Published in final edited form as:

Stat Med. 2013 May 30; 32(12): 2013–2030. doi:10.1002/sim.5652.

Time-varying Coefficient Proportional Hazards Model with Missing Covariates

Xiao Song^{1,*†} and Ching-Yun Wang²

¹Department of Epidemiology and Biostatistics, University of Georgia, U.S.A

²Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, U.S.A

SUMMARY

Missing covariates often arise in biomedical studies with survival outcomes. Existing approaches for missing covariates generally assume proportional hazards. The proportionality assumption may not hold in practice, as illustrated by data from a mouse leukemia study with covariate effects changing over time. To tackle this restriction, we study the missing data problem under the varying-coefficient proportional hazards model. Based on the local partial likelihood approach, we develop inverse selection probability weighted estimators. We consider reweighting and augmentation techniques for possible improvement of efficiency and robustness. The proposed estimators are assessed via simulation studies and illustrated by application to the mouse leukemia data.

Keywords

augmentation; inverse probability weighting; local partial likelihood; reweighting

1. Introduction

Survival data often arise in medical studies and clinical trials where the time to an endpoint is subject to censoring. In real data analysis, some covariates may be missing, either by design or by happenness. An example is the mouse leukemia data in [1]. The study was conducted in the laboratories of Dr. Robert Nowinski of the Fred Hutchinson Cancer Research Center to investigate genetic and viral factors in the development of spontaneous leukemia in mice. Information was collected on survival time and covariates for 204 mice. A total of 67 mice died of leukemia. Two covariates of interest were the Gpd-1 phenotype and viral production level. The viral production level is continuous for values less than 10^4 and truncated above. The Gpd-1 phenotype was a binary variable and only observed on a subset of mice survived beyond 400 days and was missing for other mice. It is of interest to assess the Gpd-1 phenotype and viral production level on the survival.

Among various survival models, the most widely used model is the proportional hazards model [2], which assumes the hazard of failure depending proportionally on covariates. However, in practice, the proportionality assumption may be too restricted. Using the complete data, a preliminary analysis of the mouse leukemia data was conducted. Following the analyses in [3], we dichotomize the virus level with 0 for a virus level $< 10^4$ and 1 otherwise. The two Gpd-1 phenotypes are denoted by 0 for b/b and 1 for b/a . Assuming piecewise constant covariate effects with two knots 527 and 654 days, which are equally

*Correspondence to: Xiao Song, Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Paul Coverdell Center, Room 129c, Athens, GA 30602, U.S.A. . [†]xsong@uga.edu.

spaced within the range of observed failure times, the estimated coefficients are plotted in Figure 1. It seems the effects are not constant. The SAS codes to obtain the estimates are given in the Appendix. The departure from the proportionality was further indicated by a test of proportionality (p -value=0.026). Figure 2 shows the plots of the scaled Schoenfeld residuals, which are clearly away from horizontal lines [4]. The corresponding R codes are given in the Appendix.

A useful model that can characterize such time-varying covariate effect is a modification of the proportional hazards model, the so-called time-varying coefficient proportional hazards model. Estimation of the regression coefficient has been studied by a number of authors, including [5-11]. For the mouse leukemia data, the time-varying covariate effects could be estimated using such methods using the complete data.

However, the complete data analysis ignores the missing observations. This generally causes loss in efficiency, and moreover, may lead to biased estimation. Specific approaches are needed to improve the efficiency and correct the bias. For the mouse leukemia data, the missingness was caused by study design and the missing at random (MAR) assumption seems reasonable in this problem [12]. Missing at random (MAR) implies that missingness is associated with only the observed data. Under the standard proportional hazards framework, various approaches have been proposed to deal with missing data under the MAR assumption. A simple approach to dealing with missing covariates is the regression calibration [13, 14]. However, the estimator is not consistent in general. Paik and Tsai proposed an approximate partial likelihood approach by taking the conditional expectation on the observed variables [15]. One complication of Paik and Tsai's method is that the survival time need to be discretized in order to estimate the conditional probabilities. A related method for continuous covariates using kernel estimation procedures was developed by Zhou and Wang [16]. Robins et al. proposed a class of inverse selection probability weighted (ISPW) estimators [17]. Wang and Chen developed an augmented ISPW estimator implemented by using an EM-type algorithm [18]. Qi et al. proposed an augmented ISPW estimator when both the selection probabilities and the conditional expectation of the unobserved scores are estimated nonparametrically [3]. Compared to the previous approaches, this approach has the advantage of robustness and flexibility. It is also more robust than the multiple imputation approach [19]. Xu et al. further applied a reweighting technique on the ISPW estimators by adding weights, the selection probability for the risk set, at each event time [20]. They showed that this might improve the efficiency of the ISPW estimators.

Although the missing data methodology has been extensively studied under the standard proportional hazards model, it is not yet investigated in literature for the time-varying coefficient model, as far as we know. As illustrated in the mouse leukemia data, there is a clear need to handle missingness with time-varying covariate effects. In this paper, we propose ISPW approaches for the time-varying coefficient proportional hazards model based on the local partial likelihood technique [21]. The local partial likelihood is an attractive method for the time-varying coefficient models. It does not have boundary effects and no boundary modifications are required. In addition, it is simple to implement. We first develop the local simple ISPW partial likelihood estimators. Then we apply both the augmentation and the reweighting techniques because of their robustness and efficiency gain under the standard proportional hazards model. One major difference between our approach and the approaches in [3] and [20] is the locality of the estimator due to the time-varying characteristic of the coefficients, which leads to some unique and interesting properties that are different from those under the standard proportional hazards model. We have also relaxed some key assumptions in their derivation of the asymptotic distribution (Appendix A, supplementary materials). In addition, we estimated the selection probability

nonparametrically as in [3] rather than the easier parametric (logistic regression) model based estimation as in [20]. These features, together with the locality of the estimator, pose much more challenges in deriving the asymptotic properties. Although the ISPW, reweighting and local likelihood techniques are not new, the combination of these techniques under the time-varying proportional hazards model is not trivial and important from both theoretical and applied views.

The rest of the article is organized as follows. We define the model in Section 2. The local simple ISPW partial likelihood estimator is proposed in Section 3 and the reweighting technique is considered in Section 4. The estimator is improved by augmentation in Section 5. The performance of the estimators is assessed by simulations in Section 6 and illustrated by application to real data in Section 7. The paper concludes with a discussion in Section 8. Technical proofs and variance estimates are given in the supplementary materials.

2. Model Definition

Let T denote the failure time and C denote the censoring time. The observed survival data are $V = \min(T, C)$ and $\Delta = I(T \leq C)$, where $I(\cdot)$ is the indicator function. Let $X = (X_o^T, X_m^T)^T$ denote p covariates, where X_o is always observed and X_m may be missing either by design or by happenness. Let $R = 1$ if X_m is observed and 0 otherwise. We assume MAR, that is, $P(R = 1 | X, V, \Delta) = P(R = 1 | X_o, V, \Delta)$. Let W denote the r components in $O = (X_o^T, V, \Delta)^T$ that may impact the selection probability such that $P(R = 1 | O) = P(R = 1 | W)$. Note that W may contain components that are independent of R . Let $\pi(w) = P(R = 1 | W = w)$ be the selection probability. Suppose the dataset includes n subjects and the observations are identically and independently distributed (iid) across the subjects. Use subscript i to denote variables for the i th subject, for example, $X_i = (X_{oi}^T, X_{mi}^T)^T$ denote the covariates for the i th subject.

The hazard of failure is assumed to follow the time-varying coefficient proportional hazards model

$$\begin{aligned} \lambda_i(u|X) &= \lim_{du \rightarrow 0} du^{-1} \Pr(u \leq T_i < u+du | T_i \geq u, C_i) \\ &= \lambda_0(u) \exp \{ \beta_0^T(u) X_i \}, \end{aligned} \quad (1)$$

where $\lambda_0(u)$ is an unspecified baseline hazard function, $\beta_0(u)$ is the regression coefficient which may vary over time. Model (1) may be viewed as an extension of the standard proportional hazards model where the regression coefficient is a constant. This equation makes explicit the assumption that censoring is noninformative. Our interest focuses on estimation of $\beta_0(u)$.

3. Local Simple ISPW Estimator

Since $\beta_0(\cdot)$ is a function of time, we consider estimating the regression coefficient at any given time $t \in (0, L)$. By a local linear approximation at t , we can write $\beta(u) \approx \beta(t) + (u - t)\beta'(t)$. Let $u_t = (1, u - t)^T$, $b_0 = \beta(t)$, and $b_1 = \beta'(t)$ be the derivative of $\beta(t)$. Write

$b = (b_0^T, b_1^T)^T$, $X(u, u - t) = u_t \otimes X$, where $\otimes$ is the operator for Kronecker product. For any vector c , define $c^{\otimes k}$ 1, c , cc^T for $k = 0, 1, 2$, respectively. Let $N_\lambda(u) = I(V_j \geq u, \Delta_j = 1)$ be the counting process, $Y_\lambda(u) = I(V_j \geq u)$ be the at risk process, and $L = \sup_u \{P(Y(u) = 1) > 0\}$. When there is no missing data, the local partial likelihood estimator of b [10] solves the estimating equation

$$\tilde{U}(b) = n^{-1} \sum_{i=1}^n \int_0^L K_h(u-t) (H^{*-1} u_i) \otimes \left\{ X_i - \frac{\tilde{G}_1(b, u, u-t)}{\tilde{G}_0(b, u, u-t)} \right\} dN_i(u) = 0. \quad (2)$$

Here $K_h(u) = h^{-1} K(h^{-1}u)$ with $K(\cdot)$ being a kernel function and h being the bandwidth, $H^* = \text{diag}(1, h)$, and

$$\tilde{G}_k(b, u, u-t) = n^{-1} \sum_{i=1}^n Y_i(u) X_i^{\otimes k} \exp \{b^T X_i(u, u-t)\},$$

for $k = 0, 1, 2$. In reality, part of the covariate X, X_m , may be missing. Then equation (2) can not be used directly. A common technique for dealing with missing data is the inverse selection probability weighting (ISPW) [22]. Applying this technique, we construct the local simple ISPW partial likelihood estimating equation

$$\widehat{U}_{sw}(b, \pi) = n^{-1} \sum_{i=1}^n \frac{R_i}{\pi_i} \int_0^L K_h(u-t) (H^{*-1} u_i) \otimes \left\{ X_i - \frac{\widehat{G}_{1,sw}(b, \pi, u, u-t)}{\widehat{G}_{0,sw}(b, \pi, u, u-t)} \right\} dN_i(u) = 0, \quad (3)$$

where $\pi_i = \pi(W_i)$, and

$$\widehat{G}_{k,sw}(b, \pi, u, u-t) = n^{-1} \sum_{i=1}^n \frac{R_i}{\pi_i} Y_i(u) X_i^{\otimes k} \exp \{b^T X_i(u, u-t)\}$$

for $k = 0, 1, 2$. Let $\widehat{b}_{sw}(\pi)$ be the solution to equation (3).

Next we consider the asymptotic properties of the estimator $\widehat{b}_{sw}(\pi)$. Due to the involvement of the kernel smoother in the counting processes, the empirical process arguments based on iid observations in [3] cannot be adapted directly. We instead use the empirical process theories for kernel smoothed functions and independent triangular arrays to derive the asymptotic properties.

We first define some notations. Write $b_{T0} = \beta_0(t)$, $b_T = (b_{T0}^T, b_{T1}^T)^T$ and $H = \text{diag}(I_p, hI_p)$, which denotes the $2p \times 2p$ diagonal matrix with the first p diagonal elements equal to 1 and the last p elements h ; here I_p is a $p \times p$ identity matrix. For two matrixes A and B , write $A \succ_{pd} B$ if $A - B$ is positive definite, and $A \succeq_{pd} B$ if $A - B$ is nonnegative definite. Denote the asymptotic variance of the local partial likelihood estimator from the full data by Σ . It can be shown

that, under some regularity conditions, $H \{\widehat{b}_{sw}(\pi) - b_T\}$ converges in probability to 0. Moreover,

$$(nh)^{1/2} \left[H \{\widehat{b}_{sw}(\pi) - b_T\} - \frac{h^2}{2} Q_\mu^{-1} (\mu_2, \mu_3)^T \otimes \beta'^T(t) \right] \xrightarrow{D} N \left(0, \sum_{sw} \right),$$

where μ_s ($s = 0, \dots, 3$) and Q_μ are given in Appendix A (supplementary materials), and Σ_{sw} and its estimate are given in Appendix B (supplementary materials). In addition, it can be shown that $\Sigma_{sw} \succeq_{pd} \Sigma$. This indicates that the local simple ISPW estimator loses some efficiency due to the missing covariates. From the formula of Σ_{sw} , we can see that the efficiency loss increases with the decreases of the selection probability π .

The selection probability π is usually known for design-based missing, but unknown for non-design-based missing. When the selection probability π is unknown, it can generally be estimated from the data. By analogy to [3], we estimate $\pi(w) = P(R = 1|W = w)$ by the Nadaraya-Watson estimator [23,24]

$$\widehat{\pi}(w) = \frac{\sum_{i=1}^n R_i K_{h_*}^*(W_i - w)}{\sum_{i=1}^n K_{h_*}^*(W_i - w)}.$$

Here $k_{h_*}^*(\cdot) = h_*^{-r} K^*(\cdot/h_*)$, $K^*(\cdot) \geq 0$ is a kernel function satisfying $\int K^*(w) dw = 1$, and h_* is the bandwidth. It can be shown that, with some additional conditions, $H\{\widehat{b}_{sw}(\widehat{\pi}) - b_T\}$ converges in probability to 0 (Appendix C, supplementary materials). Moreover,

$$(nh)^{1/2} \left[H\{\widehat{b}_{sw}(\widehat{\pi}) - b_T\} - \frac{h^2}{2} Q_\mu^{-1}(\mu_2, \mu_3)^T \otimes \beta'^t(t) \right] \xrightarrow{D} N\left(0, \sum_{sw}^*\right),$$

where $\sum_{sw}^*$ and its estimate are given in Appendix C (supplementary materials).

Furthermore, it can be shown that $\sum_{sw} \geq_{pd} \sum_{sw}^* \geq_{pd} \sum$. Therefore, estimating the selection probabilities may improve the efficiency of the simple ISPW estimator even if π is known. Similar results are observed for the standard proportional hazards model [3]. For simple ISPW estimators, it can be shown that including more predictors in the model of the selection probability will not decrease the efficiency. However, from the proof in Appendix C, the efficiency gain can only be achieved by estimating π when the model of the selection probability π includes the observed failure time V ; otherwise, it makes no asymptotic difference whether to estimate π or not. This is different from what has been observed under the standard proportional hazards model [3].

4. Local Simple Reweighted ISPW Estimator

For the standard proportional hazards model, Xu et al. applied the reweighting technique by adding weights, the selection probability for the risk set, at each event time [20]. They showed that this might improve the efficiency of the ISPW partial likelihood estimators. We consider using the reweighting technique for the time-varying coefficient proportional hazards model as well. Let $p(t) = \Pr(R_i = 1|V_i = t)$ be the selection probability for the risk set at time t . Then $p(t)$ can be estimated by

$$\widehat{p}(t) = \frac{\sum_{i=1}^n R_i Y_i(t)}{\sum_{i=1}^n Y_i(t)}.$$

The local simple reweighted ISPW estimator can be obtained by solving

$$\widehat{U}_{si}(b, \pi) = n^{-1} \sum_{i=1}^n \frac{R_i}{\pi_i} \int_0^L \widehat{p}(u) K_h(u-t) (H^{*-1} u_t) \otimes \left\{ X_i - \frac{\widehat{G}_{1,si}(b, \pi, u, u-t)}{\widehat{G}_{0,si}(b, \pi, u, u-t)} \right\} dN_i(u) = 0, \quad (4)$$

where $\widehat{G}_{k,si}(b, \pi, u, u-t) = \widehat{p}(u) \widehat{G}_k(b, \pi, u, u-t)$ for $k = 0, 1$.

Let $\widehat{b}_{si}(\pi)$ be the solution of (4). It can be shown that $H\{\widehat{b}_{si}(\pi) - b_\tau\}$ has the same asymptotic distributions as that of $H\{\widehat{b}_{si}(\pi) - b_\tau\}$ and $H\{\widehat{b}_{si}(\widehat{\pi}) - b_\tau\}$ has the same asymptotic distributions as that of $H\{\widehat{b}_{si}(\widehat{\pi}) - b_\tau\}$ (Appendix D, supplementary materials). Therefore, in contrast to the standard proportional hazards model, applying the reweighting technique does not affect the efficiency of the local estimator for the time-varying coefficient proportional hazards model. An intuitive explanation is that reweighting achieving effects by adding weights equal to the risk set selection probability to different event times. But the local estimator only uses local data. The weights are generally close at the local events and hence have little effect on the estimators. As the bandwidth goes to zero, the effect of reweighting eventually disappear.

5. LOCAL AUGMENTED ISPW ESTIMATOR

The simple ISPW estimator may be inefficient because it only includes the subjects with complete observations in (3) and ignores the information from subjects with missing covariates. To improve efficiency, we apply the augmentation technique [18] and construct the augmented ISPW estimating equation, which can be written as

$$\widehat{U}_{aw}(b, \pi, E) = \widehat{U}_{1,aw}(b, \pi, E) + \widehat{U}_{2,aw}(b, \pi, E) = 0, \quad (5)$$

where

$$\widehat{U}_{1,aw}(b, \pi, E) = n^{-1} \sum_{i=1}^n \frac{R_i}{\pi_i} \int_0^L K_h(u-t) (H^{*-1} u_i) \otimes \left\{ X_i - \frac{\widehat{G}_{1,aw}(b, \pi, u, u-t)}{\widehat{G}_{0,aw}(b, \pi, u, u-t)} \right\} dN_i(u),$$

$$\widehat{U}_{2,aw}(a, \pi, E) = n^{-1} \sum_{i=1}^n \left(1 - \frac{R_i}{\pi_i} \right) \int_0^L K_h(u-t) (H^{*-1} u_i) \otimes \left\{ E\{X_i | O_i\} - \frac{\widehat{G}_{1,aw}(b, \pi, u, u-t)}{\widehat{G}_{0,aw}(b, \pi, u, u-t)} \right\} dN_i(u),$$

$$\widehat{G}_{k,aw}(b, u, u-t) = n^{-1} \sum_{i=1}^n \left(\frac{R_i}{\pi_i} Y_i(u) X_i^{\otimes k} \exp\{b^T X_i(u, u-t)\} \right) + \left(1 - \frac{R_i}{\pi_i} \right) Y_i(u) E\left[X_i^{\otimes k} \exp\{b^T X_i(u, u-t)\} | O_i \right], \quad k=0, 1, 2.$$

Comparing to the simple ISPW estimating equation, equation (5) is augmented by including the augmentation term $\widehat{U}_{2,aw}(a, \pi, E)$ and substituting the augmented ratio term $\widehat{G}_{1,aw}(b, \pi, u, u-t) / \widehat{G}_{0,aw}(b, \pi, u, u-t)$. Information from subjects with incomplete observations is thus included. It is easy to see that the augmented estimator has the so-called double robust property, that is, it is consistent if either the selection probability or the conditional distribution of X_i given O_i is correctly specified.

We first consider the simple case that π is known. The conditional expectations in (5) are generally unknown and can be estimated nonparametrically. Specifically, we estimate the conditional expectation $\phi(o) = E\{g(X_i) | O_i = o\}$ by the kernel estimator

$$\widehat{\phi}(o) = \frac{\sum_{i=1}^n R_{ig}(X_i) K_{h_*}^*(O_i - o)}{\sum_{i=1}^n R_i K_{h_*}^*(O_i - o)},$$

where $g(X_i)$ denote any function of X_i . Let $\widehat{U}_{aw}(b, \pi, \widehat{E})$ be the estimating equation by replacing $\phi(o)$ by $\widehat{\phi}(o)$ in $\tilde{U}_{aw}(b, \pi, E)$ and the corresponding estimator is denoted by $\widehat{b}_{aw}(\widehat{\phi}, \pi)$. We first consider the estimator $\widehat{b}_{aw}(\phi, \pi)$ when both ϕ and π are known. It can be shown that $H\{\widehat{b}_{aw}(\phi, \pi) - b_T\}$ converges in probability to 0. Moreover,

$$(nh)^{1/2} \left[H\{\widehat{b}_{aw}(\phi, \pi) - b_T\} - \frac{h^2}{2} Q_{\mu}^{-1}(\mu_2, \mu_3)^T \otimes \beta''^T(t) \right] \xrightarrow{D} N\left(0, \sum_{aw}\right),$$

where Σ_{aw} and its estimate are given in Appendix E (supplementary materials). In addition, it can be shown that $\sum_{sw} \geq_{pd} \sum_{sw}^* \geq_{pd} \sum_{aw} \geq \sum_r$. When $O = W$, $\sum_{sw}^* = \sum_{aw}$. This indicates that the augmented estimator is at least as efficient as the simple estimator. The augmented estimators have the same efficiency as the simple ISPW estimator when $O = W$ and the selection probability is estimated. This can be explained by the inclusion of the information from subjects with incomplete observations. Similar properties have been observed for the ISPW estimators for the standard proportional hazards model [3] and the generalized linear model [17].

In practice, the selection probability may not be known and we can estimate it using the nonparametric estimator $\widehat{\pi}$ as described in Section 3. It can be shown that $H\{\widehat{b}_{aw}(\widehat{\phi}, \pi) - b_T\}$ and $H\{\widehat{b}_{aw}(\widehat{\phi}, \widehat{\pi}) - b_T\}$ have the same asymptotic distributions as that of $H\{\widehat{b}_{aw}(\phi, \pi) - b_T\}$ (Appendix F, supplementary materials). This indicates that estimating the selection probabilities does not affect the efficiency of the augmented estimators. Specifically, whether the selection probability are correctly or incorrectly estimated, the efficiency of the augmented estimator is the same as the true selection probability is used. This conforms to the results in [3] for standard proportional hazards model.

We can further apply the reweighting technique on the augmented estimating equations. However it can be shown that the reweighting technique does not improve the local augmented ISPW estimators by similar arguments as in section 4. Bandwidth Selection. Based on the asymptotic distributions, we can derive the theoretically optimal bandwidth by analogy to [10]. However, this bandwidth depends on some unknown parameters and cannot be used directly in practice. A general technique for bandwidth selection is cross validation. Specifically, the data are randomly split into S equal sized subsets. Let $\widehat{\beta}^{(-s)}$ be the estimate of β with the s th subset D_s deleted. For each approach, we choose the “optimal” bandwidth by minimizing the prediction error $-\sum_{s=1}^S \xi(\widehat{\beta}^{(-s)})$, where

$$\xi(\beta) = \sum_{i \in D_s} \int \frac{R_i}{\pi_i} \left(\beta^T(u) X_i - \log \left[\sum_{j \in D_s} R_j / \pi_j Y_j(u) \exp \{ \beta^T(u) X_j \} \right] \right) dN_i(u)$$

is the estimated partial likelihood function based on the s th subset.

6. Simulation Studies Simulation

studies were conducted to evaluate the performance of the estimators. We considered the case where there are two covariates, $X = (X_o, X_m)$, in model (1). The covariate X_o was generated from a Bernoulli distribution with the probability 0.5, and X_m from a normal distribution with mean and variance equal to 1. The corresponding regression coefficients were $\beta_1(u) = \log(1 + 6u)$ and $\beta_2(u) = 1$, respectively. The baseline hazard was a constant 0.6. The censoring distribution was generated from a uniform distribution on the interval $[0, 4]$ and truncated at 1, leading to a censoring rate of 63%. The covariate X_o was always observed, and X_m may be missing. We considered four missingness cases: 1) Case I: The selection probability π depends on the censoring status Δ , which equals 0.4 if the subject is censored and 0.9 otherwise; 2) Case II: The selection probability π depends on Δ and the observed survival or censoring time V with $\pi = [1 + \exp\{0.8 - 2\Delta - V\}]^{-1}$; 3) Case III: The selection probability π depends on V , Δ and the nonmissing covariate X_o such that $\pi = [1 + \exp\{1.5 - 2\Delta - V - X_o\}]^{-1}$; 4) Case IV: The selection probability π depends only on X_o with $\pi = 0.9I(X_o = 0) + 0.4I(X_o = 1)$. The corresponding missing rates are 41%, 37%, 42% and 35% respectively in the four cases.

For each case, 500 Monte Carlo data sets were simulated with $n = 250$ and 500. For each dataset, we fit the model using the following estimators: 1) the estimator based on the full data, 2) the estimator based on the complete data, 3) the simple ISPW estimators (SWE) with the selection probability assumed known, depending on $W = \Delta$, $W = (V, \Delta)$ and $W = O = (V, \Delta, X_o)$ for case I, on $W = V$, $W = (V, \Delta)$ and $W = O = (V, \Delta, X_o)$ for case II and III, and on $W = X_o$, $W = (\Delta, X_o)$, $W = (V, \Delta)$ and $W = O = (V, \Delta, X_o)$ for case IV. 4) the corresponding simple reweighted ISPW estimators (SRWE) 4) the corresponding augmented ISPW estimators (AWE), 5) the corresponding augmented reweighted ISPW estimators (ARWE). For each method, the 95% Wald confidence interval for $\beta(t)$ was constructed. We used the Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ here and for the application in Section 7. It is well known in the literature that the choice of the kernel function does not have much impact on the results for kernel methods. The Epanechnikov kernel is usually used in kernel approaches because it has certain theoretically optimal properties for kernel density estimation [25]. The bandwidth h and h_* were set to $5n^{-1/2}$ and $3n^{-1/3}$ (about $14\sigma_V n^{-1/2}$ and $8\sigma_V n^{-1/3}$ with σ_V being the standard deviation of V) respectively. We estimated the regression coefficients at $t = 0.3, 0.5, 0.7$.

The results under case I for $n = 250$ and $n = 500$ are shown in Tables I and II. All the ISPW estimators show negligible bias and the coverage probabilities are close to the nominal level, while the complete data estimator is obviously biased and the coverage probabilities are well below the nominal level. The empirical standard deviations are close to the standard errors. The simple estimator based on the estimated selection probability $\hat{\pi}(\Delta)$ has efficiency similar to the simple estimator based on true π , but is less efficient than the simple estimator based on $\hat{\pi}(V, \Delta, X_o)$. The simple estimators perform similar to the corresponding simple reweighted estimators. The augmented estimators and the simple estimator with the selection probability depending on O show similar efficiency. The augmented estimators are more efficient than the other simple estimators and simple reweighted estimators in general for estimation, although the efficiency gain is more obvious for $\beta_1(t)$ compared to $\beta_2(t)$. The coverage probabilities are a bit low for $\beta_2(0.7)$ when $n = 250$ but are improved at the increased sample size 500.

The results under case II for $n = 250$ are shown in Table III. In addition to similar results we observed in case I, the simple weighted estimator based on the estimated selection

probability $\hat{\pi}(\Delta, V)$ has efficiency slightly better than or comparable to the simple estimator based on true π , but slightly worse or comparable to the simple estimator based on $\hat{\pi}(V, \Delta, X_o)$. The simple estimator based on the misspecified selection probability estimate $\hat{\pi}(V)$ is biased and has worse coverage probabilities, while the corresponding augmented estimator works well. Similar phenomena have been observed for the corresponding simple reweighted estimators.

The results under case III for $n = 250$ are shown in Table A in the supplementary materials, which carry the same message as we have observed in case I and case II.

The results from the simple estimators under case IV for $n = 500$ are shown in Table IV. The complete case estimator is not biased since the missingness does not depend on the survival outcome. The simple estimators with π estimated based on (V, Δ, X_o) is more efficient than those based on (V, X_o) , which is more efficient than those based on (Δ, X_o) . The latter is slightly more efficient than the estimator based on X_o only, which has similar efficiency to the estimators based on the true selection probability or on the complete data.

In summary, the simulation results conform to the theoretical results we have derived in Sections 3 to 5. Specifically, (a) The augmented estimators and the simple estimators based on the estimated selection probability $\hat{\pi}$ involving V are more efficient than the simple estimator based on true π ; (b) The augmented estimators are more efficient than the simple estimator based on $\hat{\pi}$, except when (V, Δ, X_o) are used for estimating $\hat{\pi}$; (c) The complete-case analysis generates biased estimators when the selection probability depends on outcomes; (d) Inconsistent estimation of the selection probability π affects the consistency of the simple estimators, but not of the augmented estimators; (e) The simple estimator based on $\hat{\pi}$ not involving V does not improve the efficiency over the simple estimator based on true π asymptotically, but may improve the efficiency for finite sample; (f) The reweighting technique does not improve efficiency for either simple or augmented estimators. Items (a)-(d) conform to what have been observed in [3] for standard proportional hazards model. A similar result of item (e) was reported in [26] for logistic regression model. Item (f) is a special property for local estimators under the varying coefficient model.

7. Application

We applied the proposed methods to the mouse leukemia data described in Section 1. The data have been analyzed under the standard proportional hazards model in [3]. Since our preliminary analysis indicated a departure from proportionality, we fit the time-varying coefficient model including the dichotomous virus level X_o , 0 for a virus level $< 10^4$ and 1 otherwise, and Gpd-1 phenotype indicator $X_m = I(\text{Gpd-1 phenotype} = b/a)$. The survival time of interest was time to death from leukemia. The 5-fold cross validation is used for bandwidth selection. Based on the averaged prediction error over a grid of t , the selected bandwidth h are mostly between $13\sigma_V n^{-1/2}$ to $16\sigma_V n^{-1/2}$, and h^* between $7\sigma_V n^{-1/3}$ and $9\sigma_V n^{-1/3}$.

The estimated coefficients between 500 and 700 days for $h = 14\sigma_V n^{-1/2}$ (118 days) and $h^* = 8\sigma_V n^{-1/3}$ (156 days) are shown in Figure 3. All estimates show that higher virus level shortens the survival time, and the effect of virus level seems diminish over time. The Gpd-1 phenotype has a significant effect on survival which remains relatively stable before 630 days and the magnitude seems decreasing thereafter. It seems that mice with Gpd-1 phenotype b/a live longer than those with phenotype b/b . The simple estimates are close to the augmented estimates except for when t is close to 500, which may be due to less accurate estimation of the selection probability close to the boundary. The augmented

estimators yield narrower confidence intervals for both regression coefficients. The reweighted estimates are close to the unweighted estimates (not shown for brevity). In addition, it appears that besides the observation time V , adding other variables to estimate the selection probabilities has little effect on the estimation of the regression coefficients. This is similar to what have been observed in [3].

The varying-coefficient model offers more insights on the coefficient trajectory. The coefficient estimates for virus level decreased from around 4.5 to -0.5 for the simple estimates and from 2.2 to 0.1 for the augmented estimates, in contrast to the estimated constant effect of 1.2 and 1.0, respectively, in [3]. The coefficient estimates for Gpd-1 varies between -3.8 and -1.3 , in contrast to an estimated constant effect between -1.4 to -1.8 in [3].

8. Discussion

We have developed ISPW estimators for missing covariates data under the varying-coefficient proportional hazards model, including the simple, reweighted and augmented estimators. In contrast to under the standard proportional hazards model, the reweighting technique does not improve the efficiency of either the simple or the augmented estimators, although the augmentation may still improve the efficiency and robustness of the simple estimators.

In this article, we have focused on estimation of the regression coefficient in model (1). It may be also of interest to estimate the cumulative hazard function $\Lambda_0 = \int_0^u \lambda_0(t) dt$. Note that $\lambda_0(u)$ can be rewritten as $\Lambda_0(u) = \int_0^u dEN_i(u) / E[Y_i(u) \exp\{\beta_0^T(u) X_i\}]$, we can estimate $\Lambda_0(u)$ by $\widehat{\Lambda}_0(u) = \int_0^u \sum_{i=1}^n R_i / \pi_i dN_i(u) / \left[\sum_{i=1}^n R_i / \pi_i Y_i(u) \exp\{\widehat{\beta}^T(u) X_i\} \right]$, when π is known, and replace π by $\widehat{\pi}$ when π is unknown.

By adopting the kernel smoothing technique on estimation of the selection probability, we do not need to parametrically model the selection probability. This works well when the dimension of (V, Δ, X_0) is low. However, when the dimension of (V, Δ, X_0) increases, the performance may degrades rapidly. One strategy is to adopt a flexible selection probability model such as the generalized additive model or the partial-linear model. This will moderate the problem of high-dimensionality and meanwhile allow some flexibility on modeling the selection probability.

In this paper, we only consider the inference of $\beta(t)$ at a given point t . To formally test if $\beta(t)$ is a constant, it is desirable to obtain the simultaneous confidence band and adopt techniques as those in [11] for this purpose. But this is outside the range of this paper and will be investigated in our future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was partially supported by NIH grants R01ES017030 (Wang and Song) and CA53996 (Wang), and NSF grant DMS-1106816 (Song). The views expressed in this article are those of the authors, and do not necessarily represent the views of NIH or NSF.

Appendix

The SAS codes to fit the piecewise constant coefficient model is given as follows. Here the variables V , Δ , X_o , X_m and R denote V , Δ , X_o , X_m and R respectively. The cutoff points are given by the macro variables cutoff 1 and cutoff 2. This can be easily extended to one or more than two cutoff points.

```
data mouse;
    set mouse;
    if R=1;
run;

proc phreg data=mouse;
    model V*delta(0)=z11-z13 z21-z23;
    z11=z1*(days<=&cutoff_1);
    z12=z1*(&cutoff_1<days<=&cutoff_2);
    z13=z1*(&cutoff_2<days);
    z21=z2*(days<=&cutoff_1);
    z22=z2*(&cutoff_1<days<=&cutoff_2);
    z23=z2*(&cutoff_2<days);
run;
```

The R codes for testing proportionality are given as follows. The dataset are stored in the dataframe mouse with the variables V , Δ , X_o , X_m and R , respectively.

```
fit <- coxph(Surv(V,delta)~z1+z2,data=mouse,subset=(R==1))

cox.test <- cox.zph(fit,transform="identity")

print(cox.test)

plot(cox.test[1])

plot(cox.test[2])
```

REFERENCES

1. Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. Wiley; New York: 2002.
2. Cox DR. Regression models and lifetables (with discussion). Journal of the Royal Statistical Society, Series B. 1972; 34:187–200.
3. Qi L, Wang CY, Prentice RL. Weighted estimators for proportional regression with missing covariates. Journal of the American Statistical Association. 2005; 100:1250–1263.
4. Therneau, T.; Grambsch, P. Modelling Survival Data: Extending the Cox Model. Springer; New York: 2000.
5. Zucker D, Karr A. Nonparametric survival analysis with time-dependnet covaraite effects: a penalized likelihood approach. Annals of Statistics. 1990; 18:329–353.

6. Murphy P, Sen P. Time-dependent coefficients in a Cox-type regression models. *Theory of Probability and Its Applications*. 2004; 99:832–844.
7. Hastie T, Tibshirani R. Varying-coefficient Models. *Journal of the Royal Statistical Society, Series B*. 1993; 55:757–796.
8. Martinussen T, Scheike TH, Skovgaard IM. Efficient Estimation of Fixed and Time-varying Covariate Effects in Multiplicative Intensity Models. *Scandinavian Journal of Statistics*. 2002; 29:832–844.
9. Winnett A, Sasieni P. Iterated residuals and time-varying covariate effects in Cox regression. *Journal of the Royal Statistical Society, Series B*. 2003; 65:473–488.
10. Cai Z, Sun Y. Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics*. 2003; 30:93–111.
11. Tian L, Zucker D, Wei LJ. On the Cox model with time-varying regression coefficients. *Biometrics*. 2005; 100:172–183.
12. Rubin DB. Inference and Missing Data. *Biometrika*. 1976; 63:581–592.
13. Prentice RL. Covariate measurement errors and parameter estimates in a failure time regression model. *Biometrika*. 1982; 69:331–342.
14. Wang CY, Hsu L, Feng ZD, Prentice RL. Regression calibration in failure time regression. *Biometrics*. 1997; 53:131–145. [PubMed: 9147589]
15. Paik MC, Tsai WY. On using the Cox proportional hazards model with missing covariates. *Biometrika*. 1997; 84:579–593.
16. Zhou H, Wang CY. Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society, Series B*. 2000; 62:657–665.
17. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*. 1994; 89:846–866.
18. Wang CY, Chen HY. Augmented inverse probability weighted estimator for Cox missing covariate regression. *Biometrics*. 2001; 57:414–419. [PubMed: 11414564]
19. Qi L, Wang YF, He Y. A comparison of multiple imputation and fully augmented weighted estimators for Cox regression with missing covariates. *Statistics in Medicine*. 2010; 29:2592–2604. [PubMed: 20806403]
20. Xu Q, Paik MC, Luo X, Tsai WY. Reweighting estimators for Cox regression with missing covariates. *Journal of the American Statistical Association*. 2009; 104:1155–1167.
21. Fan J, Gijbels I, King M. Local Likelihood and Local Partial Likelihood in Hazard Regression. *Annals of Statistics*. 1997; 25:1661–1690.
22. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*. 1952; 47:663–685.
23. Nadaraya EA. On estimating regression. *Theory of Probability and Its Applications*. 1964; 9:141–142.
24. Watson GS. Smooth regression analysis. *Sankhya, Series A*. 1964; 26:359–372.
25. Wand, MP.; Jones, MC. *Kernel Smoothing*. Chapman & Hall/CRC; New York: 1995.
26. Wang C, Paik MC. Efficiencies of methods dealing with missing covariates in regression analysis. *Statistica Sinica*. 2006; 16:1169–1192.

Figure 1. Estimation of piecewise constant coefficients for the Mouse Leukemia Data. Left panel, viral production level; Right panel, Gpd-1 phenotype.

Figure 2. Scaled Schoenfeld residual against event time from a standard proportional hazards model fit to the Mouse Leukemia Data. Left panel, viral production level; Right panel, Gpd-1 phenotype.

Figure 3.

Analysis of the Mouse Leukemia Data using the varying coefficient proportional hazards model. Left panel, coefficient of viral production level; Right panel, coefficient of Gpd-1 phenotype. Pointwise 95% confidence bands are shown with the outer curves and the estimates themselves are shown with the center curves.

Table I

Simulation results for $n = 250$ under case I. $\lambda(\theta) = \lambda_0(\theta) \exp\{\beta_1(\theta)X_0 + \beta_2(\theta)X_m\}$ and $\pi = 0.9I(\Delta = 1) + 0.4I(\Delta = 0)$. B, bias; SD, empirical standard deviation across simulated data sets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

	$t = 0.3$				$t = 0.5$				$t = 0.7$			
	$\beta_1(\theta)$				$\beta_2(\theta)$				$\beta_2(\theta)$			
	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP
Full Data	-.037	.308	.302	.940	.021	.325	.330	.954	.067	.372	.373	.964
Complete Data	-.350	.328	.319	.774	-.313	.345	.347	.818	-.224	.392	.392	.892
SWE- $\pi(\Delta)$	-.048	.361	.348	.924	.012	.393	.387	.948	.075	.460	.451	.964
SWE - $\hat{\pi}(\Delta)$	-.043	.362	.346	.926	.017	.396	.386	.946	.080	.462	.450	.958
SWE - $\hat{\pi}(V, \Delta)$	-.044	.363	.345	.928	.017	.396	.385	.944	.081	.464	.449	.956
SWE - $\hat{\pi}(V, \Delta, X_0)$	-.040	.332	.326	.952	.020	.362	.362	.958	.079	.416	.417	.968
SRWE- $\pi(\Delta)$	-.048	.361	.349	.924	.012	.395	.388	.948	.075	.463	.453	.964
SWE - $\hat{\pi}(\Delta)$	-.043	.363	.347	.926	.017	.397	.388	.948	.080	.465	.453	.964
SWE - $\hat{\pi}(V, \Delta)$	-.044	.363	.347	.928	.017	.398	.387	.946	.081	.466	.452	.960
SWE - $\hat{\pi}(V, \Delta, X_0)$	-.040	.332	.336	.958	.020	.364	.374	.964	.079	.419	.435	.972
AWE- $\pi(\Delta)$	-.043	.328	.320	.940	.016	.351	.353	.962	.066	.401	.410	.966
SWE - $\hat{\pi}(\Delta)$	-.042	.329	.320	.940	.016	.353	.354	.960	.070	.404	.412	.962
SWE - $\hat{\pi}(V, \Delta)$	-.042	.329	.320	.940	.017	.353	.354	.960	.072	.408	.433	.966
SWE - $\hat{\pi}(V, \Delta, X_0)$	-.040	.329	.320	.942	.021	.357	.354	.954	.076	.416	.735	.962
ARWE- $\pi(\Delta)$	-.042	.328	.320	.940	.016	.353	.355	.962	.068	.405	.412	.958
SWE - $\hat{\pi}(\Delta)$	-.042	.329	.321	.940	.017	.354	.355	.964	.070	.406	.414	.962
SWE - $\hat{\pi}(V, \Delta)$	-.042	.329	.321	.940	.017	.354	.355	.962	.072	.413	.475	.966
SWE - $\hat{\pi}(V, \Delta, X_0)$	-.039	.330	.321	.942	.021	.358	.355	.960	.076	.417	.544	.962
	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP
Full Data	-.014	.164	.155	.930	-.020	.193	.172	.920	-.031	.212	.199	.942
Complete Data	.167	.169	.164	.778	.164	.201	.181	.768	.128	.236	.210	.868

	$t = 0.3$				$t = 0.5$				$t = 0.7$			
					$\beta_t(t)$							
	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP
SWE- $\pi(\Delta)$	-.018	.181	.175	.950	-.030	.226	.200	.934	-.053	.283	.245	.916
SWE - $\hat{\pi}(\Delta)$	-.021	.180	.173	.950	-.033	.227	.199	.928	-.056	.284	.245	.910
SWE - $\hat{\pi}(V, \Delta)$	-.021	.180	.173	.952	-.033	.225	.198	.930	-.057	.284	.244	.910
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.023	.180	.171	.948	-.035	.228	.197	.918	-.059	.284	.242	.916
SRWE- $\pi(\Delta)$	-.018	.181	.175	.950	-.030	.227	.201	.936	-.053	.284	.247	.916
SWE - $\hat{\pi}(\Delta)$	-.021	.181	.174	.950	-.033	.227	.200	.930	-.056	.285	.247	.914
SWE - $\hat{\pi}(V, \Delta)$	-.021	.180	.174	.952	-.033	.226	.200	.930	-.057	.285	.246	.916
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.023	.181	.172	.948	-.035	.228	.198	.920	-.059	.286	.245	.918
AWE- $\pi(\Delta)$	-.022	.180	.177	.956	-.034	.222	.207	.950	-.054	.273	.263	.942
SWE - $\hat{\pi}(\Delta)$	-.023	.180	.177	.956	-.035	.224	.207	.948	-.057	.277	.262	.950
SWE - $\hat{\pi}(V, \Delta)$	-.023	.179	.178	.954	-.034	.223	.207	.946	-.058	.277	.263	.946
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.024	.181	.178	.956	-.037	.228	.208	.946	-.063	.295	.277	.942
ARWE- $\pi(\Delta)$	-.022	.180	.178	.956	-.033	.223	.208	.952	-.055	.274	.264	.948
SWE - $\hat{\pi}(\Delta)$	-.023	.180	.178	.956	-.034	.225	.208	.948	-.057	.278	.264	.950
SWE - $\hat{\pi}(V, \Delta)$	-.023	.180	.178	.954	-.034	.224	.208	.950	-.058	.282	.267	.950
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.024	.181	.178	.956	-.037	.229	.209	.948	-.062	.293	.277	.944

Table II

Simulation results for $n = 500$ under case I. $\lambda_0(t) = \lambda_0(t) \exp\{\beta_1(t)X_0 + \beta_2(t)X_m\}$ and $\pi = 0.9I(\Delta = 1) + 0.4I(\Delta = 0)$. B, bias; SD, empirical standard deviation across simulated data sets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

	$t = 0.3$				$t = 0.5$				$t = 0.7$			
	$\beta_1(t)$											
	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP
Full Data	-.001	.257	.249	.948	.005	.280	.272	.934	.041	.328	.305	.938
Complete Data	-.297	.274	.264	.780	-.309	.301	.287	.780	-.236	.350	.320	.828
SWE- $\pi(\Delta)$	.005	.293	.282	.952	.020	.330	.312	.954	.071	.389	.358	.936
SWE - $\hat{\pi}(\Delta)$	.006	.293	.281	.952	.022	.330	.311	.946	.073	.389	.357	.936
SWE - $\hat{\pi}(V, \Delta)$	.006	.292	.280	.952	.021	.328	.310	.942	.072	.389	.357	.940
SWE - $\hat{\pi}(V, \Delta, X_0)$	-.002	.277	.268	.948	.010	.308	.293	.938	.058	.362	.337	.946
SRWE- $\pi(\Delta)$	.005	.293	.282	.954	.020	.330	.313	.952	.071	.390	.358	.936
SWE - $\hat{\pi}(\Delta)$	.006	.294	.281	.950	.021	.330	.312	.944	.073	.390	.358	.936
SWE - $\hat{\pi}(V, \Delta)$	.006	.292	.281	.952	.021	.328	.311	.942	.072	.389	.358	.940
SWE - $\hat{\pi}(V, \Delta, X_0)$	-.002	.278	.276	.956	.010	.309	.300	.946	.058	.363	.350	.954
AWE- $\pi(\Delta)$	-.004	.270	.260	.952	.005	.298	.286	.938	.051	.355	.328	.946
SWE - $\hat{\pi}(\Delta)$	-.004	.271	.261	.948	.005	.299	.287	.940	.052	.356	.328	.942
SWE - $\hat{\pi}(V, \Delta)$	-.004	.271	.261	.948	.005	.299	.287	.938	.052	.355	.329	.944
SWE - $\hat{\pi}(V, \Delta, X_0)$	-.003	.271	.261	.952	.005	.299	.287	.938	.052	.356	.329	.946
ARWE- $\pi(\Delta)$	-.004	.271	.261	.952	.005	.298	.287	.936	.052	.356	.329	.944
SWE - $\hat{\pi}(\Delta)$	-.004	.271	.261	.948	.005	.299	.287	.940	.052	.356	.329	.942
SWE - $\hat{\pi}(V, \Delta)$	-.004	.271	.261	.948	.005	.299	.287	.938	.052	.356	.329	.944
SWE - $\hat{\pi}(V, \Delta, X_0)$	-.003	.272	.261	.952	.005	.300	.287	.938	.052	.357	.330	.946
$\beta_2(t)$												
	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP
Full Data	-.003	.127	.127	.960	-.019	.151	.143	.944	-.011	.177	.164	.930
Complete Data	.179	.134	.134	.712	.172	.158	.149	.742	.156	.182	.171	.798

	$t = 0.3$						$t = 0.5$						$t = 0.7$					
							$\beta(t)$											
	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP	B	SD
SWE- $\pi(\Delta)$	-.006	.141	.141	.960	-.025	.172	.162	.946	-.029	.208	.195	.924	-.029	.208	.195	.924	-.029	.208
SWE - $\hat{\pi}(\Delta)$	-.007	.142	.140	.948	-.025	.172	.161	.944	-.030	.209	.194	.926	-.030	.209	.194	.926	-.030	.209
SWE - $\hat{\pi}(V, \Delta)$	-.006	.141	.140	.950	-.025	.171	.160	.942	-.029	.209	.194	.924	-.029	.209	.194	.924	-.029	.209
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.008	.140	.139	.952	-.026	.171	.159	.942	-.031	.209	.193	.920	-.031	.209	.193	.920	-.031	.209
SRWE- $\pi(\Delta)$	-.006	.141	.141	.960	-.024	.172	.162	.946	-.029	.209	.195	.926	-.029	.209	.195	.926	-.029	.209
SWE - $\hat{\pi}(\Delta)$	-.007	.142	.140	.948	-.025	.173	.161	.944	-.030	.210	.195	.928	-.030	.210	.195	.928	-.030	.210
SWE - $\hat{\pi}(V, \Delta)$	-.006	.141	.140	.950	-.025	.172	.161	.946	-.029	.210	.195	.928	-.029	.210	.195	.928	-.029	.210
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.008	.140	.139	.952	-.026	.171	.160	.944	-.031	.210	.195	.920	-.031	.210	.195	.920	-.031	.210
AWE- $\pi(\Delta)$	-.009	.140	.142	.956	-.028	.170	.165	.954	-.032	.207	.203	.944	-.032	.207	.203	.944	-.032	.207
SWE - $\hat{\pi}(\Delta)$	-.009	.140	.142	.960	-.028	.170	.165	.948	-.032	.208	.203	.940	-.032	.208	.203	.940	-.032	.208
SWE - $\hat{\pi}(V, \Delta)$	-.009	.140	.142	.956	-.028	.170	.165	.950	-.032	.208	.204	.940	-.032	.208	.204	.940	-.032	.208
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.009	.140	.142	.956	-.027	.171	.165	.952	-.032	.210	.204	.940	-.032	.210	.204	.940	-.032	.210
ARWE- $\pi(\Delta)$	-.009	.140	.142	.956	-.027	.170	.165	.956	-.032	.208	.204	.944	-.032	.208	.204	.944	-.032	.208
SWE - $\hat{\pi}(\Delta)$	-.009	.140	.142	.958	-.028	.171	.165	.948	-.032	.209	.204	.942	-.032	.209	.204	.942	-.032	.209
SWE - $\hat{\pi}(V, \Delta)$	-.009	.140	.142	.956	-.028	.171	.165	.950	-.032	.209	.204	.938	-.032	.209	.204	.938	-.032	.209
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.009	.141	.142	.956	-.027	.171	.166	.954	-.032	.211	.205	.938	-.032	.211	.205	.938	-.032	.211

Table III

Simulation results for $n = 250$ under case II. $\lambda(t) = \lambda_0(t) \exp\{\beta_1(t)X_o + \beta_2(t)X_m\}$ and $\pi = [1 + \exp\{0.8 - 2\Delta - V\}]^{-1}$. B, bias; SD, empirical standard deviation across simulated data sets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

	$t = 0.3$					$t = 0.5$					$t = 0.7$				
						$\beta_1(t)$									
	B	SD	SE	CP		B	SD	SE	CP		B	SD	SE	CP	
Full Data	-.037	.308	.302	.940		.021	.325	.330	.954		.067	.372	.373	.964	
Complete Data	-.208	.344	.337	.888		-.157	.362	.360	.920		-.076	.409	.402	.942	
SWE- $\pi(V, \Delta)$	-.033	.362	.350	.942		.032	.387	.380	.966		.097	.447	.433	.952	
SWE- $\hat{\pi}(\Delta)$	-.175	.347	.339	.906		-.130	.365	.363	.932		-.060	.413	.405	.942	
SWE- $\hat{\pi}(V, \Delta)$	-.036	.362	.348	.938		.027	.385	.378	.970		.094	.444	.430	.948	
SWE- $\hat{\pi}(V, \Delta, X_o)$	-.045	.333	.327	.948		.020	.357	.357	.962		.085	.407	.402	.952	
SRWE- $\pi(V, \Delta)$	-.033	.363	.351	.944		.032	.388	.381	.968		.097	.448	.434	.950	
SWE- $\hat{\pi}(\Delta)$	-.175	.347	.340	.906		-.130	.365	.363	.932		-.061	.413	.405	.942	
SWE- $\hat{\pi}(V, \Delta)$	-.036	.362	.349	.942		.027	.386	.379	.968		.093	.445	.432	.948	
SWE- $\hat{\pi}(V, \Delta, X_o)$	-.045	.333	.338	.954		.020	.358	.367	.966		.084	.408	.413	.960	
AWE- $\pi(V, \Delta)$	-.035	.323	.316	.944		.027	.351	.347	.960		.085	.407	.398	.960	
SWE- $\hat{\pi}(\Delta)$	-.042	.318	.343	.960		-.004	.341	.375	.976		.014	.388	.429	.980	
SWE- $\hat{\pi}(V, \Delta)$	-.034	.324	.316	.944		.027	.351	.347	.960		.085	.409	.400	.956	
SWE- $\hat{\pi}(V, \Delta, X_o)$	-.032	.323	.316	.942		.028	.351	.348	.958		.084	.408	.400	.958	
ARWE- $\pi(V, \Delta)$	-.035	.324	.316	.944		.027	.352	.348	.960		.084	.408	.399	.960	
SWE- $\hat{\pi}(\Delta)$	-.042	.318	.343	.960		-.004	.341	.375	.976		.014	.389	.430	.980	
SWE- $\hat{\pi}(V, \Delta)$	-.034	.324	.316	.944		.027	.352	.348	.962		.084	.409	.400	.954	
SWE- $\hat{\pi}(V, \Delta, X_o)$	-.032	.323	.316	.942		.028	.352	.348	.958		.084	.409	.401	.954	
$\beta_2(t)$															
	B	SD	SE	CP		B	SD	SE	CP		B	SD	SE	CP	
Full Data	-.014	.164	.155	.930		-.020	.193	.172	.920		-.031	.212	.199	.942	
Complete Data	.080	.188	.173	.890		.078	.213	.187	.876		.046	.233	.213	.916	

	$t = 0.3$						$t = 0.5$						$t = 0.7$					
							$\beta(t)$											
	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP	B	SD
SWE- $\pi(V, \Delta)$	-.024	.195	.177	.926	-.035	.226	.197	.918	-.059	.255	.232	.920	-.059	.255	.232	.920	-.059	.255
SWE - $\hat{\pi}(\Delta)$	.056	.189	.173	.898	.057	.215	.189	.888	.034	.236	.215	.916	.034	.236	.215	.916	.034	.236
SWE - $\hat{\pi}(V, \Delta)$	-.023	.194	.176	.932	-.032	.224	.195	.924	-.057	.255	.230	.920	-.057	.255	.230	.920	-.057	.255
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.025	.193	.175	.926	-.036	.224	.194	.922	-.061	.253	.229	.914	-.061	.253	.229	.914	-.061	.253
SRWE- $\pi(V, \Delta)$	-.024	.196	.178	.926	-.035	.226	.197	.922	-.059	.255	.232	.922	-.059	.255	.232	.922	-.059	.255
SWE - $\hat{\pi}(\Delta)$	.056	.189	.174	.898	.057	.216	.189	.890	.035	.236	.216	.916	.035	.236	.216	.916	.035	.236
SWE - $\hat{\pi}(V, \Delta)$	-.023	.194	.177	.932	-.032	.225	.196	.924	-.057	.255	.231	.922	-.057	.255	.231	.922	-.057	.255
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.025	.193	.176	.928	-.035	.224	.195	.928	-.061	.254	.230	.920	-.061	.254	.230	.920	-.061	.254
AWE- $\pi(V, \Delta)$	-.026	.192	.178	.932	-.039	.222	.200	.928	-.064	.252	.241	.936	-.064	.252	.241	.936	-.064	.252
SWE - $\hat{\pi}(\Delta)$	-.033	.204	.204	.954	-.014	.234	.235	.966	.010	.251	.299	.974	.010	.251	.299	.974	.010	.251
SWE - $\hat{\pi}(V, \Delta)$	-.026	.193	.178	.934	-.038	.224	.201	.940	-.062	.254	.243	.940	-.062	.254	.243	.940	-.062	.254
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.026	.194	.178	.934	-.039	.225	.201	.936	-.063	.254	.243	.942	-.063	.254	.243	.942	-.063	.254
ARWE- $\pi(V, \Delta)$	-.026	.192	.178	.932	-.038	.223	.200	.930	-.063	.253	.241	.936	-.063	.253	.241	.936	-.063	.253
SWE - $\hat{\pi}(\Delta)$	-.033	.204	.204	.954	-.014	.235	.235	.966	.010	.252	.300	.974	.010	.252	.300	.974	.010	.252
SWE - $\hat{\pi}(V, \Delta)$	-.025	.193	.178	.934	-.038	.224	.202	.940	-.062	.254	.243	.940	-.062	.254	.243	.940	-.062	.254
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.026	.194	.179	.934	-.039	.226	.202	.936	-.062	.255	.244	.944	-.062	.255	.244	.944	-.062	.255

Table IV

Simulation results for $n = 500$ under case IV. $\lambda(t) = \lambda_0(t) \exp\{\beta_1(t)X_o + \beta_2(t)X_m\}$ and $\pi = 0.9I(X_o = 0) + 0.4I(X_o = 1)$. B, bias; SD, empirical standard deviation across simulated data sets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

	$t = .3$				$t = .5$				$t = .7$			
	$\beta_1(t)$				$\beta_2(t)$				$\beta_2(t)$			
	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP
Full Data	-.001	.257	.249	.948	.005	.280	.272	.934	.041	.328	.305	.938
Complete Data	-.020	.321	.315	.944	.012	.367	.339	.940	.064	.418	.382	.950
SWE- $\pi(X_o)$	-.018	.322	.315	.944	.015	.369	.340	.942	.069	.424	.384	.946
SWE - $\hat{\pi}(X_o)$	-.018	.322	.315	.944	.016	.369	.340	.940	.070	.425	.385	.946
SWE - $\hat{\pi}(\Delta, X_o)$	-.018	.317	.310	.948	.016	.363	.334	.934	.070	.413	.374	.942
SWE - $\hat{\pi}(V, X_o)$	-.013	.304	.298	.950	.015	.349	.322	.932	.067	.406	.371	.938
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.002	.289	.289	.956	.018	.338	.316	.928	.064	.395	.361	.940
SRWE- $\pi(X_o)$	-.018	.322	.315	.944	.015	.369	.340	.942	.069	.424	.384	.948
SWE - $\hat{\pi}(X_o)$	-.018	.322	.315	.944	.016	.369	.340	.942	.069	.424	.384	.948
SWE - $\hat{\pi}(\Delta, X_o)$	-.018	.317	.312	.948	.016	.362	.335	.934	.070	.413	.375	.942
SWE - $\hat{\pi}(V, X_o)$	-.013	.303	.298	.950	.015	.349	.322	.932	.067	.406	.369	.938
SWE - $\hat{\pi}(V, \Delta, X_o)$	-.002	.289	.291	.956	.018	.338	.316	.930	.064	.394	.363	.940
	$\beta_2(t)$				$\beta_2(t)$				$\beta_2(t)$			
	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP
Full Data	-.003	.127	.127	.960	-.019	.151	.143	.944	-.011	.177	.164	.930
Complete Data	-.015	.180	.172	.954	-.033	.219	.193	.920	-.031	.258	.223	.916
SWE- $\pi(X_o)$	-.020	.196	.183	.942	-.038	.236	.204	.910	-.036	.276	.235	.904
SWE - $\hat{\pi}(X_o)$	-.020	.197	.183	.944	-.039	.236	.205	.910	-.036	.277	.236	.902
SWE - $\hat{\pi}(\Delta, X_o)$	-.019	.196	.181	.936	-.038	.234	.202	.910	-.034	.274	.234	.906
SWE - $\hat{\pi}(V, X_o)$	-.021	.195	.181	.938	-.040	.233	.202	.910	-.037	.276	.235	.900

	$t = .3$						$t = .5$						$t = .7$					
	$\beta_1(t)$						$\beta_1(t)$						$\beta_1(t)$					
	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP	B	SD	SE	CP	B	SD
$SWE - \hat{\pi}(V, \Delta, X_o)$	-.021	.196	.180	.924	-.040	.231	.200	.914	-.038	.273	.233	.906	-.038	.273	.233	.906	-.038	.273
$SRWE - \pi(X_o)$	-.020	.196	.183	.944	-.038	.236	.204	.910	-.035	.276	.235	.906	-.035	.276	.235	.906	-.035	.276
$SWE - \hat{\pi}(X_o)$	-.020	.197	.183	.944	-.039	.236	.204	.910	-.036	.276	.236	.902	-.036	.276	.236	.902	-.036	.276
$SWE - \hat{\pi}(\Delta, X_o)$	-.019	.196	.181	.936	-.037	.234	.202	.912	-.034	.274	.234	.906	-.034	.274	.234	.906	-.034	.274
$SWE - \hat{\pi}(V, X_o)$	-.021	.195	.181	.938	-.040	.233	.202	.910	-.037	.276	.235	.900	-.037	.276	.235	.900	-.037	.276
$SWE - \hat{\pi}(V, \Delta, X_o)$	-.021	.195	.179	.922	-.040	.231	.200	.914	-.038	.273	.232	.906	-.038	.273	.232	.906	-.038	.273