i)\frac{1}{E\left\{I(T_i>x_i)\mid w_i,x_i\right\}n}\int\frac{E\left\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,T_i)I(T_i>x_i)I(T_i\ge u)\mid w_i,x_i\right\}}{G(u)S(u-)} \\
& \times dM^c(u) \\
= \; & n^{-1/2}\int\frac{E\left[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,T_i)\left\{1-G(T_i)\right\}I(T_i\ge u)\right]}{G(u)S(u-)}dM^c(u)+o_p(1) \\
= \; & n^{-1/2}\sum_{j=1}^n\int\frac{B\left[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,T_i)\left\{1-G(T_i)\right\},u\right]}{G(u)}dM_j^c(u)+o_p(1).
\end{aligned}
$$

For (A.5):

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^n(1-\delta_i)\frac{\mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i)}{E\left\{I(T_i>x_i)\mid w_i,x_i\right\}n}\int\frac{E\left\{I(T_i>x_i)I(T_i\ge u)\mid w_i,x_i\right\}}{G(u)S(u-)}dM^c(u) \\
& +o_p(1) \\
= \; & n^{-1/2}\sum_{j=1}^n\int\frac{B\left\{\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,C_i)I(T_i>C_i),u\right\}}{G(u)}dM_j^c(u)+o_p(1),
\end{aligned}
$$

where $B(h,u)=E\{hI(T\ge u)\}/S(u-)$.

By combining the results from the above derivations, we have

$$
\begin{aligned}
\rho_i(\boldsymbol{\theta}) = \; & \frac{\delta_i\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,x_i)}{G(t_i)}\left\{1-G(t_i)\right\}+\frac{(1-\delta_i)E\left\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,T_j)I(T_j>x_i)\mid W_j=w_i,\mathbf{v}_i\right\}}{E\left\{I(T_i>x_i)\mid w_i,x_i\right\}} \\
- \; & E\left[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,T_i)I(T_i>C_j)\right] \\
- \; & \frac{\delta_i}{G(t_i)}E\left\{I(t_i>C_j)\mathbf{Q}_{\boldsymbol{\theta},j}(w_i,C_j)\mid \mathbf{v}_i\right\}
\end{aligned}
$$

$$
\begin{aligned}
- & \frac{(1-\delta_i)\mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i)}{E\{I(T_i > x_i) \mid w_i, x_i\}} E\{I(T_j > x_i) \mid W_j = W_i, \mathbf{v}_i\} \\
& + E\{I(T_i > C_j)\mathbf{Q}_{\boldsymbol{\theta},j}(W_j, C_j)\} \\
+ & \int \frac{B[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_j,T_j)\{1-G(T_j)\}, u]}{G(u)} dM_i^c(u) \\
- & \int \frac{B\{\mathbf{Q}_{\boldsymbol{\theta},j}(W_j,C_j)I(T_j > C_j), u\}}{G(u)} dM_i^c(u).
\end{aligned}
$$

We can further simplify ρ_i. We show that

$$
E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,T_j)I(T_j > x_i) \mid W_j = W_i, \mathbf{v}_i\} = \mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i)E\{I(T_j > x_i) \mid W_j = W_i, \mathbf{v}_i\} \tag{A.6}
$$

because

$$
\begin{aligned}
& \mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i)E\{(T_j > x_i) \mid W_j = W_i, \mathbf{v}_i\} \\
= & E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}|T_i > x_i, w_i, x_i\}E\{I(T_j > x_i)|W_j = W_i, \mathbf{v}_i\} \\
= & \frac{E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,T_i)I(T_i > x_i)|w_i,x_i\}}{E\{I(T_i > x_i)|w_i,x_i\}} E\{I(T_j > x_i)|W_j = W_i, \mathbf{v}_i\} \\
& (\text{since } (T_i, W_i),\ (T_j, W_j) \text{ are } i.i.d) \\
= & E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,T_i)I(T_i > x_i)|w_i,x_i\} \\
= & E\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,T_j)I(T_j > x_i)|W_j = W_i, \mathbf{v}_i\}.
\end{aligned}
$$

Further, by taking the expectation on both sides of the equation in (A.6), we have

$$
E[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,T_i)I(T_i > C_j)] = E\{I(T_i > C_j)\mathbf{Q}_{\boldsymbol{\theta},j}(W_j,C_j)\}.
$$

As a result, we can write $\rho_i(\boldsymbol{\theta})$ as follows because the terms leading with $1-\delta_i$ and the above two expectations are cancelled in the original form.

$$
\begin{aligned}
\rho_i(\boldsymbol{\theta}) = & \frac{\delta_i \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,x_i)}{G(t_i)}\{1-G(t_i)\} \\
- & \frac{\delta_i}{G(t_i)} E\{I(t_i > C_j)\mathbf{Q}_{\boldsymbol{\theta},j}(w_i,C_j) \mid \mathbf{v}_i\} \\
+ & \int \frac{B[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_j,T_j)\{1-G(T_j)\}, u]}{G(u)} dM_i^c(u) \\
- & \int \frac{B\{\mathbf{Q}_{\boldsymbol{\theta},j}(W_j,C_j)I(T_j > C_j), u\}}{G(u)} dM_i^c(u).
\end{aligned}
$$

This proves the results.

A.4 Proofs of theorems

Theorem 1 *Let* $\widehat{U}_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^n\{\delta_i \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,t_i) + (1-\delta_i)\widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i)\}$. *Under assumptions A1-A8,*

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_p(1),$$

where $\widehat{\boldsymbol{\theta}}$ *solves* $\widehat{U}_n(\boldsymbol{\theta}) = 0$ *and* $\boldsymbol{\theta}_0$ *is the true parameter value.*

Proof: Letting

$$\begin{aligned}
U_0(\boldsymbol{\theta}) &= E[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,T_i) - E\{\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,C_i)I(C_i<T_i)|W_i,X_i\}] + E\{(1-\Delta_i)\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,X_i)\},\\
\widehat{U}_n(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^n \delta_i \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,t_i) + (1-\delta_i)\mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i),
\end{aligned}$$

we show that

$$\sup_{\boldsymbol{\theta}\in\Theta}|\widehat{U}_n^2(\boldsymbol{\theta}) - U_0^2(\boldsymbol{\theta})| \xrightarrow{p} 0.$$

Since

$$|\widehat{U}_n^2(\boldsymbol{\theta}) - U_0^2(\boldsymbol{\theta})| \le |\widehat{U}_n(\boldsymbol{\theta}) + U_0(\boldsymbol{\theta})||\widehat{U}_n(\boldsymbol{\theta}) - U_0(\boldsymbol{\theta})|,$$

and

$$\sup_{\boldsymbol{\theta}\in\Theta}|\widehat{U}_n(\boldsymbol{\theta}) + U_0(\boldsymbol{\theta})| < \infty,$$

in probability, it is sufficient to show that

$$\sup_{\boldsymbol{\theta}\in\Theta}|\widehat{U}_n(\boldsymbol{\theta}) - U_0(\boldsymbol{\theta})| \xrightarrow{p} 0.$$

Since

$$\begin{aligned}
& \widehat{U}_n(\boldsymbol{\theta}) - U_0(\boldsymbol{\theta})\\
=\ & \widehat{U}_n(\boldsymbol{\theta}) - U_n(\boldsymbol{\theta}) + U_n(\boldsymbol{\theta}) - U_0(\boldsymbol{\theta})\\
=\ & \frac{1}{n}\sum_{i=1}^n \rho_i(\boldsymbol{\theta})\\
+\ & \frac{1}{n}\sum \delta_i \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,t_i) - E[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,T_i) - E\{\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,C_i)I(C_i<T_i)|W_i,X_i\}]\\
+\ & \frac{1}{n}\sum_{i=1}^n (1-\delta_i)\mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i) - E\{(1-\Delta_i)\mathbf{Q}_{\boldsymbol{\theta},i}(w_i,x_i)\},
\end{aligned}$$

$$= \quad L_1 + L_2 + L_3,$$

we show that

$$\sup_{\boldsymbol{\theta}\in\Theta} |L_i| \xrightarrow{p} 0 \text{ for } i = 1, 2, 3.$$

Clearly,

$$\sup_{\boldsymbol{\theta}\in\Theta} |L_3| \xrightarrow{p} 0$$

by the law of large numbers.

For L_2 : Since

$$\begin{aligned}
& E[E\{\mathbf{Q}_{\boldsymbol{\theta},i}(W_i, C_i) I(C_i < T_i) | W_i, X_i\}] \\
= \;& E[E\{E(\mathbf{S}_{i,\text{eff}}(W_i, T_i) | T_i > C_i, W_i, X_i) I(C_i < T_i) | W_i, X_i\}] \\
= \;& E[E\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i, T_i)(1 - \Delta_i) | W_i, X_i\}] \\
= \;& E\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i, T_i)(1 - \Delta_i)\},
\end{aligned}$$

therefore

$$E[\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i, T_i) - E\{\mathbf{Q}_{\boldsymbol{\theta},i}(W_i, C_i) I(C_i < T_i) | W_i, X_i\}] = E\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i, T_i)\Delta_i\},$$

which implies that

$$\sup_{\boldsymbol{\theta}\in\Theta} |L_2| \xrightarrow{p} 0$$

by the law of large numbers.

For L_1 :

$$\begin{aligned}
L_1 \;=\; & \frac{1}{n}\sum_{j=1}^{n} \rho_j(\boldsymbol{\theta}) \\
=\; & \frac{1}{n}\left[\sum_{j=1}^{n} \frac{\delta_j \mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_j, x_j)}{G(t_j)} \{1 - G(t_j)\} - \sum_{j=1}^{n} \frac{\delta_j}{G(t_j)} E\left\{I(t_j > C_i)\mathbf{Q}_{\boldsymbol{\theta},i}(w_j, C_i) \mid \mathbf{v}_j\right\}\right] \\
+\; & \left[\frac{1}{n}\sum_{j=1}^{n}\int_0^{\tau} \frac{B\left[\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i, T_i)\{1 - G(T_i)\}, u\right]}{G(u)} dM_j^c(u)\right] \\
-\; & \left[\frac{1}{n}\sum_{j=1}^{n}\int_0^{\tau} \frac{B\left\{\mathbf{Q}_{\boldsymbol{\theta},i}(W_i, C_i) I(T_i > C_i), u\right\}}{G(u)} dM_j^c(u)\right] \\
=\; & e_1 + e_2 - e_3 + o_p(1).
\end{aligned}$$

For e_1 : We have

$$\begin{aligned} & E\left[\frac{\Delta_i \mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_j,T_j)}{G(T_j)}\{1-G(T_j)\}\right] \\ = \; & E[\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_j,T_j)\{1-G(T_j)\}], \end{aligned}$$

and

$$\begin{aligned} & E\left[\frac{\Delta_j}{G(T_j)} E\{\mathbf{Q}_{\boldsymbol{\theta},j}(W_j,C_i)I(C_i<T_j)|W_j,T_j\}\right] \\ = \; & E\{I(T_j>C_i)\mathbf{Q}_{\boldsymbol{\theta},j}(W_j,C_i)\} \\ = \; & E[I(T_j>C_i)E\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i,T_i)|T_i>C_i,W_i=W_j,C_i\}] \\ = \; & E[E\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i,T_i)|T_i>C_i,W_i=W_j,C_i\}E\{I(T_j>C_i)|W_j,C_i\}] \\ = \; & E[E\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i,T_i)I(T_i>C_i)|W_i=W_j,C_i\}] \\ = \; & E[\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i,T_i)I(T_i>C_i)] \\ = \; & E[\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i,T_i)E\{I(C_i<T_i)|T_i,W_i\}] \\ = \; & E\{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i,T_i)(1-G(T_i))\}. \end{aligned}$$

Because the two terms in the summations in e_1 have the same expectation, and the summands are i.i.d., from the central limit theorem, we have

$$e_1 = O_p(n^{-1/2}), \text{ thus } \sup_{\boldsymbol{\theta}\in\Theta}|e_1| \xrightarrow{p} 0.$$

For e_2, e_3 :

Since $B(h,u) \equiv \frac{E\{hI(T>u)\}}{S(u-)}$, $B\left[\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i,T_i)\left\{1-G(T_i)\right\},u\right]$ and $B\left\{\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,C_i)I(T_i>C_i),u\right\}$ are predicable, they are continuous, as is $G(u)$, and hence locally bounded. By Corollary 3.4.1 in Fleming & Harrington (1991), we can show the uniform convergence on a bounded time interval $[0,\tau]$. We let $\widetilde{H}(u)$ stand for $\frac{B\left[\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i,T_i)\{1-G(T_i)\},u\right]}{G(u)}$ or $\frac{B\left\{\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,C_i)I(T_i>C_i),u\right\}}{G(u)}$.

Then, from Langlart's inequality, for any given $\eta, \xi > 0$, and $0 \le \tau < \infty$,

$$Pr\left[\sup_{0\le t\le\tau}\left\{\int_0^t \frac{1}{n}\widetilde{H}(u)dM^c(u)\right\}^2 \ge \xi\right] \le \frac{\eta}{\xi} + Pr\left[\int_0^\tau \left\{\frac{\widetilde{H}(u)}{n}\right\}^2 \lambda^c(u)R(u)du \ge \eta\right],$$

where

$$M^c(u) = \sum_{i=1}^n M_i^c(u),$$

and $\lambda^c(u)$ is the hazard function for the censoring time and $R(u)$ is the number of patients at risk at time u.

By Assumptions (A3), (A4), (A6) that the cumulative hazard function $\Lambda^c(u) < \infty$ and $\widetilde{H}^2(u) < \infty$, together with the fact that $\|\frac{R(u)}{n}\| < 1$, we have

$$Pr\left[\int_0^\tau \left\{\frac{\widetilde{H}(u)}{n}\right\}^2 \lambda^c(u)R(u) \geq \eta\right] \to 0.$$

Since η, ξ are arbitrary, we have

$$\sup_{t\leq\tau} |\int_0^t \frac{1}{n}\widetilde{H}(u)dM^c(u)| \xrightarrow{p} 0,$$

which implies

$$\sup_{\boldsymbol{\theta}\in\Theta} |e_j| \xrightarrow{p} 0, j = 2, 3,$$

by the martingale convergence theorem.

Therefore,

$$\sup_{\boldsymbol{\theta}\in\Theta} |\widehat{U}_n(\boldsymbol{\theta}) - U_0(\boldsymbol{\theta})| \xrightarrow{p} 0,$$

and in turn

$$\widehat{U}_n(\widehat{\boldsymbol{\theta}}) - U_0(\widehat{\boldsymbol{\theta}}) \xrightarrow{p} 0.$$

It is known that $U_0(\boldsymbol{\theta}_0) = 0$. Therefore, under Assumption (A8), we can use the Taylor expansion to expand U_0 at $\boldsymbol{\theta}_0$ to obtain $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \xrightarrow{p} 0$. This proves the results.

<u>Theorem</u> 2 *Under assumptions A1-A8, we have the asymptotic expansion*

$$\begin{aligned} & -\{A + o_p(1)\}n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ = \quad & n^{-1/2}\sum_{i=1}^n \{\delta_i \mathbf{S}_{\boldsymbol{\theta}_0,\text{eff}}(w_i, x_i) + (1-\delta_i)\mathbf{Q}_{\boldsymbol{\theta}_0,i}(w_i, x_i) \\ & +\rho_i(\boldsymbol{\theta}_0)\} + o_p(1), \end{aligned} \tag{A.7}$$

where

$$A = E\left\{\Delta_i \frac{\partial \mathbf{S}_{\boldsymbol{\theta}_0,\text{eff}}(W_i, X_i)}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + (1-\Delta_i)\frac{\partial \mathbf{Q}_{\boldsymbol{\theta}_0,i}(W_i, X_i)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right\},$$

and

$$\begin{aligned}\rho_i(\boldsymbol{\theta}) &= \frac{\delta_i \mathbf{S}_{\boldsymbol{\theta},\text{eff}}(w_i,x_i)}{G(t_i)}\{1-G(t_i)\} - \frac{\delta_i}{G(t_i)}E\{I(t_i>C_j)\mathbf{Q}_{\boldsymbol{\theta},j}(w_i,C_j)\mid \mathbf{v}_i\}\\ &+ \int \frac{B[\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_j,T_j)\{1-G(T_j)\},u]}{G(u)}dM_i^c(u)\\ &\quad -\int \frac{B\{\mathbf{Q}_{\boldsymbol{\theta},j}(W_j,C_j)I(T_j>C_j),u\}}{G(u)}dM_i^c(u).\end{aligned}$$

Consequently, when $n\to\infty$,

$$n^{1/2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\to N\{0,A^{-1}\boldsymbol{\Omega}(A^{-1})^{\mathrm{T}}\}$$

in distribution, where

$$\boldsymbol{\Omega} = E\left(J_1(\boldsymbol{\theta}_0)^{\otimes 2} + \int E\left[\{\boldsymbol{\Omega}_1(\boldsymbol{\theta}_0,u)+\boldsymbol{\Omega}_2(\boldsymbol{\theta}_0,u)+\boldsymbol{\Omega}_3(\boldsymbol{\theta}_0,u)\}^{\otimes 2}\right]\lambda^c R_i(u)du\right),$$

and

$$\begin{aligned}J_1(\boldsymbol{\theta}) &\equiv \mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i,X_i) - E\{I(X_i>C_j)\mathbf{Q}_{\boldsymbol{\theta},j}(W_i,C_j)\mid \mathbf{V}_i\} + \{1-G(X_i)\}\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,X_i),\\ \boldsymbol{\Omega}_1(\boldsymbol{\theta},u) &\equiv -\frac{\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_i,X_i) - E\{I(X_i>C_j)\mathbf{Q}_{\boldsymbol{\theta},j}(W_i,C_j)\mid \mathbf{V}_i\} + G(X_i)\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,X_i)}{G(u)},\\ \boldsymbol{\Omega}_2(\boldsymbol{\theta},u) &\equiv \frac{B[\mathbf{S}_{\boldsymbol{\theta},\text{eff}}(W_j,T_j)\{1-G(T_j)\},u]}{G(u)},\\ \boldsymbol{\Omega}_3(\boldsymbol{\theta},u) &\equiv -\frac{B\{\mathbf{Q}_{\boldsymbol{\theta},j}(W_j,C_j)I(T_j>C_j),u\}}{G(u)}.\end{aligned}$$

Here, $\mathbf{v}_i = (w_i,t_i,\delta_i)^{\mathrm{T}}$ is the observation of the ith individual, M_i^c and λ^c denote the martingale representation and hazard rate for the censoring time respectively and $R_i(t)\equiv I(X_i\geq t)$.

In practice, we approximate the matrix A by using the numeric derivatives of the estimating equations. To obtain $\boldsymbol{\Omega}$, we first estimate $E\{J_1(\boldsymbol{\theta}_0)^{\otimes 2}\}$ and

$$E\left[\{\boldsymbol{\Omega}_1(\boldsymbol{\theta}_0,u)+\boldsymbol{\Omega}_2(\boldsymbol{\theta}_0,u)+\boldsymbol{\Omega}_3(\boldsymbol{\theta}_0,u)\}^{\otimes 2}\right]$$

via their empirical counterparts, which are respectively denoted by $\widehat{\mathcal{E}}_J$ and $\widehat{\mathcal{E}}(\boldsymbol{\theta}_0,u)$. We then approximate

$$E\left(\int E\left[\{\boldsymbol{\Omega}_1(\boldsymbol{\theta}_0,u)+\boldsymbol{\Omega}_2(\boldsymbol{\theta}_0,u)+\boldsymbol{\Omega}_3(\boldsymbol{\theta}_0,u)\}^{\otimes 2}\right]\lambda^c R_i(u)du\right)$$

using

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathcal{E}}(\boldsymbol{\theta}_0, X_i)(1-\Delta_i).$$

Proof:

$$\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0 = -\left\{\frac{\partial \widehat{U}_n(\widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right\}^{-1}\widehat{U}_n(\boldsymbol{\theta}_0),$$

where $\widetilde{\boldsymbol{\theta}}$ is the point on the line connecting $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$. First, we have

$$\begin{aligned}
& \frac{\partial \widehat{U}_n(\widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \\
= & \left[n^{-1}\sum_{i=1}^{n}\delta_i\frac{\partial \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(w_i,t_i)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}+(1-\delta_i)\frac{\partial \widehat{\mathbf{Q}}_{\boldsymbol{\theta},i}(w_i,x_i)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right]_{\widetilde{\boldsymbol{\theta}}} \\
\xrightarrow{p} & E\left\{\Delta_i\frac{\partial \mathbf{S}_{\widetilde{\boldsymbol{\theta}},\mathrm{eff}}(w_i,x_i)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}+(1-\Delta_i)\frac{\partial \mathbf{Q}_{\widetilde{\boldsymbol{\theta}}_n,i}(w_i,x_i)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right\} \\
\to & E\left\{\Delta_i\frac{\partial \mathbf{S}_{\boldsymbol{\theta}_0,\mathrm{eff}}(w_i,x_i)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}+(1-\Delta_i)\frac{\partial \mathbf{Q}_{\boldsymbol{\theta}_0,i}(w_i,x_i)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right\}.
\end{aligned}$$

The first convergence follows the weak law of large numbers. Further, because $\widehat{\boldsymbol{\theta}}$ is a consistent estimator for $\boldsymbol{\theta}_0$, while $|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0| \le |\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0|$, hence $|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0| = o_p(1)$. Note that $\widetilde{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}$ depend on the sample size n, and the inequality holds for any n. By the continuous mapping theorem, the second convergence follows. Second, by the central limit theorem, we have

$$\sqrt{n}\widehat{U}_n(\boldsymbol{\theta}_0) \xrightarrow{D} N(\mu, \boldsymbol{\Omega}),$$

where

$$\begin{aligned}
\mu &= \lim_n E\{\Delta_i \mathbf{S}_{\boldsymbol{\theta}_0,\mathrm{eff}}(W_i,X_i)+(1-\Delta_i)\widehat{\mathbf{Q}}_{\boldsymbol{\theta}_0,i}(W_i,X_i)\} \\
&= \lim_n E\{U_n(\boldsymbol{\theta}_0)+L_1\} = 0,
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\Omega} &= E[\{\Delta_i \mathbf{S}_{\boldsymbol{\theta}_0,\mathrm{eff}}(W_i,T_i)+(1-\Delta_i)\widehat{\mathbf{Q}}_{\boldsymbol{\theta}_0,i}(W_i,X_i)\}^{\otimes 2}]-\mu^2 \\
&= E[\{\Delta_i \mathbf{S}_{\boldsymbol{\theta}_0,\mathrm{eff}}(W_i,T_i)+(1-\Delta_i)\mathbf{Q}_{\boldsymbol{\theta}_0,i}(W_i,X_i)+\rho_i(\boldsymbol{\theta})\}^{\otimes 2}].
\end{aligned}$$

Plugging in the expression for $\rho_i(\boldsymbol{\theta})$, we define

$$J(\boldsymbol{\theta}) \equiv \Delta_i \mathbf{S}_{\boldsymbol{\theta}_0,\mathrm{eff}}(W_i,X_i)+(1-\Delta_i)\mathbf{Q}_{\boldsymbol{\theta}_0,i}(W_i,X_i)+\rho_i(\boldsymbol{\theta})$$

$$\begin{aligned} = & \frac{\Delta_i \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,X_i)}{G(X_i)} - \frac{\Delta_i}{G(X_i)} E\left\{I(X_i > C_j)\mathbf{Q}_{\boldsymbol{\theta}_0,j}(W_i,C_j) \mid \mathbf{V}_i\right\} \\ & +(1-\Delta_i)\mathbf{Q}_{\boldsymbol{\theta}_0,i}(W_i,X_i) \\ & + \int \frac{B\left[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_j,T_j)\left\{1-G(T_j)\right\},u\right]}{G(u)} dM_i^c(u) \\ & - \int \frac{B\left\{\mathbf{Q}_{\boldsymbol{\theta},j}(W_j,C_j)I(T_j>C_j),u\right\}}{G(u)} dM_i^c(u). \end{aligned}$$

By (A.1), in which

$$\frac{\Delta_i}{G(X_i)} = 1 - \int \frac{dM_i^c(u)}{G(u)}$$

and

$$(1-\Delta_i) = 1 - G(X_i) + G(X_i)\int \frac{dM_i^c(u)}{G(u)},$$

we have

$$\begin{aligned} & \frac{\Delta_i \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,X_i)}{G(X_i)} - \frac{\Delta_i}{G(X_i)} E\left\{I(X_i > C_j)\mathbf{Q}_{\boldsymbol{\theta}_0,j}(W_i,C_j) \mid \mathbf{V}_i\right\} \\ = & \left\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,X_i) - \int \frac{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,X_i)}{G(u)} dM_i^c(u)\right\} \\ & - \left\{E\left\{I(X_i > C_j)\mathbf{Q}_{\boldsymbol{\theta}_0,j}(W_i,C_j) \mid \mathbf{V}_i\right\} - \int \frac{E\left\{I(X_i > C_j)\mathbf{Q}_{\boldsymbol{\theta}_0,j}(W_i,C_j) \mid \mathbf{V}_i\right\}}{G(u)} dM_i^c(u)\right\} \end{aligned}$$

and

$$\begin{aligned} & (1-\Delta_i)\mathbf{Q}_{\boldsymbol{\theta}_0,i}(W_i,X_i) \\ = & \{1-G(X_i)\}\mathbf{Q}_{\boldsymbol{\theta}_0,i}(W_i,X_i) + G(X_i)\int \frac{\mathbf{Q}_{\boldsymbol{\theta}_0,i}(W_i,X_i)}{G(u)} dM_i^c(u). \end{aligned}$$

Therefore, $J(\boldsymbol{\theta})$ can be written as

$$\begin{aligned} J(\boldsymbol{\theta}) = & \mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,X_i) - E\left\{I(X_i > C_j)\mathbf{Q}_{\boldsymbol{\theta},j}(W_i,C_j) \mid \mathbf{V}_i\right\} + \{1-G(X_i)\}\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,X_i) \\ & - \int \frac{\left\{\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_i,X_i) - E\left\{I(X_i > C_j)\mathbf{Q}_{\boldsymbol{\theta},j}(W_i,C_j) \mid \mathbf{V}_i\right\} + G(X_i)\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,X_i)\right\}}{G(u)} \\ & \times dM_i^c(u) + \int \frac{B\left[\mathbf{S}_{\boldsymbol{\theta},\mathrm{eff}}(W_j,T_j)\left\{1-G(T_j)\right\},u\right]}{G(u)} dM_i^c(u) \\ & - \int \frac{B\left\{\mathbf{Q}_{\boldsymbol{\theta},j}(W_j,C_j)I(T_j>C_j),u\right\}}{G(u)} dM_i^c(u) \\ = & J_1(\boldsymbol{\theta}) + J_2(\boldsymbol{\theta}) + J_3(\boldsymbol{\theta}) + J_4(\boldsymbol{\theta}). \end{aligned}$$

As shown in Ma & Yin (2010), $J_1(\boldsymbol{\theta})$ is uncorrelated with the rest of the terms. Letting $\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3$ be defined as in the theorem, then

$$\left\{\Delta_i \mathbf{S}_{\boldsymbol{\theta}_0,\mathrm{eff}}(W_i,X_i) + (1-\Delta_i)\mathbf{Q}_{\boldsymbol{\theta},i}(W_i,X_i) + \rho_i(\boldsymbol{\theta})\right\}^{\otimes 2}$$

$$= \quad J_1(\boldsymbol{\theta})^{\otimes 2} + \left\{ \int \boldsymbol{\Omega}_1(\boldsymbol{\theta}_0, u) + \boldsymbol{\Omega}_2(\boldsymbol{\theta}_0, u) + \boldsymbol{\Omega}_3(\boldsymbol{\theta}_0, u) dM_i^c(u) \right\}^{\otimes 2}.$$

Further, we know that

$$\begin{aligned} & E\left(\left[\int \left\{ \boldsymbol{\Omega}_1(\boldsymbol{\theta}, u) + \boldsymbol{\Omega}_2(\boldsymbol{\theta}, u) + \boldsymbol{\Omega}_3(\boldsymbol{\theta}, u) \right\} dM_i^c(u) \right]^{\otimes 2} \right) \\ = \quad & E\left[\int \left\{ \boldsymbol{\Omega}_1(\boldsymbol{\theta}, u) + \boldsymbol{\Omega}_2(\boldsymbol{\theta}, u) + \boldsymbol{\Omega}_3(\boldsymbol{\theta}, u) \right\}^{\otimes 2} \lambda^c(u) R_i(u) du \right], \end{aligned}$$

therefore, we have

$$\boldsymbol{\Omega} = E\left(J_1(\boldsymbol{\theta}_0)^{\otimes 2} + E\left[\int \left\{ \boldsymbol{\Omega}_1(\boldsymbol{\theta}_0, u) + \boldsymbol{\Omega}_2(\boldsymbol{\theta}_0, u) + \boldsymbol{\Omega}_3(\boldsymbol{\theta}_0, u) \right\}^{\otimes 2} \lambda^c R_i(u) du \right] \right).$$

This proves the results.