ADJUSTING FOR DEPENDENT CENSORING USING MANY COVARIATES

Donglin Zeng *

September 23, 2002

Abstract

Right-censored data are common in many epidemiological studies. One main goal is to estimate the survival function of lifetime. However, if this right-censoring is dependent and is explained by high-dimensional covariates, estimating the survival function of lifetime by using either semiparametric models or nonparametric methods can be problematic. In this paper, we condense these high-dimensional covariates through two working models for both lifetime and censoring time; therefore, an estimator of the survival function can be derived nonparametrically. We show that this estimator has the following properties: when either working model is correct, the estimator is consistent and asymptotically Gaussian; when both models are correct, its asymptotic variance attains the generalized Cramér-Rao bound. Simulations using small samples are performed to verify the robustness of the proposed estimator.

KEY WORDS: dependent censoring, double robustness, semiparametric efficiency.

* Donglin Zeng is Assistant Professor with Department of Biostatistics, University of North Carolina at Chapel Hill, NC 27599. Research is supported in part by NIDA Grant 2P50 DA 10075 and NSF-DMS 9802885.
1 Introduction.

Right-censored data are common in many epidemiological studies. Often, one main goal is to estimate the survival function of lifetime $T$, which can be defined as the time from the entry to the death of a patient in one study or which may be defined as the time from the entry to the occurrence/re-occurrence of a disease in another study. $T$ can only be observed for the patients whose events have already occurred during the study; for those who dropped out or those who had not experienced the events by the end of the study, their drop-out times or the ending time of the study, which are denoted as $C$, are observed instead of $T$. $C$ is called censoring time. Additionally, the patients’ information such as treatments they received, their demographic information, and much other auxiliary information are also collected and entered into right-censored data as covariates. These covariates are denoted as $L$ in this paper.

The Kaplan-Meier estimator is a standard estimator for the survival function of $T$. However, it is well-known that, when $T$ and $C$ are dependent, this estimator is inconsistent. If the dependence between $T$ and $C$ is explained by the covariates $L$, an intuitive approach to estimate the survival function of $T$ is to estimate the conditional distribution of $T$ given $L$ using a semiparametric model or via nonparametric estimation then to average this conditional distribution over all the observed covariates. For instance, a semiparametric model for the conditional distribution of $T$ given $L$ may be a Cox’s proportional hazard model (Cox (1972)), or a proportional odds model (Bennett (1983)); nonparametric estimation of this distribution may be using a local likelihood function (Tibshirani and Hastie (1987)), or using counting process (Aalen (1975), Beran (1981)).

However, the above intuitive approaches can be problematic when the covariates, $L$, is high-dimensional. This is because that when $L$ is high-dimensional (for instance, $L$ contains at least 3 continuous variables), nonparametric estimation of the distribution of $T$ given $L$ is impossible in moderate sample due to the curse of dimensionality; while any semiparametric model used for $T$ given $L$ may be easily misspecified: for example, if a Cox’s proportional hazard model is assumed for $T$ given $L$, we may hesitate about what interactions among $L$’s should also be included in the model since the number of possibilities is overwhelming (e.g., if the dimension of $L$ is 10, the number of their two-way interactions is 50). Using a misspecified model for $T$ given $L$ inevitably brings bias to the estimation of the distribution of $T$ given $L$, which is a nuisance parameter for the parameter of the survival function of $T$. So the estimator for the survival function of $T$ may also be inconsistent.

The situation that $L$ is high-dimensional is very common in many epidemiological studies. $L$ often contains much other auxiliary information which researchers are not primarily interested in but which
is useful for either predicting patients’ survival probability or explaining why patients dropped out or both. For example, during the study, patients might be asked about how they felt about participation and their answers were input into the data, because that if a patient did not feel the participation was useful, then it might imply that either he/she was potentially healthy, or he/she was likely to leave the study; $L$ might also include the patients’ accessibility to hospitals, because that the more difficult the accessibility to hospitals was, the more likely patients would discontinue participation; information about patients’ social support from either their families or their communities, as well as patients’ genetic types or family disease histories, might also be collected due to the possible influence these variables had on patients’ participation or lifetime.

To reduce the limitation of either semiparametric or nonparametric estimation in the above intuitive approaches, an innovative way is proposed in this paper in deriving an estimator for the survival function of the lifetime $T$. Especially, two working models are supposed for both the lifetime $T$ and the censoring time $C$ given all the covariates $L$. Then two-dimensional condensed information of $L$ are extracted from the working models and they are used as the new covariates in place of $L$. The estimator of the survival function is obtained by maximizing a pseudo-likelihood function. Under the assumption that the dependence between $T$ and $C$ is fully explained by $L$, it is shown that if either working model is correct, the estimator of the marginal survival function is consistent and asymptotically Gaussian; if both working models are correct, the asymptotic variance of the estimator attains the generalized Cramér-Rao bound of the full model space (cf. Bickel, Klaassen, Ritov and Wellner (1993)). The first property is named “double robustness” by Robins et al. (2000), since the estimator remains consistent if one working model is misspecified but the other one is correct. These properties can be summarized in Table 1, in which $Model(T|L)$ denotes the model of $T$ given $L$ while $Model(C|L)$ denotes the model of $C$ given $L$. Recall that in semiparametric estimation of the intuitive approaches described previously, the consistency requires $Model(T|L)$ to be correct. The advantages of our approach proposed in this paper are obvious: our estimator not only is consistent when the model for the lifetime is correct; but can be consistent even when the model for the lifetime is misspecified.

This paper is organized as follows: in Section 2, we first illustrate our ideas and approach using

| $Model(C|L)$ correct | $Model(T|L)$ correct | $Model(T|L)$ wrong |
|----------------------|----------------------|-------------------|
| consistent & C-R bound attained | consistent & C-R bound attained | consistent & C-R bound attained |
| consistent | inconsistent | inconsistent |

Table 1:
A simple example of estimating mean response subject to informative missingness; we then list the
detailed steps of estimating the survival function using right-censored data in Section 3; the asymptotic
properties of our estimator are given in Section 4; Section 5 provides an algorithm of estimating the
estimator’s asymptotic variance; some simulation results are reported in Section 6; finally, the paper
concludes with discussions and future work. Most of proofs are deferred to the appendix at the end
of the paper.

2 An Illustrative Example

For the purpose of illustration, we first take a look at an example of estimating mean response when
some responses are missing.

A common complication in many experimental studies is that some subjects leave the study pre-
maturely, i.e., they drop out of the study. In this case, any responses that were to be collected after the
person drops out are not observed. In interpreting the phenomenon of the dropout, researchers usually
collect a large amount of auxiliary information in the hope that, conditional on this auxiliary infor-
mation, the dropout will be random. That is, conditional on the auxiliary information, the random
variable indicating the status of the dropout will be independent of the future (unobserved) responses.

Let \( Y \) denote the response. Although the complete data include the response \( Y \) and high-dimensional
covariates (denoted as \( L \)), \( Y \) can be observed only if subjects are not missing cases. Let \( R \) be the
indicator of missing status (\( R \) takes either 0 or 1 and \( R = 0 \) indicates the missing case and vice versa).

Then the observed incomplete data consist of \((R_i, R_i Y_i, L_i), i = 1, ..., n\). As we have already assumed
that a lack of information is explained by \( L \), we suppose that \( R \) and \( Y \) are independent given \( L \). Our
goal is to estimate the mean of the response, \( \mathbb{E}Y \).

We make two working models: we tentatively assume that the model for \( R \) given \( L \) is a generalized
linear model with logit link, i.e., \( \pi(R = 1|l) = \frac{e^{\gamma l}}{1 + e^{\gamma l}} \), where \( \pi(R = 1|l) \) is the probability of
\( R = 1 \) given \( L = l \). We also tentatively assume that the conditional density of \( Y \) given \( L = l \), denoted
as \( p(y|l) \), is a normal density with a mean \( \beta' l \) and a variance \( \sigma^2 \). Thus, we consider the two-dimensional
covariates \((\beta' L, \gamma' L)\) as the condensed information of \( L \).

At first, the two parameters \((\beta, \gamma)\) introduced in these working models need to be estimated: based
on the tentatively assumed generalized linear model for \( R \) given \( L \), we can obtain the estimator of \( \gamma \),
denoted as \( \hat{\gamma}_n \), by performing a logistic regression; or equivalently, by maximizing

\[
\prod_{i=1}^{n} \pi(R_i = 0|L_i)^{1-R_i} \pi(R_i = 1|L_i)^{R_i},
\]

3
where \[ \pi(R = 1|L) = e^{\gamma L}/(1 + e^{\gamma L}) \]. At the same time, the estimator \( \hat{\beta}_n \) of \( \beta \) can be acquired by normal regression using the complete observations; or equivalently, by maximizing the following function \( \prod_{i=1}^{n} p(Y_i|L_i)^{R_i} \), where \( p(Y|L) \) is equal to \( e^{-(Y-\beta L)^2/2\sigma^2}/\sqrt{2\pi}\sigma \). It can be easily shown (Appendix A) that there exist two constants \( \beta^* \) and \( \gamma^* \) such that \( \hat{\beta}_n \) and \( \hat{\gamma}_n \) converge to \( \beta^* \) and \( \gamma^* \) in probability, respectively. Moreover, if the working model for \( Y \) given \( L \) is correct, then \( \beta^* \) is the correct constant for the parameter \( \beta \) in the working model; if the working model for \( R \) given \( L \) is correct, then \( \gamma^* \) is the correct constant for the parameter \( \gamma \) in the working model. Furthermore, we obtain that, when either working model is correct, \( R \) and \( Y \) are independent given \( (\beta^* L, \gamma^* L) \) (Appendix A). That is, if either working model is correct, the two-dimensional measurements \( (\beta^* L, \gamma^* L) \) are sufficient to explain the dependence between \( Y \) and \( R \).

Hence, if supposing that \( \beta^* \) and \( \gamma^* \) are known beforehand, then we can replace the \( L \) observations with the observations of \( (\beta^* L, \gamma^* L) \). We thus condense the high-dimensional \( L \) and obtain a reduced data set \( (R_i, Y_i, (\beta^* L_i, \gamma^* L_i)), i = 1, ..., n. \) Since \( R \) and \( Y \) are independent given \( (\beta^* L, \gamma^* L) \), the observed log-likelihood function for this reduced data is

\[
\sum_{i=1}^{n} \{ R_i \ln p(Y_i|\beta^* L_i, \gamma^* L_i) + R_i \ln P(R_i = 1|\beta^* L_i, \gamma^* L_i) \\
+ (1 - R_i) \ln P(R_i = 0|\beta^* L_i, \gamma^* L_i) + f_{\beta^* L_i, \gamma^* L_i}(\beta^* L_i, \gamma^* L_i) \}.
\]

Therefore, nonparametric estimation of the distribution of \( Y \) given \( (\beta^* L, \gamma^* L) \) is feasible. One way is to estimate the conditional distribution of \( Y \) given \( (\beta^* L, \gamma^* L) = (z_1, z_2) \) by maximizing a local version of the above log-likelihood function

\[
\sum_{i=1}^{n} R_i K\left( \frac{\beta^* L_i - z_1}{a_n}, \frac{\gamma^* L_i - z_2}{a_n} \right) \ln p(Y_i|(\beta^* L, \gamma^* L) = (z_1, z_2)),
\]

where \( K(\cdot) \) is a symmetric kernel function in \( R^2 \) and \( a_n \) is a bandwidth. If denote \( Z^* = (\beta^* L, \gamma^* L) \), then maximizing the above function gives that the estimator of the conditional density of \( Y \) given \( Z^* = (z_1, z_2) \) is an empirical function at the observed \( Y_i, i = 1, ..., n \), and the mass at \( Y_i \) is

\[
\frac{R_i K\left( \frac{z_1 - z_1}{a_n}, \frac{z_2 - z_2}{a_n} \right)}{\sum_{k=1}^{n} R_k K\left( \frac{z_1 - z_1}{a_n}, \frac{z_2 - z_2}{a_n} \right)}.
\]

Hence, we can estimate \( EY \) by

\[
\frac{1}{n} \sum_{j=1}^{n} \left( \sum_{i=1}^{n} \frac{Y_i R_i K\left( \frac{z_1 - z_1}{a_n}, \frac{z_2 - z_2}{a_n} \right)}{\sum_{k=1}^{n} R_k K\left( \frac{z_1 - z_1}{a_n}, \frac{z_2 - z_2}{a_n} \right)} \right).
\]
However, since \( \beta^* \) and \( \gamma^* \) are unknown but they can be consistently estimated by \( \hat{\beta}_n \) and \( \hat{\gamma}_n \) respectively, we substitute \((\hat{\beta}_n, \hat{\gamma}_n)\) into the above equation. Thus, we obtain the final estimator for \( EY \)

\[
\hat{EY} = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{Y_i R_i K(\frac{\hat{Z}_j - \hat{Z}_i}{a_n})}{\sum_{k=1}^{n} R_k K(\frac{\hat{Z}_j - \hat{Z}_k}{a_n})},
\]

where \( \hat{Z}_i = (\hat{\beta}_n' L_i, \hat{\gamma}_n' L_i) \).

It can be shown that, under some regularity conditions, if either working model is correct, the above estimator of \( EY \) is consistent; if both working models are correct, the asymptotic variance of the estimator is the same as the Cramér-Rao bound. The proof is sketched in Appendix A.

Why does this estimator have the above properties? In the following, we will explain the intuitions behind from the view of the likelihood functions.

For simplicity of illustration, we assume both \( Y \) and \( L \) are discrete random variables. For any constant \((\beta, \gamma)\), let \( Z = (\beta' L, \gamma' L) \) and \( z = (\beta' l, \gamma' l) \). Also denote \( Z^* = (\beta'^* L, \gamma'^* L) \) and \( z^* = (\beta'^* l, \gamma'^* l) \), where \((\beta^*, \gamma^*)\) are the limits of \((\hat{\beta}_n, \hat{\gamma}_n)\). Then the likelihood function of \((R = r, RY = ry, L = l)\) can be written as

\[
P(R = r, RY = ry, L = l) = P(R = r, RY = ry, Z = z, L = l) = P(Z = z)P(Y = y | Z = z)^r P(R = 1 | Y = y, Z = z) P(L = l | R = 1, Y = y, Z = z)^r \left( \sum_y P(L = l | R = 0, Y = y, Z = z) P(R = 0 | Y = y, Z = z) P(Y = y | Z = z) \right)^{1-r}.
\]

The whole parameters consist of

\[
P(Z = z), P(Y = y | Z = z), P(R = 1 | Y = y, Z = z), P(L = l | R = r, Y = y, Z = z)
\]

and each parameter space of them is assumed to be distinct from one another.

For \( n \) i.i.d observations \((R_i, R_i Y_i, L_i)\), we denote

\[
lik_n(\beta, \gamma) = \prod_{i=1}^{n} P(Z = Z_i) P(Y = Y_i | Z = Z_i)^{R_i}.
\]

We thus have the following claims:

1. If the working model for \( Y \) given \( L \) is correct, i.e., \( \beta^* \) is the correct constant in the working model, then for any \( \gamma \), \( lik_n(\beta^*, \gamma) \) is a partial like likelihood function for the inference of \( EY \).
2. If the working model for \( R \) given \( L \) is correct, i.e., \( \gamma^* \) is the correct constant in the working model, then for any \( \beta \), \( \text{lik}_n(\beta, \gamma^*) \) is a partial likelihood function for the inference of \( EY \).

3. If both working models are correct, \( \text{lik}_n(\beta^*, \gamma^*) \) is a full likelihood function concerning the inference of \( EY \).

The truth of these claims can be justified from the following arguments: if either of the two working models is correct, i.e., either \( \beta = \beta^* \) or \( \gamma = \gamma^* \) is the correct constant, \( R \) and \( Y \) are independent given \( Z = (\beta' L, \gamma' L) \). Therefore, the likelihood function of \( (R = r, RY = ry, L = l) \) can be further simplified as

\[
P(Z = z)P(Y = y|Z = z)^r P(R = 1|Z = z)^r P(R = 0|Z = z)^{1-r} \\
P(L = l|R = 1, Y = y, Z = z)^r \\
|\sum_y P(L = l|R = 0, Y = y, Z = z)P(Y = y|Z = z)|^{1-r}.
\]

Since the parameter space for each of \( P(Z = z), P(Y = y|Z = z), P(R = 1|Z = z), \) and \( P(L = l|R = r, Y = y, Z = z) \) is distinct from one another, we conclude that, for \( n \) i.i.d observations \( (R_i, R_iY_i, L_i), i = 1, ..., n \),

\[
\text{lik}_n(\beta, \gamma) = \prod_{i=1}^n P(Z = (\beta' L_i, \gamma' L_i))P(Y_i|Z = (\beta' L_i, \gamma' L_i))^{R_i}
\]

is a partial likelihood function for \( EY \) (cf. Wong (1986)). Furthermore, if both working models are correct, i.e, both \( \beta^* \) and \( \gamma^* \) are the correct constants in the working models, then \( (Y, R) \) and \( L \) are also independent given \( Z^* \), since

\[
P(Y = y, R = r|L = l, Z^* = z^*) \\
= P(Y = y|L = l, Z^* = z^*)P(R = r|Y = y, L = l, Z^* = z^*) \\
= P(Y = y|Z^* = z^*)P(R = r|L = l, Z^* = z^*) \\
= P(Y = y|Z^* = z^*)P(R = r|Z^* = z^*).
\]

Consequently, the likelihood function of \( (R = r, RY = ry, L = l) \) can be expressed as

\[
P(Z^* = z^*)P(Y = y|Z^* = z^*)^r P(R = 1|Z^* = z^*)^r P(R = 0|Z^* = z^*)^{1-r} \\
P(L = l|Z^* = z^*)^r P(L = l|Z^* = z^*)^{1-r}.
\]

Hence, for \( n \) i.i.d observations \( (R_i, R_iY_i, L_i), i = 1, ..., n \), the function \( \text{lik}_n(\beta^*, \gamma^*) \) is in fact the full likelihood function for the inference concerning \( EY \).
Now, we further define an estimator of \( EY \) by maximizing \( \text{lik}_n(\beta, \gamma) \) using the similar procedure as we obtained \( \hat{EY} \), except that \( \beta \) and \( \gamma \) are assumed to be known. Such an estimator is denoted by \( \hat{Y}(\beta, \gamma) \). Clearly,

\[
\hat{Y}(\beta, \gamma) = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{Y_i R_i K\left(\frac{Z_j - Z_i}{a_n}\right)}{\sum_{k=1}^{n} R_k K\left(\frac{Z_j - Z_k}{a_n}\right)}.
\]

Corresponding to Claims (1-3), we can easily make the following conclusions.

C.1 If the working model for \( Y \) for given \( L \) is correct, then for any \( \gamma \), \( \hat{Y}(\beta^*, \gamma) \) essentially maximizes a partial likelihood function \( \text{lik}_n(\beta^*, \gamma) \). Thus, \( \hat{Y}(\beta^*, \gamma) \) is asymptotically consistent with \( EY \) (cf. Wong (1986)). Moreover, since \( \hat{Y}(\beta^*, \gamma) \) is consistent for any \( \gamma \), \( \hat{Y}(\beta^*, \gamma^*) \) has the same asymptotic distribution with \( \hat{Y}(\beta^*, \hat{\gamma}_n) \) for \( \hat{\gamma}_n \to P \gamma^* \). That is, whether \( \gamma^* \) is known or estimated consistently has no influence on the asymptotic variance of the corresponding estimator.

C.2 If the working model for \( R \) given \( L \) is correct, then for any \( \beta \), \( \hat{Y}(\beta, \gamma^*) \) essentially maximizes a partial likelihood function \( \text{lik}_n(\beta, \gamma^*) \). Thus, it is asymptotically consistent with \( EY \). Moreover, \( \hat{Y}(\beta^*, \gamma^*) \) has the same asymptotic distribution with \( \hat{Y}(\beta_n, \gamma^*) \) for \( \beta_n \to P \beta^* \). That is, whether \( \beta^* \) is known or estimated consistently has no influence on the asymptotic distribution.

C.3 If both working models are correct, then \( \hat{Y}(\beta^*, \gamma^*) \) essentially maximizes a full likelihood function. That is, at this point, it is exactly the maximum likelihood estimator for \( EY \) so it is consistent and its asymptotic variance is the same as the generalized Cramér-Rao bound.

Recall that our proposed estimator, \( \hat{EY} \), is the same as \( \hat{Y}(\hat{\beta}_n, \hat{\gamma}_n) \), where \( \hat{\beta}_n \) and \( \hat{\gamma}_n \) are both consistent estimators of \( \beta^* \) and \( \gamma^* \), respectively. We conclude from (C.1) and (C.2) that, when either working model is correct, \( \hat{EY} \) approximates \( \hat{Y}(\beta^*, \gamma^*) \) so it must be consistent as well. When both working models are correct, from (C.1) and (C.2), we know that \( \hat{EY} = \hat{Y}(\hat{\beta}_n, \hat{\gamma}_n) \) has the same distribution with \( \hat{Y}(\beta^*, \gamma^*) \) so \( \hat{EY} \)'s asymptotic variance is the same as the generalized Cramér-Rao bound by (C.3).

3 Estimating Marginal Survival Function

Right-censored observations from \( n \) patients are

\[
(Y_i = T_i \wedge C_i, R_i = I_{T_i \leq C_i, L_i}), i = 1, ..., n,
\]

where \( L \) is time-independent but may include at least three continuous covariates. We assume that \( T \) and \( C \) are independent given \( L \). The goal is to estimate the marginal survival function for \( T \), which is denoted by \( S(t) = P(T > t) \).
The observed likelihood function for \( n \) observations can be written as:

\[
\prod_{i=1}^{n} h_{T|L}(Y_i|L_i)^{R_i} e^{-H_{T|L}(Y_i|L_i)} h_{C|L}(Y_i|L_i)^{1-R_i} e^{-H_{C|L}(Y_i|L_i)} f_{L}(L_i),
\]

where \( h_{T|L}(.|L) \) and \( h_{C|L}(.|L) \) are the hazard rate functions for \( T \) and \( C \) given \( L \), respectively; \( H_{T|L}(.|L) \) and \( H_{C|L}(.|L) \) are their respective cumulative hazard functions.

Our estimation procedure consists of the following steps.

Step 1. We make two working models for both the lifetime \( T \) and the censoring time \( C \) given \( L \). Our working models for \( T \) given \( L \) and \( C \) given \( L \) are the Cox’s proportional hazard models (these hypothesized models are used in this paper; however, any other model consisting of low-dimensional functions of \( L \) can also be used.), i.e., we tentatively assume that

\[
h_{T|L}(y|l) = \lambda_T(y)e^{\beta l}, \quad h_{C|L}(y|l) = \lambda_C(y)e^{\gamma l},
\]

for some unknown functions \( \lambda_T(.) \), \( \lambda_C(.) \) and some parameters \( \beta, \gamma \).

Step 2. We derive the estimator of \((\beta, \gamma)\) by performing the Cox’s regressions; equivalently, we maximize the logarithm of a pseudo partial likelihood (in fact, it is the same as the Cox’s partial likelihood function if the working models are correct) given in the following

\[
\hat{L}_1^{(n)}(\beta) = \frac{1}{n} \sum_{i=1}^{n} [R_i (\beta L_i - \ln( \sum_{Y_j \geq Y_i} e^{\beta L_i} ))],
\]

\[
\hat{L}_2^{(n)}(\gamma) = \frac{1}{n} \sum_{i=1}^{n} [(1 - R_i) (\beta L_i - \ln( \sum_{Y_j \geq Y_i} e^{\gamma L_i} ))].
\]

Basically, we substitute \( h_{T|L}(y|l) \) and \( h_{C|L}(y|l) \) in the observed likelihood function by \( \lambda_T(y)e^{\beta l} \) and \( \lambda_C(y)e^{\gamma l} \) respectively then maximize over \( \lambda_T(y), \lambda_C(y) \), and \((\beta, \gamma)\). We denote the estimator of \((\beta, \gamma)\) as \((\hat{\beta}_n, \hat{\gamma}_n)\). It can be shown in the next section that there exist two constants \( \beta^* \) and \( \gamma^* \) such that \( \hat{\beta}_n \) and \( \hat{\gamma}_n \) converge to \( \beta^* \) and \( \gamma^* \) in probability, respectively. Moreover, if the working model for \( T \) given \( L \) is correct, \( \beta^* \) is the true vector of the regression coefficients; if the working model for \( C \) given \( L \) is correct, \( \gamma^* \) is the true vector of the regression coefficients.

Step 3. Acting as if the two constants \( \beta^* \) and \( \gamma^* \) were known, we obtain the estimator of the hazard rate function of \( T \) given \((\beta^* L, \gamma^* L)\). Denote \( Z^* = (\beta^* L, \gamma^* L) \). When one of our working models is right, it can be shown that \( T \) and \( C \) are independent given \( Z^* \) (Lemma 4.1). That is, the two-dimensional covariate \( Z^* \) is sufficient to explain the dependence between \( T \) and \( C \). Therefore, we can replace the covariates \( L \) by \( Z^* \) in the observations; we thus obtain a reduced data set

\[
(Y_i, R_i, Z^*_i = (\beta^* L_i, \gamma^* L_i)), i = 1, \ldots, n.
\]
Clearly, the likelihood function for this reduced data can be verified to be:

$$\prod_{i=1}^{n} h_{T|Z^*}(Y_i|Z_i^*)e^{-H_{T|Z^*}(Y_i|Z_i^*)}h_{C|Z^*}(Y_i|Z_i^*)1-R_i e^{-H_{C|Z^*}(Y_i|Z_i^*)}f_Z(Z_i^*)$$

where $h_{T|Z^*}(.|Z^*), h_{C|Z^*}(.|Z^*)$ are the hazard rate functions of $T$ and $C$ given $Z^*$, respectively and $H_{T|Z^*}(.|Z^*), H_{C|Z^*}(.|Z^*)$ are their corresponding cumulative hazard function. So intuitively, we can estimate $h_{T|Z^*}(y|z)$ nonparametrically by maximizing the above observed likelihood function. One technique to obtain an estimator for the hazard rate function of $T$ given $Z^* = z$ is to maximize a local version of the observed log-likelihood function, which is

$$\sum_{i=1}^{n} K\left(\frac{Z_i^*-z}{a_n}\right)(R_i \ln h_{T|Z^*}(Y_i|z) - H_{T|Z^*}(Y_i|z)),$$

where $K(., .)$ is a symmetric two-dimensional kernel function with its support in a unit disk and $a_n$ is a bandwidth to be chosen later. The maximizer for $h_{T|Z^*}(y|z)$ is given as an empirical function. It has point mass at each observed $Y_i$ and the mass is

$$\frac{R_i K\left(\frac{Z_i^*-z}{a_n}\right)}{\sum_{Y_j \geq Y_i} K\left(\frac{Z_j^*-z}{a_n}\right)}.$$

Step 4. Since the two constants $\beta^*$ and $\gamma^*$ are unknown but they can be consistently estimated by $\hat{\beta}_n$ and $\hat{\gamma}_n$, we substitute $(\beta^*, \gamma^*)$ with $(\hat{\beta}_n, \hat{\gamma}_n)$ in the estimator obtained in Step 3. Thus, we obtain the estimator for the hazard rate function

$$\hat{h}_{T|Z^*}(Y_i|z) = \frac{R_i K\left(\frac{\hat{Z}_i-z}{a_n}\right)}{\sum_{Y_j \geq Y_i} K\left(\frac{\hat{Z}_j-z}{a_n}\right)},$$

So the estimator for the cumulative hazard function is

$$\hat{H}_{T|Z^*}(t|z^*) = \sum_{Y_i \leq t} \hat{h}_{T|Z^*}(Y_i|z).$$

Here, $\hat{Z}_i = (\hat{\beta}_n L_i, \hat{\gamma}_n L_i)$. In addition, since $S_{T|Z^*}(t|z) = \prod_{s \leq t} (1 - H_{T|Z^*}([s]|z))$, the estimator of the conditional survival function of $T$ given $Z^* = z$ is given by

$$\hat{S}_{T|Z^*}(t|z) = \prod_{j=1}^{n} (1 - \frac{K\left(\frac{\hat{Z}_j-z}{a_n}\right)I_{Y_j \leq t}R_j}{\sum_{m=1}^{n} K\left(\frac{\hat{Z}_m-z}{a_n}\right)I_{Y_j \leq Y_m}}).$$

Step 5. The final estimator of the survival function $S(t)$ is simply the average of $\hat{S}_{T|Z^*}(t|z^*)$ over all the $\hat{Z}_i$’s, i.e.,

$$\hat{S}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \hat{S}_{T|Z^*}(t|\hat{Z}_i) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{n} (1 - \frac{K\left(\frac{\hat{Z}_j-\hat{Z}_i}{a_n}\right)I_{Y_j \leq t}R_j}{\sum_{m=1}^{n} K\left(\frac{\hat{Z}_m-\hat{Z}_i}{a_n}\right)I_{Y_j \leq Y_m}}).$$
4 Main Results

We need the following assumptions.

(A1.) $T$ and $C$ are independent given $L$.

(A2.) The joint density of $(T, C, L)$ is continuously twice-differentiable inside their support. Moreover, it has a positive lower bound in its support and its second derivatives are uniformly bounded.

(A3.) There exists an unknown constant $\theta$ such that

$$\inf_L P(T > \tau | L) > \theta > 0, \inf_L P(C = \tau | L) > \theta > 0 \text{ a.s.}$$

where $\tau = \max\{T \wedge C\}$.

(A4.) The kernel function $K(x_1, x_2)$ has a bounded support in the unit disk of $R^2$ and it is continuously twice differentiable in its support. Moreover, it satisfies

$$K(-x_1, -x_2) = K(x_1, x_2), \int x_j K(x_1, x_2) dx_j = 0,$$

$$|\nabla x_j K(x_1, x_2)| \leq CK(x_1, x_2), \text{ where } x_1^2 + x_2^2 < 1, j = 1, 2.$$

(A5.) $\frac{\ln a_n}{na_n} \to 0, na_n^2 \to \infty, na_n^4 \to 0$.

Our first result is the convergence of the two estimators $\hat{\beta}_n$ and $\hat{\gamma}_n$.

Theorem 4.1 Under Assumptions (A1)-(A5), there exist $\beta^*$ and $\gamma^*$ such that

$$\sqrt{n}(\hat{\beta}_n - \beta^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_{\beta}(\beta^*, Y_i, R_i, L_i) + o_p(1),$$

$$\sqrt{n}(\hat{\gamma}_n - \gamma^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_{\gamma}(\gamma^*, Y_i, R_i, L_i) + o_p(1),$$

for some influence functions $S_{\beta}$ and $S_{\gamma}$. Thus, both $\sqrt{n}(\hat{\beta}_n - \beta^*)$ and $\sqrt{n}(\hat{\gamma}_n - \gamma^*)$ weakly converge to some multinormal distributions.

Theorem 4.1 shows $(\hat{\beta}_n, \hat{\gamma}_n)$ converge to some constants even if using the Cox’s proportional hazard models as working models is wrong. Obviously, if the model of $T$ given $L$ is correct, $\beta^*$ is the correct constant in this model; if the model of $C$ given $L$ is correct, $\gamma^*$ is the correct constant in this model. Moreover, the following lemma says, if either working model is correct, these condensed variables $(\beta^* L, \gamma^* L)$ are sufficient to explain the dependence between the lifetime and the censoring time.

Lemma 4.1 Suppose either of the working models to be right, i.e., the distribution of $T$ given $L$ is given by a Cox’s proportional hazard model or the distribution of $C$ given $L$ is given by another Cox’s
proportional hazard model. Let $Z^*$ be a vector function of $L$ consisting of $(\beta^* L, \gamma^* L)$, where $\beta^*$ and $\gamma^*$ are the constants in the previous theorem. Then $T$ and $C$ are independent given $Z^*$, and moreover,

$$H_{T|Z^*}(t|z) = -\int_0^t \frac{d_u P(T \wedge C > u, R = 1|Z^* = z)}{P(T \wedge C \geq u|Z^* = z)}.$$

**Proof.** We only prove the results hold if the censoring time $C$ depends on $L$ only through $\gamma^* L$. The proof of the other situation is similar. Since $C$ is fully explained by $\gamma^* L$,

$$P(T < t_1, C < t_2|Z^*) = E_{L|Z^*}[P(T < t_1, C < t_2|L)] = E_{L|Z^*}[P(T < t_1|L)P(C < t_2|L)] = P(C < t_2|Z^*)E_{L|Z^*}[P(T < t_1|L)] = P(C < t_2|Z^*)P(T < t_1|Z^*).$$

Therefore,

$$-\int_0^t \frac{d_u P(T \wedge C > u, R = 1|Z^* = z)}{P(T \wedge C \geq u|Z^* = z)}$$

$$= -\int_0^t \frac{d_u P(C > T|Z^* = z) - P(u > T|Z^* = z)}{P(C \geq u, T \geq u|Z^* = z)}dF_C(c)du$$

$$= \int_0^t P(C > u|Z^* = z)d_u P(T < u|Z^* = z)$$

$$= \int_0^t P(T \geq u|Z^* = z)du$$

$$= H_{T|Z^*}(t|z).$$

We then obtain the asymptotic result for $\hat{S}_n(t)$.

**Theorem 4.2** (Consistency and Asymptotic Normality of $\hat{S}_n(t)$) Under Assumptions (A1)-(A5), if either of the two working models is correct, i.e., either the model for $T$ given $L$ is a Cox’s proportional hazard model or the model for $C$ given $L$ is a Cox’s proportional hazard model,

$$\sqrt{n}(\hat{S}_n(.) - S(.) ) \Rightarrow G(\cdot),$$

where the weak convergence is in $L^\infty([0, \tau])$ space and $G(\cdot)$ is a Gaussian process with covariance

$$r(s, t) = E[A(s; Y, R, L)A(t; Y, R, L)].$$

The function $A(\cdot)$ is given as

$$A(t; Y, R, L)$$
\[ e^{-H_{T|Z^*(t)|Z^*}} - S(t) \]
\[ = -RI_{Y} \leq \{ e^{-H_{T|Z^*(t)|Z^*}} + H_{C|Z^*(Y|Z^*)} - H_{T|Z^*(t)|Z^*} \} \]
\[ + \int_{0}^{t} e^{-H_{T|Z^*(t)|Z^*}} d_{u} H_{T|Z^*(u)|Z^*} \]
\[ -E[ e^{-H_{T|Z^*(t)|Z^*}} \frac{d}{d\gamma} \mid \gamma = \gamma^* \int_{0}^{t} d_{u} P(Y \leq u, R = 1|\gamma^*, L, \beta^* L) \] \[ S_{\gamma} (\gamma^*, Y, R, L) \]
\[ -E[ e^{-H_{T|Z^*(t)|Z^*}} \frac{d}{d\beta} \mid \beta = \beta^* \int_{0}^{t} d_{u} P(T \leq u, R = 1|\beta^* L) / P(T \leq u|\beta^* L) ^{\gamma^*} \] \[ S_{\beta} (\beta^*, Y, R, L), \]

where \( Y = T \land C, Z^* = (\beta^* L, \gamma^* L) \), and \( S_{\beta} (\beta^*, Y, R, L), S_{\gamma} (\gamma^*, Y, R, L) \) are the influence functions in estimating \( \beta^* \) and \( \gamma^* \) in Theorem 4.1. Furthermore, for the last two terms in the above expression, the first term is zero when the working model for \( T \) given \( L \) is true; the second term is zero when the working model for \( C \) given \( L \) is true.

**Remark** Theorem 4.2 concludes that the estimator, \( \hat{S}_n(t) \), is a consistent estimator of \( S(t) \) when either working model is correct. To fulfill its assumptions, one choice of \( K(x_1, x_2) \) can be
\[ K(x_1, x_2) = e^{-(x_1^2 + x_2^2)} I_{x_1^2 + x_2^2 \leq 1} \]
and the choice of the bandwidth \( a_n \) is \( a_n = Mn^{-1/3} \) for some constant \( M \).

The next theorem shows that if both working models are correct, then the asymptotic variance of \( \hat{S}_n(t) \) attains the efficiency bound.

**Theorem 4.3** (Efficiency of \( \hat{S}_n(t) \)) Under Assumptions (A1)-(A5), when the model of the lifetime \( T \) given the covariates \( L \) and the model of \( C \) given the covariates \( L \) are both the Cox’s proportional hazard models, the asymptotic variance of \( \hat{S}_n(t) \) is the same as the generalized Cramér-Rao bound of the full model space for any fixed \( t \in [0, \tau] \).

**Proof.** In Theorem 4.2, we have seen \( \hat{S}_n(t) \) is an asymptotic linear estimator with influence function \( A(t; L, Y, R) \). To prove the efficiency, it is sufficient to verify this influence function is the same as the efficient influence function of the full model space when the underlying models of the lifetime given the covariates and the censoring time given the covariates happen to be coincident with the Cox’s proportional hazard models. In other words, we need to verify that the influence function \( A(t; L, Y, R) \) is on the linear space spanned by the score functions.

For this purpose, from the result in Theorem 4.2, the last two terms of \( A(t; Y, R, L) \) vanish when both working models are correct. Hence, it is sufficient if we can show the remaining term in the influence function, which is
\[ e^{-H_{T|Z^*(t)|Z^*}} - S(t) \]
In this section, we propose an algorithm to estimate the asymptotic variance for estimating a Fréchet differentiable functional of \( S(t) \), denoted as \( \Psi(S(t)) \). Here \( \Psi(\cdot) \) is a functional on \( BV[0, \tau] \) space. Such examples include the probability that one survives longer than \( t_0, S(t_0) \) \( (t_0 < \tau) \), the observed mean lifetime \( E[T|T \leq \tau] \), and etc.

Especially, we choose

\[
\tilde{h}(Y, L) = -h_{T|L}(Y|L)I_{Y \leq t}e^{H_{T|Z^*}(Y|Z^*)+H_{C|Z^*}(Y|Z^*)-H_{T|Z^*}(t|Z^*)} + \int_0^Y \tilde{h}(s, L)ds : E\tilde{h}^2(Y, L) < \infty.
\]

On the other hand, by differentiating the log-likelihood of \( (Y, R, L) \) with respect to the parameter \( h_{T|L}(Y|L) \) along the direction \( \tilde{h}(Y, L) \), it is easy to calculate that in full model space, the score functions for \( h_{T|L}(Y|L) \) are

\[
\{ \frac{R\tilde{h}(Y, L)}{h_{T|L}(Y|L)} - \int_0^Y \tilde{h}(s, L)ds : E\tilde{h}^2(Y, L) < \infty \}.
\]

Moreover, when both working models are correct, \( h_{T|L}(Y|L) = h_{T|Z^*}(Y|Z^*) \). So

\[
A(t; Y, R, L) = e^{-H_{T|Z^*}(t|Z^*)} - Ee^{-H_{T|Z^*}(t|Z^*)} + \frac{R\tilde{h}(Y, Z^*)}{h_{T|Z^*}(Y|Z^*)} - \int_0^Y \tilde{h}(s, Z^*)ds,
\]

which is on the tangent space spanned by the score functions. Therefore, this influence function must be the efficient influence function.

5 Variance Estimation

In this section, we propose an algorithm to estimate the asymptotic variance for estimating a Fréchet differentiable functional of \( S(t) \), denoted as \( \Psi(S(t)) \). Here \( \Psi(\cdot) \) is a functional on \( BV[0, \tau] \) space. Such examples include the probability that one survives longer than \( t_0, S(t_0) \) \( (t_0 < \tau) \), the observed mean lifetime \( E[T|T \leq \tau] \), and etc.

Obviously, a good estimator for \( \Psi(S(t)) \) is \( \Psi(\hat{S}_n(t)) \). To obtain the asymptotic distribution of \( \Psi(\hat{S}_n(t)) \), we need to consider the first derivative of \( \Psi(\cdot) \) at \( S(t) \), denoted by \( \Psi'(S(t))(\cdot) \), which is a linear functional on the tangent space of \( S(t) \). Often, we can write this linear functional in a form of

\[
\Psi'(S(t))(h) = \int_0^T h(t)d\psi(t)
\]

where \( h(\cdot) \) is any tangent vector for \( S(t) \) and \( \psi(t) \) is in \( BV[0, \tau] \). For example, for estimating the survival probability at \( t_0 \in [0, \tau] \), \( \psi(t) \) thus is given by a Heaviside function \( I_{t \geq t_0} \); for estimating \( E[T|T \leq \tau] \), simple calculation gives that

\[
\psi(t) = t/(1 - S(\tau)) - I_{t \geq \tau}(\tau - E[T|T \leq \tau])/(1 - S(\tau)).
\]
Therefore, if either working model is correct, by the delta method, \( \sqrt{n}(\hat{S}_n(t)) - \Psi(S(t)) \) has an asymptotically normal distribution, and its asymptotic variance is given by

\[
\sigma^2 = \int_0^\tau \int_0^\tau r(s,t) d\psi(s) d\psi(t),
\]

where \( r(s,t) \) is the covariance function given in Theorem 4.2, i.e.,

\[
r(s,t) = E[\mathcal{A}(s;Y,R,L)\mathcal{A}(t;Y,R,L)].
\]

The function \( \mathcal{A}(t;Y,R,L) \) is given as

\[
\mathcal{A}(t;Y,R,L) = e^{-H_{T|Z^*}(t|Z^*)} - S(t)
\]

\[
-RI_{Y<e^{H_{T|Z^*}(Y^*)}+H_{C|Z^*}(Y^*)-H_{T|Z^*}(t|Z^*)} + \int_0^t \int_0^\tau e^{H_{T|Z^*}(u|Z^*)+H_{C|Z^*}(u|Z^*)-H_{T|Z^*}(t|Z^*)} du H_{T|Z^*}(u|Z^*)
\]

\[
-\int_0^t \frac{d}{d\gamma} P(Y \land C \leq u, R = 1|\gamma', L, \beta'^* L) \big| S_\gamma(\gamma^*, Y, R, L)
\]

\[
-\int_0^t \frac{d}{d\beta} P(Y \land C \leq u, R = 1|\gamma'^* L, \beta L) \big| S_\beta(\beta^*, Y, R, L),
\]

where \( Y = T \land C, Z^* = (\beta'^* L, \gamma'^* L) \), and \( S_\beta(\beta^*, Y, R, L), S_\gamma(\gamma^*, Y, R, L) \) are the influence functions in estimating \( \beta^* \) and \( \gamma^* \) in Theorem 4.1.

One consistent estimator of \( \sigma^2 \) is given by

\[
\int_0^\tau \int_0^\tau \hat{r}(s,t) d\psi(s) d\psi(t),
\]

in which \( \hat{r}(s,t) = \mathbf{P}_n[\hat{A}(t;Y,R,L)\hat{A}(t;Y,R,L)] \) and \( \hat{A}(t;Y,R,L) \) is a consistent estimator of \( \mathcal{A}(t;Y,R,L) \).

To find such an \( \hat{A}(t;Y,R,L) \), we substitute \( H_{T|Z^*}(t|Z^*) \) and \( H_{C|Z^*}(t|Z^*) \) by their corresponding estimators \( \hat{H}_{T|Z^*}(t|\hat{Z}) \) and \( \hat{H}_{C|Z^*}(t|\hat{Z}) \) following Step 3 of Section 3; furthermore, according to the expressions in the proof of Theorem 4.1, we can consistently estimate the influence functions for \( \hat{\beta}_n \) and \( \hat{\gamma}_n \), which are \( S_\beta(\beta^*, Y, R, L) \) and \( S_\gamma(\gamma^*, Y, R, L) \), by \( \hat{S}(\hat{\beta}_n, Y, R, L) \) and \( \hat{S}(\hat{\gamma}_n, Y, R, L) \) respectively.

Here,

\[
\hat{S}_\beta(\hat{\beta}_n, y, r, l) \quad \text{(1)}
\]

\[
= -\left\{ \mathbf{P}_n\left[ \frac{R(\mathbf{P}_n[I_{Y \geq y}L'e^{\hat{\beta}_n L}]}{\mathbf{P}_n[I_{Y \geq y}e^{\hat{\beta}_n L}]^2} - \mathbf{P}_n[I_{Y \geq y}Le^{\hat{\beta}_n L}][\mathbf{P}_n[I_{Y \geq y}L'e^{\hat{\beta}_n L}]_{|y'=y}] \right] \right\}^{-1}
\]

\[
= \frac{-rl - r \mathbf{P}_n[I_{y \leq y}Le^{\hat{\beta}_n L}]}{\mathbf{P}_n[I_{y \leq y}e^{\hat{\beta}_n L}]} - le^{\hat{\beta}_n L}\mathbf{P}_n\left[ \frac{I_{y \leq y}}{\mathbf{P}_n[I_{y \leq y}e^{\hat{\beta}_n L}]_{|y'=y}} \right]
\]

\[
+ e^{\hat{\beta}_n L} \mathbf{P}_n\left[ \frac{I_{y \leq y}}{\mathbf{P}_n[I_{y \leq y}e^{\hat{\beta}_n L}]_{|y'=y}} \right]
\]

\[
+ e^{\hat{\beta}_n L}\mathbf{P}_n\left[ \frac{I_{y \leq y}}{\mathbf{P}_n[I_{y \leq y}e^{\hat{\beta}_n L}]_{|y'=y}} \right]
\]
\[ \hat{S}_n(\hat{\gamma}_n, y, r, t) = -\{P_n[1 - R(P_n[I_{Y \geq y} LL'e_{\hat{\gamma}_n}^L] - P_n[I_{Y > y'} LL'e_{\hat{\gamma}_n}^L]P_n[I_{Y > y'} LL'e_{\hat{\gamma}_n}^L])|y' = Y]\}^{-1} \]

\[ \{1 - r\}l - (1 - r) P_n[I_{Y \leq Y} e_{\hat{\gamma}_n}^L] - le_{\hat{\gamma}_n}^L P_n[(1 - R) P_n[I_{Y \leq y} e_{\hat{\gamma}_n}^L]|y' = Y] \]
\[ + e_{\hat{\gamma}_n}^L P\{1 - R\} P_n[I_{Y \leq y} e_{\hat{\gamma}_n}^L]|y' = Y\} \} \].

Additionally, if we can find two statistics, denoted by \( \hat{V}_{\gamma_n} \) and \( \hat{V}_{\beta_n} \), such that they are consistent estimators for

\[ -E[e^{-H_T|Z^t(t)Z^*}} \frac{d}{d\gamma}|_{\gamma = \gamma^*} \int_0^t \frac{du}{P(T \wedge C \leq u, R = 1|\gamma', L, \beta^* L)} \]

and

\[ -E[e^{-H_T|Z^t(t)Z^*}} \frac{d}{d\beta}|_{\beta = \beta^*} \int_0^t \frac{du}{P(T \wedge C \leq u|\gamma^* L, \beta^* L)} \]

respectively, then one consistent estimator for \( \hat{A}(t; Y, R, L) \) is given by

\[ \hat{A}(t; Y, R, L) = e^{-\hat{H}_T|Z^t(t)Z^*}} - \hat{S}_n(t) \]

\[ -R[I_{Y \leq Y} e^{\hat{H}_T|Z^t(t)Z^*} + \hat{H}_C|Z^t(t)Z^*}) - \hat{H}_T|Z^t(t)Z^*} \]
\[ + \int_0^t \frac{du}{P(T \wedge C \leq u, R = 1|\gamma^* L, \beta^* L)} \]
\[ + \hat{V}_{\gamma_n} S_{\gamma}(\hat{\gamma}_n, Y, R, L) + \hat{V}_{\beta_n} S_{\beta}(\hat{\beta}_n, Y, R, L). \]

The remaining thing is to find the estimators \( \hat{V}_{\beta_n} \) and \( \hat{V}_{\gamma_n} \). However, since in order to estimate

\[ -E[e^{-H_T|Z^t(t)Z^*}} \frac{d}{d\gamma}|_{\gamma = \gamma^*} \int_0^t \frac{du}{P(T \wedge C \leq u|\gamma^* L, \beta^* L)} \]

and

\[ -E[e^{-H_T|Z^t(t)Z^*}} \frac{d}{d\beta}|_{\beta = \beta^*} \int_0^t \frac{du}{P(T \wedge C \leq u|\gamma^* L, \beta^* L)} \],

it needs much effort to obtain some smooth estimators for \( P(T \wedge C \leq u, R = 1|\beta^* L, \gamma^* L) \) and \( P(T \wedge C \geq u|\beta^* L, \gamma^* L) \) and differentiate with respect to either \( \beta \) or \( \gamma \), we would rather estimate them by using numerical computation, which is given in the following lemma.

**Lemma 5.1** For any constants \((\beta, \gamma)\), we define an estimator of \( S(t) \), denoted by \( \hat{S}_n(t; \beta, \gamma) \) by repeating Steps (1-3) in Section 3 for fixed \( \beta \) and \( \gamma \). Let \( e_1, ..., e_{\text{dim}(\beta^*)} \) be the unit vectors in \( R^{\text{dim}(\beta^*)} \), i.e., \( e_i \) takes 1 at \( i \)th position while 0’s elsewhere. Similarly, let \( d_1, ..., d_{\text{dim}(\gamma^*)} \) be the unit vectors in \( R^{\text{dim}(\gamma^*)} \). Moreover, we select a constant \( \epsilon_n \) such that

\[ \epsilon_n \to 0, \epsilon_n = o(a_n), \sqrt{n} \epsilon_n \to \infty. \]
Then when one of the working models is correct, the consistent estimators, $\hat{V}_{\beta_n}$ and $\hat{V}_{\gamma_n}$ can be chosen as

$$
\hat{V}_{\beta_n} = \frac{1}{\epsilon_n} \left( \begin{array}{c}
\hat{S}_n(t; \beta_n + \epsilon_n e_1, \gamma_n) - \hat{S}_n(t) \\
\vdots \\
\hat{S}_n(t; \beta_n + \epsilon_n \epsilon_{dim(\beta^*)}, \gamma_n) - \hat{S}_n(t)
\end{array} \right),
$$

and

$$
\hat{V}_{\gamma_n} = \frac{1}{\epsilon_n} \left( \begin{array}{c}
\hat{S}_n(t; \beta_n, \gamma_n + \epsilon_n d_1) - \hat{S}_n(t) \\
\vdots \\
\hat{S}_n(t; \beta_n, \gamma_n + \epsilon_n \epsilon_{dim(\beta^*)}) - \hat{S}_n(t)
\end{array} \right).
$$

**Proof.** Obviously, our true estimator $\hat{S}_n(t)$ is the same as $\hat{S}_n(t; \beta_n, \gamma_n)$. By repeating the proof of Theorem 4.2, we can obtain that if

$$
|\beta - \beta^*|, |\gamma - \gamma^*| = o(a_n),
$$

then

$$
\hat{S}_n(t; \beta, \gamma) - S(t) = (P_n - P)[S_{t|Z^*}(t|Z^*) - S(t) - R I_{Y \leq 1} S_{t|Z^*}(t|Z^*) e^{H_{T|Z^*}(Y|Z^*)} + H_{C|Z^*}(Y|Z^*)] + o_P\left(\frac{1}{\sqrt{n}}\right),
$$

where $Z^* = (\beta^* L, \gamma^* L)$ and

$$
B(\beta, \gamma, z, t) = \int_0^t P(Y \leq u, R = 1| (\beta^* L, \gamma^* L) = z) \\
P(Y \geq u| (\beta^* L, \gamma^* L) = z).
$$

We especially choose $\gamma = \gamma_n$ and $\beta = \beta_n + \epsilon_n v$ where $v$ is any constant vector on $R^{dim(\beta^*)}$ with norm one. After linearizing the $B(\beta, \gamma, Z, t)$ around $\beta = \beta^*, \gamma = \gamma^*$ and applying the Donsker theorem, we find that

$$
\hat{S}_n(t; \beta_n + \epsilon_n v, \gamma_n) - S(t) = (P_n - P)[S_{t|Z^*}(t|Z^*) - S(t) - R I_{Y \leq 1} S_{t|Z^*}(t|Z^*) e^{H_{T|Z^*}(Y|Z^*)} + H_{C|Z^*}(Y|Z^*)] + o_P\left(\frac{1}{\sqrt{n}}\right).
$$
Moreover, the conclusions in the lemma hold.

Similarly, for any constant vector \( \tilde{v} \) in \( R^{\text{dim}(\gamma^*)} \) with norm 1,

\[
\frac{\hat{S}_n(t; \beta_n + \epsilon_n \tilde{v}, \hat{\gamma}_n) - \hat{S}_n(t)}{\epsilon_n} \to^p \mathbb{P}\{S_{T|Z^*}(t|Z^*) - B(\beta^*, \gamma^*, Z^*, t)\} \tilde{v}.
\]

So the conclusions in the lemma hold.

Hence, we obtain the following algorithm to estimate the asymptotic variance of any differential functional \( \Psi(S(t)) \):

1. obtain the estimators \( \hat{H}_{T|Z^*}(t|z), \hat{H}_{C|Z^*}(t|z), \hat{S}_n(t) \);

2. calculate \( \hat{S}_\beta(\hat{\beta}_n, y, r, l) \) and \( \hat{S}_\gamma(\hat{\gamma}_n, y, r, l) \) using the expressions (5.1) and (5.2);

3. for each \( e_i, d_j, i = 1, \ldots, \text{dim}(\beta^*), j = 1, \ldots, \text{dim}(\gamma^*) \), compute \( \hat{S}_n(t; \beta_n + \epsilon_n e_i, \hat{\gamma}_n) \) and \( \hat{S}_n(t; \beta_n, \hat{\gamma}_n + \epsilon_n d_j) \);

4. obtain

\[
\hat{V}_{\beta_n} = \frac{1}{\epsilon_n} \left( \begin{array}{c}
\hat{S}_n(t; \beta_n + \epsilon_n e_1, \hat{\gamma}_n) - \hat{S}_n(t) \\
\vdots \\
\hat{S}_n(t; \beta_n + \epsilon_n e_{\text{dim}(\beta^*)}, \hat{\gamma}_n) - \hat{S}_n(t)
\end{array} \right),
\]

and

\[
\mathbb{P}\{S_{T|Z^*}(t|Z^*) - H_{T|Z^*}(t|Z^*)\} = 0.
\]
\[ \hat{V}_{\hat{\gamma}_n} = \frac{1}{\epsilon_n} \begin{pmatrix} \hat{S}_n(t; \hat{\beta}_n, \hat{\gamma}_n + \epsilon_n d_1) - \hat{S}_n(t) \\ \vdots \\ \hat{S}_n(t; \hat{\beta}_n, \hat{\gamma}_n + \epsilon_n d_{\dim(j)}) - \hat{S}_n(t) \end{pmatrix}; \]

5. calculate

\[
\hat{A}(t; Y, R, L) = e^{-\hat{H}_T^*Z^*(t|\hat{Z})} - \hat{S}_n(t) \\
-RI_{Y \leq t}e^{\hat{H}_T^*Z^*(Y|\hat{Z}) + \hat{H}_C^*Z^*(Y|\hat{Z})} - \hat{H}_T^*Z^*(t|\hat{Z}) \\
+ \int_0^{t \wedge Y} e^{\hat{H}_T^*Z^*(u|\hat{Z}) + \hat{H}_C^*Z^*(u|\hat{Z})} d_u \hat{H}_T^*Z^*(u|\hat{Z}) \\
+ \hat{V}_{\hat{\gamma}_n} \hat{S}_{\gamma}(\hat{\gamma}_n, Y, R, L) + \hat{V}_{\hat{\beta}_n} \hat{S}_{\beta}(\hat{\beta}_n, Y, R, L); 
\]

6. the asymptotic variance of \( \Psi(\hat{S}_n(t)) \) can be estimated by

\[
\int_0^T \int_0^T \mathbf{P}_n[\hat{A}(t, Y, R, L)\hat{A}(s, Y, R, L)] d\psi(s) d\psi(t). 
\]

Remark One choice of \( a_n, \epsilon_n \) is \( a_n = Mn^{-1/3}, \epsilon_n = Mn^{-5/12} \) for some constant \( M \). In particular, to estimate the asymptotic variance of the survival probability at \( t = t_0 \), the last step of the above algorithm is simply computing \( \mathbf{P}_n[\hat{A}(t_0; Y, R, L)]^2 \). Except that the last step requires a numerical double integration, the computing time in the other steps is a linear order of the computing time for computing \( \hat{S}_n(t) \), which is \( O(n^3a_n^2) \). The storage in the computation is the same order as storing a \( n \times n \) numerical array.

6 Simulations

We have performed a number of simulations to evidence the advantages of our approach in small sample. In particular, two goals are achieved by these simulations. One goal is to verify the double robustness of the estimator from our approach. That is, if either of the working models is misspecified but the other is correct, the estimator from our approach produces little bias in a small sample size; moreover, even if both working models are misspecified, the estimator performs at least as well as the Kaplan-Meier estimator. Thus, we conclude that when high-dimensional auxiliary information is present, it is always worth it to condense this information by modeling both lifetime and censoring time given covariates. In detail, among the following simulations,

Simulation 1 indicates that, in situation in which ignoring the dependent censoring does not really bias the Kaplan-Meier estimator (we think these are situations in which the sample variability
dominates the bias caused by ignoring the dependent censoring), our estimator also performs well; this is true even when neither working model is correct; i.e. it is worthwhile to use two working models. Moreover, using covariates which do not predict $T$ or $C$ in the working models does not increase bias.

*Simulation 2* indicates that, in situation in which the dependent censoring causes bias in the Kaplan-Meier estimator (i.e., sample size is sufficiently large so that sample variability does not dominate the bias caused by ignoring the dependent censoring), our estimator out-performs the Kaplan-Meier estimator; this is true even when neither working model is correct; i.e. it is worthwhile to use two working models. Moreover, using covariates which do not predict $T$ or $C$ in the models does not increase bias.

*Simulation 3* indicates that, in situation in which the sample size is sufficiently large so that sample variability does not dominate the bias caused by ignoring the dependent censoring and in which two incorrect working models lead to appreciable bias (but not significantly larger than that evidenced in the Kaplan-Meier), getting either working model correct produces less bias than getting neither working model correct. Moreover, using covariates which do not predict $T$ or $C$ in the models does not increase bias.

*Simulation 4* indicates that, when the underlying models for $T$ or $C$ given $L$ are not proportional hazard models, the above conclusions are also correct. That is, the Kaplan-Meier estimator has appreciable bias but our estimator produces less bias. It is true even if we make two incorrect working models. In other words, it is worth it to make two working models even if the underlying models are not proportional hazard models.

In *Simulation 1, Simulation 2, Simulation 3 and Simulation 4*, the working models used in our approach are the Cox’s proportional hazard models with either main effects or with additional two-way interactions of all the covariates. The estimated averages of median lifetime and their Monte-Carlo variances based on 1000 repeated samples are reported. These results are then compared among the estimators based on different working models and the Kaplan-Meier estimator. In the tables, the symbol $T(L_1, ..., L_k)$ denotes that the working model for lifetime ($T$ denotes lifetime) is the Cox’s model with $L_1, ..., L_k$ as covariates; the symbol $C(L_1, ..., L_k)$ ($C$ denotes the censoring time) denotes that the working model for censoring time is the Cox’s model with $L_1, ..., L_k$ as covariates. “KM” denotes the Kaplan-Meier estimator.
Our second goal is to indicate how our approach can do a better job than the intuitive approach; the intuitive approach estimates the conditional survival function based on a Cox’s proportional hazard model then averages over all the covariates. That is, the estimator from the intuitive approach is given by
\[ \frac{1}{n} \sum_{i=1}^{n} e^{-\hat{\Lambda}_n(t)e^{\hat{\beta}_n L_i}}, \]
where \( \hat{\Lambda}_n(.) \) and \( \hat{\beta}_n \) are obtained from the usual Cox’s regression. Therefore, in the following simulations, Simulation 5, Simulation 6 and Simulation 7 are designed from different underlying models in order to compare the results from this intuitive approach with the results from our approach. In addition, it is assumed that the modeling abilities in the intuitive approach and in our approach are the same; i.e., both the intuitive approach and our approach model event times given all the covariates using the Cox’s proportional hazard models with only main effects of the covariates. The estimated averages of median lifetime and their Monte-Carlo standard errors based on 1000 repeated samples are reported in Table 5 to Table 7.

The details of each simulation are in the following.

**Simulation 1.** The underlying models for lifetime and censoring time given covariates are both the Cox’s proportional hazard models and their hazard rate functions are
\[ h_{T|L}(t|L) = t^4 \exp\{2L_2 - 0.2L_3 - 2L_1L_2 + 3L_1L_3 + 3L_2L_3\}, \]
and
\[ h_{C|L}(t|L) = t^4 \exp\{-0.6 - 0.1L_1 + L_2 - L_3 - 2L_1L_2 + 2L_1L_3 + 2L_2L_3\}, \]
where \( L_1, ..., L_3 \) are independent covariates generated from a uniform distribution between 0 and 1. Moreover, another independent covariate \( \bar{L} \) is generated from a uniform distribution between 0 and 1. The maximum of the censoring time is 0.745. About 30% of the observations are censored. The marginal correlation between \( T \) and \( C \) is only 0.23. The kernel function is chosen as \( K(x_1, x_2) = e^{-(x_1^2+x_2^2)}I_{x_1^2+x_2^2<1} \) and the bandwidth \( a_n = 6n^{-1/3} \). Table 1 reports the results using sample size 100. From Table 1, we see this is the situation when the sample variability dominates the sample bias so even the Kaplan-Meier estimator produces little bias. As evidenced in Table 1, even if we misspecify both working models or add an irrelevant covariate in the working models, the bias from our estimators is also small.

**Simulation 2.** The underlying models for lifetime and censoring time given covariates are both the Cox’s proportional hazard models and their hazard rate functions are
\[ h_{T|L}(t|L) = t^4 \exp\{-5 + 4L_1 + 6L_2 + 3L_3 + 0.1L_1L_2 + 0.1L_2L_3\}, \]
Simulation 1: true median=0.563 (n=100, repetition num.=1000)

<table>
<thead>
<tr>
<th>working models</th>
<th>median($T$) (MC variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T(L_1 - L_3, L_1L_2, L_1L_3, L_2L_3, L)$, $C(L_1 - L_3, L_1L_2, L_1L_3, L_2L_3, \bar{L})$</td>
<td>0.560(0.00100)</td>
</tr>
<tr>
<td>$T(L_1 - L_3, L_1L_2, L_1L_3, L_2L_3)$, $C(L_1 - L_3, L_1L_2, L_1L_3, L_2L_3)$</td>
<td>0.560(0.00100)</td>
</tr>
<tr>
<td>$T(L_1 - L_3, L_1L_2, L_1L_3, L_2L_3)$, $C^*$</td>
<td>0.559(0.00100)</td>
</tr>
<tr>
<td>$T^*, C(L_1 - L_3, L_1L_2, L_1L_3, L_2L_3)$</td>
<td>0.565(0.00104)</td>
</tr>
<tr>
<td>$T(L_1 - L_3, L_1L_2, L_1L_3, L_2L_3)$, $C(L_1 - L_3)^*$</td>
<td>0.559(0.00099)</td>
</tr>
<tr>
<td>$T(L_1 - L_3)^*, C(L_1 - L_3, L_1L_2, L_1L_3, L_2L_3)$</td>
<td>0.560(0.00100)</td>
</tr>
<tr>
<td>$T(L_1 - L_3)^<em>, C(L_1 - L_3)^</em>$</td>
<td>0.560(0.00099)</td>
</tr>
<tr>
<td>$T^<em>, C(L_1 - L_3)^</em>$</td>
<td>0.561(0.00100)</td>
</tr>
<tr>
<td>$T(L_1 - L_3)^<em>, C^</em>$</td>
<td>0.560(0.00099)</td>
</tr>
<tr>
<td>KM</td>
<td>0.560(0.00099)</td>
</tr>
</tbody>
</table>

Table 2:
(* indicates that the working model is misspecified)
Simulation 2: true median=0.525 (n=100, repetition num.=1000)

<table>
<thead>
<tr>
<th>working models</th>
<th>average of median(T) (MC variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T(L_1 - L_4, L_1L_2, L_1L_3, L_2L_3, L) ), ( C(L_1 - L_4, L_1L_2, L_1L_3, L_2L_3, \hat{L}) )</td>
<td>0.525(0.00174)</td>
</tr>
<tr>
<td>( T(L_1 - L_4, L_1L_2, L_1L_3, L_2L_3) ), ( C(L_1 - L_4, L_1L_2, L_1L_3, L_2L_3) )</td>
<td>0.525(0.00174)</td>
</tr>
<tr>
<td>( T(L_1 - L_4, L_1L_2, L_1L_3, L_2L_3), C(L_1 - L_4)^* )</td>
<td>0.526(0.00172)</td>
</tr>
<tr>
<td>( T(L_1 - L_4)^*, C(L_1 - L_4, L_1L_2, L_1L_3, L_2L_3) )</td>
<td>0.529(0.00180)</td>
</tr>
<tr>
<td>( T(L_1 - L_4, L_1L_2, L_1L_3, L_2L_3), C^* )</td>
<td>0.524(0.00173)</td>
</tr>
<tr>
<td>( T^*, C(L_1 - L_4, L_1L_2, L_1L_3, L_2L_3) )</td>
<td>0.525(0.00173)</td>
</tr>
<tr>
<td>( T(L_1 - L_4)^<em>, C(L_1 - L_4)^</em> )</td>
<td>0.524(0.00174)</td>
</tr>
<tr>
<td>( T^<em>, C(L_1 - L_4)^</em> )</td>
<td>0.529(0.00178)</td>
</tr>
<tr>
<td>( T(L_1 - L_4)^<em>, C^</em> )</td>
<td>0.526(0.00171)</td>
</tr>
<tr>
<td>KM</td>
<td>0.555(0.00200)</td>
</tr>
</tbody>
</table>

Table 3: (* indicates that the working model is misspecified)

Table 3:

\( h_{C|L}(t|L) = t^4 \exp\{-2.5 + L_1 + 5L_2 + L_3 + 0.2L_2L_3\}, \)

where \( L_1, L_2, L_3 \) are independent covariates generated from a uniform distribution between 0 and 1.

Moreover, another independent covariate \( \hat{L} \) is generated from a uniform distribution between 0 and 1. The maximum of the censoring time is 1.25. About 33% of the observations are censored. The marginal correlation between \( T \) and \( C \) is 0.58. The kernel function is chosen as \( K(x_1, x_2) = e^{-(x_1^2 + x_2^2)}I_{x_1^2 + x_2^2 < 1} \) and the bandwidth \( a_n = 8n^{-1/3} \). Table 2 reports the results using sample size 100. From Table 2, we see this is the situation that the sample variability is relatively small so the Kaplan-Meier estimator has a significant bias. However, the estimators from our approach by using either correct or wrong working models produce little bias; moreover, adding another irrelevant covariate does not increase bias.
Simulation 3. The underlying models for lifetime and censoring time given covariates are both the Cox’s proportional hazard models and their hazard rate functions are

\[ h_{T\mid L}(t\mid L) = t^4 \exp\{2L_2 - 0.2L_3 + 0.01L_6 - 2L_1L_2 + 3L_1L_3 + 3L_2L_3\}, \]

and

\[ h_{C\mid L}(t\mid L) = t^3 \exp\{-0.1L_1 + L_2 - L_3 + 0.01L_4 - 2L_1L_2 + 2L_1L_3 + 2L_2L_3\}, \]

where \( L_1, L_2, L_3 \) are independent covariates generated from a Bernoulli distribution with probability 0.5 at 0 and 1 and \( L_4, L_5, L_6 \) are independently generated from a uniform distribution between 0 and 1 and \( L_4 \) is an independent Bernoulli variable with probability 0.5 at 0 and 1. The maximum of the censoring time is 1.20. About 32% of the observations are censored. The marginal correlation between \( T \) and \( C \) is 0.55. The kernel function is chosen as \( K(x_1, x_2) = e^{-(x_1^2 + x_2^2)}I_{x_1^2 + x_2^2 < 1} \) and the bandwidth \( a_n = 3n^{-1/3} \). Table 3 reports the results using sample size 100. From Table 3, we observe that this is the situation when the sample bias is more dominating than the sample variability so the Kaplan-Meier estimator has a very significant bias. Moreover, the estimator from our approach by using two incorrect working models also produce significant bias but not as large as the Kaplan-Meier estimator. However, if either working model is correct, the estimators turn out to be very little biased.

Simulation 4. The underlying models for lifetime and censoring time given covariates are not the Cox’s proportional hazard models and \( T \) and \( C \) are generated as follows

\[ \ln T \sim N(-2 + L_1 + L_3 + L_1L_3 + L_2L_3, 0.25), \]

and

\[ \ln C \sim N(-1.5 + L_3 + L_1L_3 + 2L_2L_3, 0.25), \]

where \( L_1, L_2, L_3 \) are independent covariates generated from a uniform distribution between 0 and 1 and \( L_4 \) is an independent Bernoulli variable with probability 0.5 at 0 and 1. The maximum of the censoring time is 2.00. About 42% of the observations are censored. The marginal correlation between \( T \) and \( C \) is 0.53. The kernel function is chosen as \( K(x_1, x_2) = e^{-(x_1^2 + x_2^2)}I_{x_1^2 + x_2^2 < 1} \) and the bandwidth \( a_n = 8n^{-1/3} \). Table 4 reports the results using sample size 100. From Table 4, we observe that each estimator is significantly biased. Interestingly, the Kaplan-Meier estimator has the largest bias; the estimator from our approach by using either correct or incorrect working models also produce significant bias but not as large as the Kaplan-Meier estimator.
Simulation 3. true median=0.605 (n=100, repetition num.=1000) working models average of median(T) (MC variance)

<table>
<thead>
<tr>
<th>working models</th>
<th>average of median(T) (MC variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T(L_1 - L_6, L_1 L_2, L_1 L_3, L_2 L_3)$, $C(L_1 - L_6, L_1 L_2, L_1 L_3, L_2 L_3)$</td>
<td>0.606(0.00207)</td>
</tr>
<tr>
<td>$T(L_1 - L_6, L_1 L_2, L_1 L_3, L_2 L_3), C(L_1 - L_6)$</td>
<td>0.605(0.00205)</td>
</tr>
<tr>
<td>$T(L_1 - L_6), C(L_1 - L_6, L_1 L_2, L_1 L_3, L_2 L_3)$</td>
<td>0.606(0.00203)</td>
</tr>
<tr>
<td>$T^<em>, C(L_1 - L_6, L_1 L_2, L_1 L_3, L_2 L_3), C^</em>$</td>
<td>0.602(0.00200)</td>
</tr>
<tr>
<td>$T^*, C(L_1 - L_6, L_1 L_2, L_1 L_3, L_2 L_3)$</td>
<td>0.604(0.00203)</td>
</tr>
<tr>
<td>$T(L_1 - L_6), C(L_1 - L_6)$</td>
<td>0.623(0.00219)</td>
</tr>
<tr>
<td>$T^<em>, C(L_1 - L_6)^</em>$</td>
<td>0.644(0.00222)</td>
</tr>
<tr>
<td>$T(L_1 - L_6)^<em>, C^</em>$</td>
<td>0.630(0.00221)</td>
</tr>
<tr>
<td>KM</td>
<td>0.645(0.00217)</td>
</tr>
</tbody>
</table>

Table 4: (∗ indicates that the working model is misspecified)

Simulation 5. The underlying models for lifetime and censoring time given covariates are both the Cox’s proportional hazard models and their hazard rate functions are

$$h_{T|L}(t|L) = t^4 \exp\{-5 + 4L_1 + 6L_2 + 4L_3\},$$

and

$$h_{C|L}(t|L) = t^4 \exp\{-2.5 + L_1 + 5L_2 + L_3\},$$

where $L_1, L_2, L_3$ are independent covariates from a uniform distribution between 0 and 1. The maximum of the censoring time is 1.257. About 33% of the observations are censored. The marginal correlation between $T$ and $C$ is 0.57. The kernel function is chosen as $K(x_1, x_2) = e^{-(x_1^2 + x_2^2)}I_{x_1^2 + x_2^2 < 1}$ and the bandwidth $a_n = 8n^{-1/3}$. Table 5 reports the results using sample size 200. In Table 5, our approach (I) means that we use both the Cox’s models with main effects as the working models for $T$ and $C$; thus we use two correct working models; our approach (II) means that we only use the correct model for $T$ but treat $C$ as completely independent; the intuitive approach estimates the survival function based on the Cox’s model with main effects. From Table 5, obviously, the intuitive estimator has a small bias as expected and we also observe that, the Kaplan-Meier estimator has a significant
Simulation 4: true median=0.579 (n=100, repetition num.=1000)

<table>
<thead>
<tr>
<th>working models</th>
<th>average of median(T) (MC variance)</th>
</tr>
</thead>
</table>
| $T(L_1 - L_3, L_1 L_2, L_1 L_3, L_2 L_3, L)$* ,  
$C(L_1 - L_3, L_1 L_2, L_1 L_3, L_2 L_3, \tilde{L})$* | 0.590(0.00290) |
| $T(L_1 - L_3, L_1 L_2, L_1 L_3, L_2 L_3)$* ,  
$C(L_1 - L_3, L_1 L_2, L_1 L_3, L_2 L_3)$* | 0.590(0.00290) |
| $T(L_1 - L_3, L_1 L_2, L_1 L_3, L_2 L_3)$* ,  
$C(L_1 - L_3)$* | 0.590(0.00291) |
| $T(L_1 - L_3)$* ,  
$C(L_1 - L_3, L_1 L_2, L_1 L_3, L_2 L_3)$* | 0.593(0.00291) |
| $T(L_1 - L_3, L_1 L_2, L_1 L_3, L_2 L_3)$* ,  
$C$* | 0.591(0.00266) |
| $T^*$ ,  
$C(L_1 - L_3, L_1 L_2, L_1 L_3, L_2 L_3)$* | 0.595(0.00287) |
| $T(L_1 - L_3)$* ,  
$C(L_1 - L_3)$* | 0.593(0.00295) |
| $T^*$ ,  
$C(L_1 - L_3)$* | 0.599(0.00297) |
| $T(L_1 - L_3)$* ,  
$C$* | 0.598(0.00268) |
| KM | 0.702(0.00357) |

Table 5:  
(* indicates that the working model is misspecified)
Simulation 5: true median=0.529 (n=200, repetition num.=1000)

<table>
<thead>
<tr>
<th>working models</th>
<th>average of median(T) (MC variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Approach (I)</td>
<td>$T(L_1 - L_3), C(L_1 - L_3)$</td>
</tr>
<tr>
<td>Our Approach (II)</td>
<td>$T(L_1 - L_3), C^*$</td>
</tr>
<tr>
<td>Intuitive Approach</td>
<td></td>
</tr>
<tr>
<td>KM</td>
<td></td>
</tr>
</tbody>
</table>

Table 6:

(* indicates that the working model is misspecified)

and the largest bias; the estimators from both our approach (I) and (II) produce very small bias, which are competent with the intuitive estimator.

Simulation 6. The underlying models for lifetime and censoring time given covariates are both the Cox’s proportional hazard models and their hazard rate functions are

$$h_{T|L}(t|L) = t^4 \exp \{-2 - 2L_1L_2 + 6L_1L_3 + 6L_2L_3\},$$

and

$$h_{C|L}(t|L) = t^4 \exp \{1 - 3L_2 - 3L_3\},$$

where $L_1, L_2, L_3$ are independent covariates from a uniform distribution between 0 and 1. The maximum of the censoring time is 1.600. About 26% of the observations are censored. The marginal correlation between $T$ and $C$ is -0.46. The kernel function is chosen as $K(x_1, x_2) = e^{-(x_1^2 + x_2^2)}I_{x_1^2 + x_2^2 < 1}$ and the bandwidth $a_n = 4n^{-1/3}$. Table 6 reports the results using sample size 200. In Table 6, ”our approach” means that we use both the Cox’s models with only main effects as the working models for $T$ and $C$ and the intuitive approach estimates the survival function based on the Cox’s model with only main effects. From Table 6, the intuitive estimator has a large bias and it does as poorly as the Kaplan-Meier estimator; but if we use the correct working model for the censoring time, the estimator from our approach produces very small bias, which are much better than the intuitive estimator.

Simulation 7. The underlying models for lifetime and censoring time given covariates are not the Cox’s proportional hazard models and $T$ and $C$ are generated from the following distribution

$$\ln T \sim N(-1.5 + L_1 + L_2 + L_3, 0.16),$$

26
**Simulation 6:** true median=0.819 (n=200, repetition num.=1000)

<table>
<thead>
<tr>
<th>working models</th>
<th>average of median(T) (MC variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Approach</td>
<td>0.822(0.00153)</td>
</tr>
<tr>
<td>(T(L_1 - L_3)^*, C(L_1 - L_3))</td>
<td></td>
</tr>
<tr>
<td>Intuitive Approach</td>
<td>0.807(0.00156)</td>
</tr>
<tr>
<td>KM</td>
<td>0.806(0.00143)</td>
</tr>
</tbody>
</table>

Table 7:  
(* indicates that the working model is misspecified)

**Simulation 7:** true median=1.002 (n=200, repetition num.=1000)

<table>
<thead>
<tr>
<th>working models</th>
<th>average of median(T) (MC variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Approach</td>
<td>1.143(0.00395)</td>
</tr>
<tr>
<td>(T(L_1 - L_3)^<em>, C(L_1 - L_3)^</em>)</td>
<td></td>
</tr>
<tr>
<td>Intuitive Approach</td>
<td>1.377(0.00528)</td>
</tr>
<tr>
<td>KM</td>
<td>1.149(0.00247)</td>
</tr>
</tbody>
</table>

Table 8:  
(* indicates that the working model is misspecified)

\[
\ln C \sim N(-2 + 4L_1 + L_2, 0.16),
\]

where \(L_1, L_2, L_3\) are independent covariates from a uniform distribution between 0 and 1. The maximum of the censoring time is 2.299. About 41% of the observations are censored. The marginal correlation between \(T\) and \(C\) is 0.42. The kernel function is chosen as \(K(x_1, x_2) = e^{-(x_1^2 + x_2^2)x_1^2 + x_2^2 < 1}\) and the bandwidth \(a_n = 24n^{-1/3}\). Table 7 reports the results using sample size 200. In Table 7, ”our approach” means that we use both the Cox’s models with only main effects as the working models for \(T\) and \(C\) and the intuitive approach estimates the survival function based on the Cox’s model with only main effects. As expected, every estimator is significantly biased. However, the estimator from our approach has the least bias, compared with the intuitive estimator and the Kaplan-Meier estimator.

One additional simulation (Simulation 8) has also been performed using the algorithm of estimating asymptotic variance which was given in Section 5. The data were generated using the same design as in Simulation 3. For each sample (n=100) of 1000 samples, we used the algorithm in Section 5.
Simulation 8: \( t_0 = 0.605, S(t_0) = 0.500 \) (n=100, repetition num.=1000)

<table>
<thead>
<tr>
<th>working models</th>
<th>( \hat{S}_n(t_0) ) average of</th>
<th>Monte-Carlo std. dev.</th>
<th>estimated std. dev.</th>
<th>coverage prob. 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T(L_1 - L_6, L_1L_2, L_1L_3, L_2L_3) ), ( C(L_1 - L_6, L_1L_2, L_1L_3, L_2L_3) )</td>
<td>0.502</td>
<td>0.0542</td>
<td>0.0524</td>
<td>0.93</td>
</tr>
<tr>
<td>( T(L_1 - L_6)^<em>, C(L_1 - L_6)^</em> )</td>
<td>0.502</td>
<td>0.0539</td>
<td>0.0521</td>
<td>0.93</td>
</tr>
<tr>
<td>( T(L_1 - L_6), C(L_1 - L_6)^*, C(L_1 - L_6, L_1L_2, L_1L_3, L_2L_3) )</td>
<td>0.503</td>
<td>0.0541</td>
<td>0.0529</td>
<td>0.94</td>
</tr>
<tr>
<td>( T^*, C(L_1 - L_6, L_1L_2, L_1L_3, L_2L_3) )</td>
<td>0.502</td>
<td>0.0537</td>
<td>0.0520</td>
<td>0.93</td>
</tr>
<tr>
<td>( T^*, C(L_1 - L_6, L_1L_2, L_1L_3, L_2L_3) )</td>
<td>0.504</td>
<td>0.0548</td>
<td>0.0550</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table 9: (* indicates that the working model is misspecified)

\((a_n = 3n^{-1/3}, \epsilon_n = n^{-5/12})\) to calculate the asymptotic variance for the estimator of the survival probability at \( t = 0.605 \) (the true survival probability \( S(0.605) \) is 0.50); thus, we gave approximated 95% confidence intervals and calculated their coverage probability. The result is in Table 9. Table 9 indicates that when either working model is correct, our algorithm of estimating the asymptotic variance gives a fairly correct coverage and that the estimated standard deviation is close to the Monte-Carlo standard deviation. It is also observed that the estimated standard deviation using our algorithm is about 4% smaller than the Monte-Carlo standard deviation and we think this may be due to the large variability in such a small sample.

7 Discussion

The simulation results in the previous section indicate that, in small samples when one of the sample variability and the sample bias is dominating over the other (as evidenced by the performance of the Kaplan-Meier estimator), the estimators from our approach always produce little bias if either working model is correct; moreover, even if both models are misspecified, the estimator from our approach produce less bias than the bias of the Kaplan-Meier estimator. The simulations also indicate that including irrelevant covariates in the working models does not significantly increase the bias. Combined with our large sample results, we thus conclude that, when right-censored data include high-dimensional auxiliary covariates, condensing such information by utilizing working models for
both lifetime and censoring time given covariates can make adjusting for dependent censoring possible
and produce an estimator which is robust to the misspecification of either working model, robust to
the sample variability in a small sample, and robust to accidentally using irrelevant information.

It is observed in our simulations that the choice of the bandwidth $a_n$ plays a very important role
in influencing the bias. A large $a_n$ may over-smooth the conditional hazard rate estimator (in fact,
with simple calculation, for fixed $n$, if $a_n$ is close to infinity, our estimator is close to the Kaplan-Meier
estimator); while a small $a_n$ may under-smooth the conditional hazard rate estimator thus introduce
large variation in estimation. Both situations will increase the mean square errors of the estimated
hazard rate function thus will cause bias in the estimator of the survival function. In some of our
simulations, inappropriate choice of $a_n$ can even produce as large bias in our estimator as in the
Kaplan-Meier estimator. So far, we let $a_n$ be a constant only depending on $n$ (but it may change
from one simulation to another simulation) and no general selection rule is followed; however, one of
our future goals is to make the choice of $a_n$ be data-adaptive; for example, $a_n$ can be chosen as the
Euclidean distance between each location and its $k$-th nearest observation ($k$ is a fixed integer).

Though we hope that our working models are correct, we never know this in reality. To make this
hope more likely, we may use more general models other than the Cox’s models as working models,
such as, using generalized additive model, using splines as covariates in working models, and etc.
At the same time, a model selection rule is necessary so that the optimal working models can be
determined eventually. For instance, suppose the Cox’s models are used as the working models; if
after adding one covariate or one high-order interaction to the working models, the estimated median
changes much, then we may consider such a covariate or interaction should be included. Therefore,
a test for goodness of fit as well as a test for comparing two different sets of working models will be
useful in practical applications.

A final comment is, the ideas of our approach are not just limited to the application to right-
censored data. The similar ideas can be applied to many other missing data problems, such as missing
covariates, current status data, missingness in causal inference, and etc. Another important future
work is to generalize these ideas to be able to adjust for dependent censoring using time-dependent
covariates.
APPENDIX

A  Asymptotic Properties of Mean Response Estimator.

Recall
\[
\hat{E}Y = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{Y_i R_i K(\frac{Z_j - \hat{Z}_i}{a_n})}{\sum_{k=1}^{n} R_k K(\frac{Z_j - \hat{Z}_k}{a_n})}.
\]

Before giving the asymptotic properties of \( \hat{E}Y \), we list all the assumptions as below.

(A’1.) \( R \) and \( Y \) are independent given \( L \).

(A’2.) \( L \)’s distribution is nondegenerate; moreover, there exists an unknown constant \( \delta > 0 \) such that
\[
\pi(R = 1|L) \geq \delta, \text{ a.s.}
\]

(A’3.) For \( r = 0 \) or \( r = 1 \), the joint density of
\[
(R = r, Y = y, L = l)
\]
has a bounded support and it is continuously twice-differentiable with respect to \((y, l)\). In addition, the density has a positive lower bound in its support.

(A’4.) The kernel function \( K(., .) \) has a support in the unit disk of \( R^2 \) and its gradient, \( \nabla K(., .) \), satisfies
\[
|\nabla K(., .)| \leq CK(., .).
\]
Moreover, it holds that
\[
\int_{(y_1, y_2) \in R^2} y_i K(y_1, y_2) dy_1 dy_2 = 0, i = 1, 2.
\]

(A’5.) \( na_n^2 \to \infty, na_n^4 \to 0 \).

**Theorem A.1** (Asymptotic properties of \( \hat{E}Y \)) Under Assumptions (A’1)-(A’5), if either working model is correct, \( \hat{E}Y \) is a consistent estimator for \( EY \) and \( \sqrt{n}(\hat{E}Y - EY) \) weakly converges to a normal distribution; if both working models are correct, the asymptotic variance of \( \sqrt{n}(\hat{E}Y - EY) \) is the same as the semiparametric efficiency bound.

**Proof.** Step 1. We show that there exist two constants \( \beta^* \) and \( \gamma^* \) such that \( \hat{\beta}_n \) and \( \hat{\gamma}_n \) are asymptotically linear estimators for \( \beta^* \) and \( \gamma^* \), respectively. In other words,
\[
\sqrt{n}(\hat{\beta}_n - \beta^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_\beta(\beta^*, R_i, R_i Y_i, L_i) + o_p(1),
\]

and
\[
\sqrt{n}(\hat{\gamma}_n - \gamma^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_{\gamma}(\gamma^*, R_i, R_iY_i, L_i) + o_p(1),
\]
for some influence functions \(S_{\beta}\) and \(S_{\gamma}\).

The \(\hat{\gamma}_n\) maximizes the function

\[
l_n(\gamma) = P_n[R\gamma' L - \ln(1 + e^{\gamma' L})].
\]

Define \(l_0(\gamma) = P[R\gamma' L - \ln(1 + e^{\gamma' L})]\). It is obvious from the Glivenko-Cantelli Theorem that in probability, for \(\gamma\) in any compact set \(A\),

\[
\sup_{\gamma \in A} |l_n(\gamma) - l_0(\gamma)| \to 0;
\]

\[
\sup_{\gamma \in A} |\nabla_{\gamma} l_n(\gamma) - \nabla_{\gamma} l_0(\gamma)| \to 0;
\]

\[
\sup_{\gamma \in A} |\nabla_{\gamma \gamma}^2 l_n(\gamma) - \nabla_{\gamma \gamma}^2 l_0(\gamma)| \to 0.
\]

Additionally,

\[
\nabla_{\gamma \gamma}^2 l_0(\gamma) = -P[\frac{e^{\gamma' L}}{(1 + e^{\gamma' L})^2} LL']
\]
is a strictly negative matrix since \(L\) is nondegenerate. So \(l_0(\gamma)\) has a unique maximizer \(\gamma^*\) and \(\hat{\gamma}_n \to \gamma^*\) in probability (Corollary II.2, Andersen and Gill (1982)).

\(\hat{\gamma}_n\) also solves the estimating equation: \(l'_n(\gamma) = 0\). With little difficulty, we have

\[
\sqrt{n}(\hat{\gamma}_n - \gamma^*) = -\sqrt{n}(P_n - P) \nabla_{\gamma \gamma}^2 l_0(\gamma^*)^{-1} [RL - \frac{Le^{\gamma' L}}{1 + e^{\gamma' L}}] + o_p(1).
\]

Similar proof can be applied to show that there exists a constant \(\beta^*\), such that \(\hat{\beta}_n\) converges to \(\beta^*\) in probability and moreover,

\[
\sqrt{n}(\hat{\beta}_n - \beta^*) = -\sqrt{n}(P_n - P) \{E[RLL']\}^{-1} R(Y - \beta^* L)L + o_p(1).
\]

Step 2. When either working model is correct, \(R\) and \(Y\) are independent given \((\beta^* L, \gamma^* L)\).

Clearly, if either working model is correct then either \(\beta^*\) or \(\gamma^*\) is the correct constant in the working models. That is, either \(p_{Y|L}(y|L = l) \sim N(\beta^* L, \sigma^2)\) or \(P(R = 1|L) = e^{\gamma^* L}/(1 + e^{\gamma^* L})\).

Hence, if \(p_{Y|L}(y|L = l) \sim N(\beta^* L, \sigma^2)\),

\[
P(R = 1, Y \leq y|\beta^* L, \gamma^* L) = E[E[R1_{Y \leq y}|L]|\beta^* L, \gamma^* L]
\]

\[
= E[E[P(R = 1|L)P(Y \leq y|L)|\beta^* L, \gamma^* L]
\]
Then $\hat{\beta}'_n = \arg\min_{\beta'} \frac{1}{n} \sum_{i=1}^{n} g^2_{\beta'}(Y_i, Z_i)$. On the other hand, if $P(R = 1|L) = e^{\gamma' L}/(1 + e^{\gamma' L})$,

$$P(R = 1, Y \leq y|\beta' L, \gamma' L) = E[P(R = 1|L)P(Y \leq y|L)] = E[Y, R, \hat{Y}, \hat{R}, \hat{E}Y] = \Phi(\frac{\hat{Y} - \beta' L}{\sigma})$$

Thus, $R$ and $Y$ are independent given $(\beta' L, \gamma' L)$.

Step 3. We write $\hat{E}Y$ in terms of empirical process $\sqrt{n}(P_n - P)$ as follows.

First, define

$$\hat{Z} = (\hat{\beta}'_n L, \hat{\gamma}'_n L), Z^* = (\beta' L, \gamma' L),$$

$$g^3_n(y, r, z; \hat{z}) = yrK(\frac{\hat{z} - z}{a_n}),$$

$$g^3_n(y, r, z; \hat{z}) = rK(\frac{\hat{z} - z}{a_n}).$$

and

$$g^3_n(\hat{z}) = \frac{P_n[g^3_n(Y, R, \hat{Z}; \hat{z})]}{P_n[g^3_n(R, \hat{Z}; \hat{z})]}.$$

Then $\hat{E}Y = P_n[g^3_n(\hat{Z})]$. On the other hand,

$$g^3_n(\hat{z}) = \frac{(P_n - P)[g^3_n(Y, R, \hat{Z}; \hat{z})]}{P_n[g^3_n(R, Z; \hat{z})]} + \frac{P[g^3_n(Y, R, \hat{Z}; \hat{z})]}{P_n[g^3_n(R, Z; \hat{z})]}$$

$$= \frac{(P_n - P)[g^3_n(Y, R, \hat{Z}; \hat{z})]}{P_n[g^3_n(R, Z; \hat{z})]} + \frac{P[g^3_n(Y, R, \hat{Z}; \hat{z})]}{P_n[g^3_n(R, Z; \hat{z})]} - \frac{P[g^3_n(Y, R, \hat{Z}; \hat{z})]}{P_n[g^3_n(R, Z; \hat{z})]}$$

$$= (P_n - P)\frac{g^3_n(Y, R, \hat{Z}; \hat{z})}{P_n[g^3_n(R, Z; \hat{z})]}.$$
Then  

\[
\frac{P[g_1^n(Y, R, \hat{Z}; \hat{z})](P_n - P)[g_2^n(R, \hat{Z}; \hat{z})]}{P_n[g_2^n(R, \hat{Z}; \hat{z})]P[g_2^n(R, \hat{Z}; \hat{z})]}
\]

+ \frac{P[g_1^n(Y, R, \hat{Z}; \hat{z})]}{P[g_2^n(R, \hat{Z}; \hat{z})]}.  

If further define  

\[
h_1^n(y, r, z) = P[g_1^n(y, r, z; \hat{z})]\frac{P[g_2^n(R, \hat{Z}; \hat{z})]}{P_n[g_2^n(R, \hat{Z}; \hat{z})]P[g_2^n(R, \hat{Z}; \hat{z})]} |_{\hat{z} = \hat{z}},
\]

and  

\[
h_2^n(r, z) = P[g_2^n(R, \hat{Z}; \hat{z})|_{\hat{z} = \hat{z}},
\]

then  

\[
\hat{E}Y = (P_n - P)g_3^n(\hat{Z}) + P g_3^n(\hat{Z})
\]

= (P_n - P)g_3^n(\hat{Z})

+ (P_n - P)h_1^n(Y, R, \hat{Z}) - (P_n - P)h_2^n(R, \hat{Z})

+ P\left\{\frac{P[g_1^n(Y, R, \hat{Z}; \hat{z})]}{P[g_2^n(R, \hat{Z}; \hat{z})]} |_{\hat{z} = \hat{z}}\right\}.

Denote \(w_n(Y, R, L; \hat{\beta}_n, \hat{\gamma}_n)\) as

\[
g_3^n(\hat{Z}) - h_1^n(Y, R, \hat{Z}) + h_2^n(R, \hat{Z}).
\]

Then

\[
\sqrt{n}(\hat{E}Y - EY)
\]

= \sqrt{n}(P_n - P)[w_n(Y, R, L; \hat{\beta}_n, \hat{\gamma}_n)]

+ \sqrt{n}\{P\left\{\frac{P[g_1^n(Y, R, \hat{Z}; \hat{z})]}{P[g_2^n(R, \hat{Z}; \hat{z})]} |_{\hat{z} = \hat{z}}\right\} - EY\}.

Step 4. Prove that there exists a measurable function \(w(Y, R, L; \beta^*, \gamma^*)\) (independent of \(n\)) such that

\[
\sqrt{n}(P_n - P)[w_n(Y, R, L; \hat{\beta}_n, \hat{\gamma}_n)] = \sqrt{n}(P_n - P)[w(Y, R, L; \beta^*, \gamma^*)] + o_p(1).
\]

This can be seen from the following results.

1. By Bernstein’s inequality, for any \(x > 0\),

\[
P\left\{\left|\frac{1}{a_n^2} RYK(\hat{Z} - \hat{z})\right| > x\right\} \leq 2\exp\left\{-\frac{1}{4}\left|\frac{1}{a_n^2} RYK(\hat{Z} - \hat{z})\right|^2 + x/a_n^2\right\}
\]

\[
\leq C_1\exp\{-C_2n a_n^2 x\} \to 0,
\]

\[
P_n[\frac{1}{a_n^2} g_1^n(Y, R, \hat{Z}; z)] = P[\frac{1}{a_n^2} g_1^n(Y, R, \hat{Z}; z)] + o_p(1).
\]
Similarly,
\[ P_n \left[ \frac{1}{a_n} g_2^n (R, \hat{Z} ; z) \right] = P \left[ \frac{1}{a_n^2} g_2^n (R, \hat{Z} ; z) \right] + o_p(1). \]

Hence,
\[
\begin{align*}
g_3^n (z) &= P \left[ \frac{g_1^n (Y, R, \hat{Z} ; z)}{g_2^n (R, \hat{Z} ; z)} \right] + o_p(1) = P \left[ \frac{RY K (\hat{Z} - \tilde{z})}{RY (\hat{Z} - \tilde{z})} \right] + o_p(1) \\
&= \frac{E[Y | \hat{Z} = z]}{E[Y | \hat{Z} = z]} + o_p(1). \\
\end{align*}
\]

Indeed, this convergence in probability can be further shown to be uniform in \( z \) over any compact set of the support of \( Z^* \). Its proof imitates the proof in Dabrowska (1987) and we skip it here.

2. Similarly, we obtain that uniformly in \((y, z)\) over a compact set of the support of \((Y, Z^*)\), in probability,
\[
\begin{align*}
h_1^n (y, r, z) &\to \frac{ry}{E[R | Z^* = z]} , \\
h_2^n (r, z) &\to \frac{r E[Y | Z^* = z]}{E[R | Z^* = z]^2} , \\
\end{align*}
\]

3. \( w_n (y, r, l; \hat{\beta}_n, \hat{\gamma}_n) \in \mathcal{F}_n \), where
\[
\mathcal{F}_n = \{ w_n (y, r, l; \beta^*, \gamma^*) + \frac{1}{\sqrt{n}} \theta_1, \gamma^* + \frac{1}{\sqrt{n}} \theta_2) : (\theta_1, \theta_2) = O_p(1) \}.
\]

We can apply the Donsker theorem of changing classes (Theorem 2.11.23, van der Vaart and Wellner (1996)) to the class \( \mathcal{F}_n \) which is indexed by \((n, \theta_1, \theta_2)\). It is sufficient to verify that
\[ w_n (y, r, l; \beta^* + \frac{1}{\sqrt{n}} \theta_1, \gamma^* + \frac{1}{\sqrt{n}} \theta_2) \]
is Lipschitz with respect to \((\theta_1, \theta_2)\) in any bounded set of \((\theta_1, \theta_2)\) and such Lipschitz coefficients are bounded by \( O_p (\frac{1}{\sqrt{n a_n}}) \).

From (1-3), we also obtain
\[
w(Y, R, L; \beta^*, \gamma^*) = E[Y | Z^*] - EY + \frac{R(Y - E[Y | Z^*])}{E[R | Z^*]} .
\]

Step 5. Work out the term
\[
\sqrt{n} \left\{ P \left[ \frac{g_1^n (Y, R, \hat{Z} ; \hat{z})}{g_2^n (R, \hat{Z} ; \hat{z})} \right] \right\}_{\hat{z}=\hat{z}} = EY .
\]
\[
\sqrt{n}\{P\left[\frac{g^n(Y, R, \hat{Z}; \tilde{z})}{g^n(R, \tilde{Z}; \hat{z})}|\tilde{z} = \hat{Z}\right] - EY\} \\
= \sqrt{n}\{P\left[E[R\hat{Y}|\hat{Z}]\right] + O_p(a_n^2) - EY\} \\
= \sqrt{n}\left(E\left[\frac{RY}{E[R|\gamma^*L, \beta^*L]}\right] - EY\right) \\
+ \frac{d}{d\gamma}|_{\gamma = \gamma^*}E\left[\frac{RY}{E[R|\gamma^*L, \beta^*L]}\right]\sqrt{n}(\gamma_n - \gamma^*) \\
+ \frac{d}{d\beta}|_{\beta = \beta^*}E\left[\frac{RY}{E[R|\gamma^*L, \beta^*L]}\right]\sqrt{n}(\beta_n - \beta^*) \\
+ o_p(1).
\]

Finally,

\[
\sqrt{n}(E\hat{Y} - EY) \\
= \sqrt{n}(P_n - P)w(Y, R, L; \beta^*, \gamma^*) \\
+ \frac{d}{d\gamma}|_{\gamma = \gamma^*}E\left[\frac{RY}{E[R|\gamma^*L, \beta^*L]}\right]\sqrt{n}(\gamma_n - \gamma^*) \\
+ \frac{d}{d\beta}|_{\beta = \beta^*}E\left[\frac{RY}{E[R|\gamma^*L, \beta^*L]}\right]\sqrt{n}(\beta_n - \beta^*) \\
+ \sqrt{n}(E\left[\frac{RY}{E[R|\gamma^*L, \beta^*L]}\right] - EY) \\
+ o_p(1).
\]

Step 6. When either working model is correct, \(R\) and \(Y\) are independent given \((\beta^*L, \gamma^*L)\). So the asymptotic bias

\[
\sqrt{n}(E\hat{Y} - EY) = 0.
\]

Hence, the consistency and the asymptotic normality of \(\hat{EY}\) hold.

Step 7. When both working models are correct, for any \((\beta, \gamma)\),

\[
E\left[\frac{RY}{E[R|\gamma^*L, \beta^*L]}\right] = E\left[\frac{RY}{E[R|\gamma^*L, \beta^*L]}\right] = EY.
\]

So

\[
\frac{d}{d\gamma}|_{\gamma = \gamma^*}E\left[\frac{RY}{E[R|\gamma^*L, \beta^*L]}\right] = 0
\]

and

\[
\frac{d}{d\beta}|_{\beta = \beta^*}E\left[\frac{RY}{E[R|\gamma^*L, \beta^*L]}\right] = 0.
\]

Hence, \(\hat{EY}\) has an influence function

\[
w(Y, R, L; \beta^*, \gamma^*) = E[Y|Z^*] - EY + \frac{R(Y - E[Y|Z^*])}{E[R|Z^*]}.
\]

35
Moreover, the conditional distribution of $Y$ given $L$ depends on $L$ through $Z^*$. So the influence function for $\hat{E}Y$ is also equal to

$$\frac{R}{E[R|L]}(Y - EY) - \frac{R - E[R|L]}{E[R|L]}(E[Y|L] - EY),$$

which is the efficient influence function for the parameter $EY$.

**B Proof of Theorem 4.1.**

We only consider the estimator $\hat{\beta}_n$ in the following. The arguments for the estimator $\hat{\gamma}_n$ is similar. Obviously, $\hat{\beta}_n$ maximizes

$$\tilde{L}_1^{(n)}(\beta) = \frac{1}{n} \sum_{i=1}^{n} R_i \beta' L_i - \frac{1}{n} \sum_{i=1}^{n} R_i \ln(\sum_{Y_i \geq Y} e^{\beta' L_i}).$$

For $\beta$ in any compact set, it is easy to show by using the Glivenko-Cantelli theorem that $\tilde{L}_1^{(n)}(\beta)$ converges to a function $\tilde{L}_1(\beta)$ uniformly in $\beta$, where

$$\tilde{L}_1(\beta) = \mathbb{P}[R\beta' L - R \ln E[I_{Y \geq y} e^{\beta' L}]|_{y=Y}].$$

The similar uniform convergence also holds for their first and second derivatives with respect to $\beta$. Moreover, $\tilde{L}_1(\beta)$ is a strictly concave function since

$$\nabla_{\beta \beta}^2 \tilde{L}_1(\beta) = \mathbb{P}\{R(E[I_{Y \geq y} L_L e^{\beta' L}] - E[I_{Y \geq y} L e^{\beta' L}]E[I_{Y \geq y} L e^{\beta' L}])|_{y=Y}|_{y=Y} \}
\leq 0.$$

Let $\beta^*$ be the unique maximizer of the limit function $\tilde{L}_1(\beta)$. Then from Theorem II.1 and Corollary II.2 in Andersen and Gill (1982), under the regularity conditions in Assumptions (A2) and (A3), we obtain

$$\hat{\beta}_n \rightarrow^P \beta^*.$$

Hence,

$$0 = \nabla_{\beta} \tilde{L}_1^{(n)}(\hat{\beta}_n) = \nabla_{\beta} \tilde{L}_1^{(n)}(\beta^*) + \nabla_{\beta \beta}^2 \tilde{L}_1^{(n)}(\beta^{**}_n)(\hat{\beta}_n - \beta^*)$$

where $\beta^{**}_n$ is on the line between $\beta^*$ and $\hat{\beta}_n$. Moreover,

$$\nabla_{\beta \beta}^2 \tilde{L}_1^{(n)}(\beta^{**}_n) \rightarrow \nabla_{\beta \beta}^2 \tilde{L}_1(\beta^*)$$

36
in probability. If define

\[ \mu_n(y; \beta) = P_n[I_{Y \geq y e^{\beta^* L}}], \mu(y; \beta) = P[I_{Y \geq y e^{\beta^* L}}], \]

then

\[
\nabla_\beta \tilde{L}_1^{(n)}(\beta^*)
= (P_n - P)[RL - R \frac{\nabla_\beta \mu_n(Y; \beta^*)}{\mu_n(Y; \beta^*)}]
+ P[RL - R \frac{\nabla_\beta \mu_n(Y; \beta^*)}{\mu_n(Y; \beta^*)}] - P[RL - R \frac{\nabla_\beta \mu(Y; \beta^*)}{\mu(Y; \beta^*)}]
+ P[RL - R \frac{\nabla_\beta \mu(Y; \beta^*)}{\mu(Y; \beta^*)}],
\]

Since \( \nabla_\beta \tilde{L}_1(\beta^*) = 0, \)

\[ P[RL - R \frac{\nabla_\beta \mu(Y; \beta^*)}{\mu(Y; \beta^*)}] = 0. \]

We notice that \( \mu_n(y; \beta^*) \rightarrow \mu(y; \beta^*) \) and \( \nabla_\beta \mu_n(y; \beta^*) \rightarrow \nabla_\beta \mu(y; \beta^*) \) uniformly in \( y \) by the Glivenko-Cantelli theorem. So

\[
\nabla_\beta \tilde{L}_1^{(n)}(\beta^*)
= (P_n - P)[RL - R \frac{\nabla_\beta \mu_n(Y; \beta^*)}{\mu_n(Y; \beta^*)}]
- P[RL - R \frac{\nabla_\beta \mu_n(Y; \beta^*)}{\mu_n(Y; \beta^*)}]
+ P[RL - R \frac{\nabla_\beta \mu_n(Y; \beta^*)}{\mu_n(Y; \beta^*)}] - P[RL - R \frac{\nabla_\beta \mu(Y; \beta^*)}{\mu(Y; \beta^*)}]
+ P[RL - R \frac{\nabla_\beta \mu(Y; \beta^*)}{\mu(Y; \beta^*)}]
= (P_n - P) f_n^1(L, Y, R; \beta^*),
\]

where,

\[
f_n^1(l, y, r; \beta^*)
= rl - r \frac{P_n[I_{y \leq Y} L e^{\beta^* L}]}{P[I_{y \leq y} e^{\beta^* L}]} - \frac{e^{\beta^* l} P[I_{Y \leq y}]}{P[I_{y \leq y} e^{\beta^* L}]} - \frac{P[I_{Y \leq y} e^{\beta^* L}]}{P[I_{y \leq y} L e^{\beta^* L}]} + \frac{P[I_{Y \leq y} e^{\beta^* L}]}{P[I_{y \leq y} L e^{\beta^* L}]}. \]

37
Therefore,

$$\sqrt{n}(\hat{\beta}_n - \beta^*) = -\{\nabla^2_{\beta} L_1(\beta^*)\}^{-1} \sqrt{n}(\mathbb{P}_n - \mathbb{P}) f_1^*(L, Y; \beta^*) + o_p(1).$$

Define a class of functions

$$\mathcal{F}(\epsilon) = \{rl - r\frac{\xi_1(y)}{\xi_2(y)} - le^{\beta^* L} \mathbb{P}[R I_{Y \leq y} | \frac{\xi_2(Y)}{\xi_2(y)}] + e^{\beta^* L} \mathbb{P}[R I_{Y \leq y} \mathbb{P}[|y' \leq \gamma L|_{y' = Y}] :$$

$$\xi_1(y), \xi_2(y) \text{ are monotone in } y \text{ and bounded by } 1/\epsilon;$$

$$\xi_2(y) > \epsilon > 0\}.$$ 

Then for \(\epsilon\) small enough, \(f^*_1(l, y, r; \beta^*) \in \mathcal{F}(\epsilon)\). Since any function in \(\mathcal{F}(\epsilon)\) is Lipschitz with respect to \(\xi_1(y)\) and \(\xi_2(y)\) in \(L_2\)-norm, \(\mathcal{F}(\epsilon)\) is a Donsker class. We thus apply the Donsker theorem and obtain

$$\sqrt{n}(\hat{\beta}_n - \beta^*) = -\{\nabla^2_{\beta} L_1(\beta^*)\}^{-1} \sqrt{n}(\mathbb{P}_n - \mathbb{P}) f_1(L, Y; \beta^*) + o_p(1),$$

where

$$f_1(l, y, r; \beta^*) = rl - r\frac{\mathbb{P}[R I_{y \leq L}] - le^{\beta^* L} \mathbb{P}[R I_{y \leq L} \mathbb{P}[|y' \leq \gamma L|_{y' = Y}]}{\mathbb{P}[R I_{y \leq L} \mathbb{P}[|y' \leq \gamma L|_{y' = Y}]}. $$

The theorem is proved.

**C Proof of Theorem 4.2.**

**Definition C.1** Recall \(Z^* = (\beta^* L, \gamma^* L)\) and \(\hat{Z} = (\hat{\beta}_n L, \hat{\gamma}_n L)\). One compact set \(Q\) is called a compact interior subset of \(Z^*\) if one of the following cases is true:

**Case 1.** Both \(\beta^* L\) and \(\gamma^* L\) are discrete. Then \(Q\) contains all the discrete values of \(Z^*\).

**Case 2.** \(\beta^* L\) is discrete and takes values in a finite set \(Q_1\); but \(\gamma^* L\) is continuous with a support \(Q_2\). Then \(Q = Q_1 \times Q_2\), where \(Q_2\) is a compact set and is in the interior of \(Q_2\).

**Case 3.** \(\gamma^* L\) is discrete and takes values in a finite set \(Q_2\); but \(\beta^* L\) is continuous with a support \(Q_1\). Then \(Q = Q_1 \times Q_2\), where \(Q_1\) is a compact set and is in the interior of \(Q_1\).
Case 4. Both $\beta' L$ and $\gamma' L$ are continuous. Then $Q$ is a compact set and is in the interior of the support of $Z^*$.

In addition, we define a norm $\| \cdot \|_{[0,\tau] \times Q}$ as

$$\|g(t,z)\|_{[0,\tau] \times Q} = \sup_{t \in [0,\tau], z \in Q} |g(t,z)|.$$

We assume one of the working models is correct so $T$ and $C$ are independent given $Z^*$ from Lemma 4.1. The whole proof consists of four steps: in the first step, we show the consistency of $\hat{H}T|Z^* (t|z)$ thus $\hat{S}_n(t)$; then we write $\sqrt{n}(\hat{S}_n(t) - S(t))$ by using the empirical processes; in the third step, we apply the Donsker theorem to $\sqrt{n}(\hat{S}_n(t) - S(t))$ to obtain the asymptotic properties; finally, we show one of the last two terms in the expression of $A()$ is zero.

First, the consistency of $\hat{H}T|Z^* (t|z)$ and $\hat{S}_n(t)$ can be seen from the following lemmas.

**Lemma C.1** For any compact interior set $I$ of $Z^*$,

$$\left\| \frac{P_n[K(Z_0^* - z\alpha_n)Y \geq t]}{P_n[K(Z_0^*-z\alpha_n)]} - P(Y \geq t | Z^* = z) \right\|_{[0,\tau] \times Q} \rightarrow P 0,$$

$$\left\| \frac{P_n[K(Z_0^*-z\alpha_n)Y \leq tR]}{P_n[K(Z_0^*-z\alpha_n)]} - P(Y \leq t, R = 1 | Z^* = z) \right\|_{[0,\tau] \times Q} \rightarrow P 0.$$

**Proof.**

Let $z = (z_1, z_2)$ and we only need to verify the first result.

For Case 1 in which both $\beta' L$ and $\gamma' L$ are discrete,

$$\frac{P_n[K(Z_0^*-z\alpha_n)Y \geq t]}{P_n[K(Z_0^*-z\alpha_n)]} = \frac{P_n[I_{Z^*=z} Y \geq t]}{P_n[I_{Z^*=z}]},$$

in probability uniformly in $t \in [0, \tau]$. The result holds.

For Case 2, $Q = Q_1 \times Q_2'$,

$$\frac{P_n[K(Z_0^*-z\alpha_n)Y \geq t]}{P_n[K(Z_0^*-z\alpha_n)]} = \frac{P_n[I_{\beta^* L = z_1} K(0, \frac{\gamma' L - z_2}{\alpha_n}) Y \geq t]}{P_n[I_{\beta^* L = z_1} K(0, \frac{\gamma' L - z_2}{\alpha_n})]}.$$
By using the result from Dabrowska (1987), if $Q'_2$ is a rectangle, the above expression converges uniformly in $z_2 \in Q'_2$ in probability to

$$\frac{P(Y \geq t, \beta' L = z_1 | \gamma' L = z_2)}{P(\beta' L = z_1 | \gamma' L = z_2)} = P(Y \geq t | Z = z).$$

For any compact $Q'_2$, the convergence holds since $Q'_2$ can be covered by a finite number of rectangles.

The result for Case 3 can be shown to hold similar to Case 2.

In Case 4, by Dabrowska (1987), the result holds when $Q$ is a rectangle. However, for any compact interior subset $Q$ of $Z^*$, since $Q$ can be covered by a finite number of such rectangles, the result holds as well.

**Lemma C.2** For any compact interior set $Q$ of $Z^*$,

$$\| \frac{P_n[K(\frac{\hat{z} + a}{a_n})I_{Y \geq t}]}{P_n[K(\frac{\hat{z} - a}{a_n})]} - P(Y \geq t | Z^* = z)\|_{[0, \tau] \times Q} \to^{P} 0,$$

$$\| \frac{P_n[K(\frac{\hat{z} + a}{a_n})I_{Y \leq t R}]}{P_n[K(\frac{\hat{z} - a}{a_n})]} - P(Y \leq t, R = 1 | Z^* = z)\|_{[0, \tau] \times Q} \to^{P} 0.$$

**Proof.**

We only prove the first result holds.

Denote

$$g(\beta, \gamma) = \frac{1}{a_n^2} P_n[K(\frac{\beta' L, \gamma' L - z}{a_n})I_{Y \geq t}].$$

For any large constant $M$ and any pairs $(\beta_1, \gamma_1)$, $(\beta_2, \gamma_2)$ such that

$$|\beta_j - \beta^*| \leq \frac{M}{\sqrt{n}}, |\gamma_j - \gamma^*| \leq \frac{M}{\sqrt{n}},$$

we have that there exist another pair $(\tilde{\beta}, \tilde{\gamma})$, which is on the segment between $(\beta_1, \gamma_1)$ and $(\beta_2, \gamma_2)$, so that

$$g(\beta_1, \gamma_1) - g(\beta_2, \gamma_2)$$

$$= \frac{1}{a_n^2} P_n[\nabla_\beta K(\frac{\tilde{\beta}' L, \tilde{\gamma}' L - z}{a_n})I_{Y \geq t}] (\beta_1 - \beta_2)$$

$$+ \frac{1}{a_n^2} P_n[\nabla_\gamma K(\frac{\tilde{\beta}' L, \tilde{\gamma}' L - z}{a_n})I_{Y \geq t}] (\gamma_1 - \gamma_2)$$

$$\leq 2CM \frac{1}{\sqrt{n}a_n} g(\tilde{\beta}, \tilde{\gamma})$$

$$\leq 2CM \frac{1}{\sqrt{n}a_n} g(\beta, \gamma).$$

$$\leq \frac{2CM}{\sqrt{n}a_n} \max_{|\beta - \beta^*| \leq \frac{M}{\sqrt{n}}, |\gamma - \gamma^*| \leq \frac{M}{\sqrt{n}}} g(\beta, \gamma).$$
\[ \max_{|\beta - \beta^*| \leq \frac{M}{\sqrt{n}}, |\gamma - \gamma^*| \leq \frac{M}{\sqrt{n}}} g(\beta, \gamma) - \min_{|\beta - \beta^*| \leq \frac{M}{\sqrt{n}}, |\gamma - \gamma^*| \leq \frac{M}{\sqrt{n}}} g(\beta, \gamma) \leq 2CM \]

Thus,

\[ (1 - \frac{2CM}{\sqrt{n}a_n}) \max_{|\beta - \beta^*| \leq \frac{M}{\sqrt{n}}, |\gamma - \gamma^*| \leq \frac{M}{\sqrt{n}}} g(\beta, \gamma) \leq \min_{|\beta - \beta^*| \leq \frac{M}{\sqrt{n}}, |\gamma - \gamma^*| \leq \frac{M}{\sqrt{n}}} g(\beta, \gamma). \]

We obtain

\[ \max_{|\beta - \beta^*| \leq \frac{M}{\sqrt{n}}, |\gamma - \gamma^*| \leq \frac{M}{\sqrt{n}}} g(\beta, \gamma) - \min_{|\beta - \beta^*| \leq \frac{M}{\sqrt{n}}, |\gamma - \gamma^*| \leq \frac{M}{\sqrt{n}}} g(\beta, \gamma) \leq \frac{2CM}{\sqrt{n}a_n(1 - \frac{M}{\sqrt{n}})}. \]

Therefore,

\[
P(\left\| \frac{1}{a_n^2} P_n[K(\frac{\hat{Z} - z}{a_n})I_{Y \geq t}] - \frac{1}{a_n^2} P_n[K(\frac{Z^* - z}{a_n})I_{Y \geq t}] \right\|_{[0, \tau]} > \epsilon) \
\leq P(\left\| \frac{1}{a_n^2} P_n[K(\frac{\hat{Z} - z}{a_n})I_{Y \geq t}] - \frac{1}{a_n^2} P_n[K(\frac{Z^* - z}{a_n})I_{Y \geq t}] \right\|_{[0, \tau]} > \epsilon, |\hat{\beta}_n - \beta^*| \leq \frac{M}{\sqrt{n}}) \\
+ P(\left\| \frac{1}{a_n^2} P_n[K(\frac{\hat{Z} - z}{a_n})I_{Y \geq t}] - \frac{1}{a_n^2} P_n[K(\frac{Z^* - z}{a_n})I_{Y \geq t}] \right\|_{[0, \tau]} > \epsilon, |\hat{\gamma}_n - \gamma^*| > \frac{M}{\sqrt{n}}) \\
\leq P\left(\frac{2CM}{\sqrt{n}a_n(1 - \frac{M}{\sqrt{n}})} \left\| \min_{|\beta - \beta^*| \leq \frac{M}{\sqrt{n}}, |\gamma - \gamma^*| \leq \frac{M}{\sqrt{n}}} \frac{1}{a_n^2} P_n[K(\frac{(\beta' L, \gamma' L) - z}{a_n})I_{Y \geq t}] \right\|_{[0, \tau]} > \epsilon\right) \\
+ P(|\hat{\beta}_n - \beta^*| > \frac{M}{\sqrt{n}}) + P(|\hat{\gamma}_n - \gamma^*| > \frac{M}{\sqrt{n}}) \\
\leq P\left(\frac{2CM}{\sqrt{n}a_n(1 - \frac{M}{\sqrt{n}})} \left\| \frac{1}{a_n^2} P_n[K(\frac{Z^* - z}{a_n})I_{Y \geq t}] \right\|_{[0, \tau]} > \epsilon\right) \\
+ P(|\hat{\beta}_n - \beta^*| > \frac{M}{\sqrt{n}}) + P(|\hat{\gamma}_n - \gamma^*| > \frac{M}{\sqrt{n}}). 
\]

The last two terms converge to zero as \( M, n \) go to infinity. For the first term, from the previous lemma, it goes to zero as \( n \) goes to infinity. The result of Lemma C.2 holds.

**Lemma C.3** Denote

\[ \hat{H}_{T|Z^*}(t|z) = \int_0^t d_s P_n[K(\frac{\hat{Z} - z}{a_n})RI_{Y \leq s}] P_n[K(\frac{\hat{Z} - z}{a_n})I_{Y \geq s}], \]

41
and

\[ \hat{S}_{T|Z^*}(t|z) = \prod_{s \leq t} (1 - \hat{H}_{T|Z^*}({s}|z)). \]

Then for any interior compact set \( Q \) of \( Z^* \), in probability,

\[ \| \hat{H}_{T|Z^*}(t|z) - H_{T|Z^*}(t|z)\|_{[0,\tau] \times Q} \to 0, \]

and

\[ \| \hat{S}_{T|Z^*}(t|z) - S_{T|Z^*}(t|z)\|_{[0,\tau] \times Q} \to 0. \]

**Proof.**

The first result follows from Lemma C.2 and the following inequality:

\[
\begin{align*}
\| \hat{H}_{T|Z^*}(t|z) - H_{T|Z^*}(t|z)\|_{[0,\tau] \times Q} &= \left\| \int_0^t \frac{dP_n[K(\frac{\hat{Y}_{\tau}}{an})]_{Y \leq s}}{P_n[K(\frac{\hat{Y}_{\tau}}{an})]_{Y \geq s}} - \int_0^t \frac{dE[RI_{Y \leq s}|Z^*]}{E[I_{Y \geq s}|Z^*]} \right\|_{[0,\tau] \times Q} \\
&\leq \left\| \frac{P_n[K(\frac{\hat{Y}_{\tau}}{an})]_{Y \leq s}}{P_n[K(\frac{\hat{Y}_{\tau}}{an})]_{Y \geq s}} - P(Y \leq t, R = 1|Z^* = z) \right\|_{[0,\tau] \times Q} \\
&\quad + \left\| \frac{P_n[K(\frac{\hat{Y}_{\tau}}{an})]_{Y \geq s}}{P_n[K(\frac{\hat{Y}_{\tau}}{an})]_{Y \geq s}} - P(Y \geq t|Z^* = z) \right\|_{[0,\tau] \times Q}.
\end{align*}
\]

For the second result, we use the Duhamel equation and integration by part: for any \( t \in [0, \tau] \) and \( z \in Q \),

\[
\begin{align*}
|\hat{S}_{T|Z^*}(t|z) - S_{T|Z^*}(t|z)| &= |S_{T|Z^*}(t|z) \int_0^t \frac{\hat{S}_{T|Z^*}(u - |z|)}{S_{T|Z^*}(u|z)} d(\hat{H}_{T|Z^*}(u|z) - H_{T|Z^*}(u|z))| \\
&= |\hat{S}_{T|Z^*}(t - |z|)(\hat{H}_{T|Z^*}(t|z) - H_{T|Z^*}(t|z)) \\
&\quad - \int_0^t (\hat{H}_{T|Z^*}(u|z) - H_{T|Z^*}(u|z))(\frac{d\hat{S}_{T|Z^*}(u - |z|)}{S_{T|Z^*}(u|z)} - \frac{\hat{S}_{T|Z^*}(u - |z|) dS_{T|Z^*}(u|z)}{S_{T|Z^*}(u|z)^2})| \\
&\leq (1 + \frac{2}{\min_{z \in Q} P(T > \tau|Z^* = z)^2}) \max_{0 \leq s \leq t} |\hat{H}_{T|Z^*}(s|z) - H_{T|Z^*}(s|z)|.
\end{align*}
\]

Second, we will write \( \hat{H}_{T|Z^*}(t|z) - H_{T|Z^*}(t|z) \) and \( \hat{S}_n(t) - S(t) \) in terms of the empirical process of \( (P_n - P) \).

\[ \hat{H}_{T|Z^*}(t|z) - H_{T|Z^*}(t|z) \]
$$\begin{align*}
&= \mathbf{P}_n \left[ \frac{1}{a_n^2} K \left( \frac{Z - z}{a_n} \right) I_{Y \leq t R} \right] - H_{T|Z}(t|z) \\
&= (\mathbf{P}_n - \mathbf{P}) \left[ \frac{1}{a_n^2} K \left( \frac{Z - z}{a_n} \right) I_{Y \leq t R} \right] \\
&\quad + \mathbf{P} \left[ \frac{1}{a_n^2} K \left( \frac{Z - z}{a_n} \right) I_{Y \geq t R} \right] - H_{T|Z}(t|z) \\
&= (\mathbf{P}_n - \mathbf{P}) \left[ \frac{1}{a_n^2} K \left( \frac{Z - z}{a_n} \right) I_{Y \geq t R} \right] \\
&\quad - \mathbf{P} \left[ \frac{1}{a_n^2} K \left( \frac{Z - z}{a_n} \right) I_{Y \leq t R} \right] (\mathbf{P}_n - \mathbf{P}) \left[ \frac{1}{a_n^2} K \left( \frac{Z - z}{a_n} \right) I_{Y \geq t R} \right] \\
&\quad + \mathbf{P} \left[ \frac{1}{a_n^2} K \left( \frac{Z - z}{a_n} \right) I_{Y \geq t R} \right] - H_{T|Z}(t|z) \\
&= (I) + (II) + (III).
\end{align*}$$

Let $f_{R,Y,Z}$ be the joint density of $(R,Y,Z)$ and let $f_{Y,Z}$ be the joint density of $(Y,Z)$. Then for $(III)$, since uniformly in $z \in Q$ ($Q$ is any interior compact subset of $Z^*$) and $y \in [0,\tau]$, 

$$\begin{align*}
&\mathbf{P} \left[ \frac{1}{a_n^2} K \left( \frac{Z - z}{a_n} \right) I_{Y \geq y} \right] \\
&= \frac{1}{a_n^2} \int_y \int_v f_{Y,Z}(u,v) K \left( \frac{Z - v}{a_n} \right) dudv \\
&= \int_y \int_v f_{Y,Z}(u,z + a_n \bar{v}) K(\bar{v}) d\bar{v} du \\
&= \int_y [f_{Y,Z}(u,z) + a_n \sum_{j=1}^2 f_{Y,Z}(u,z) \int_{\bar{v}} \bar{v}_j K(\bar{v}) d\bar{v}] du + O_p(a_n^2) \\
&= \int_y f_{Y,Z}(u,z) du + O_p(a_n^2),
\end{align*}$$

$$(III) = \mathbf{P} \left[ \frac{1}{a_n^2} K \left( \frac{Z - z}{a_n} \right) I_{Y \leq t R} \right] \\
= \int_0^t \int_{f(Y \geq u \mid Z = z)} f_{R,Y,Z}(u,z) du + O_p(a_n^2) - H_{T|Z}(t|z) \\
= \int_0^t d_u P(R = 1, Y \leq u | Z = z) + O_p(a_n^2) - H_{T|Z}(t|z) \\
= \int_0^t d_u P(R = 1, Y \leq u | Z^* = z) + O_p(a_n^2) - H_{T|Z}(t|z) \\
= \int_0^t d_u P(Y \leq u | R = 1, (\beta' L, \gamma^{*'} L) = z) + \nabla_{\beta} |_{\beta = \beta^*} \int_0^t d_u P(Y \leq u | R = 1, (\beta' L, \gamma^{*'} L) = z) |(\hat{\beta}_n - \beta^*)
\]
Hence, by Lemma 4.1,
\[ \int_0^t \frac{d_u P(R = 1, Y \leq u | Z^* = z)}{P(Y \geq u | (\beta' L, \gamma' L) = z)} = H_{T|Z^*}(t|z). \]
Hence,
\[(III) = \nabla_\beta |_{\beta = \beta^*} \left[ \int_0^t \frac{d_u P(Y \leq u, R = 1 | (\beta' L, \gamma' L) = z)}{P(Y \geq u | (\beta' L, \gamma' L) = z)} (\hat{\beta}_n - \beta^*) \right. \]
\[+ \nabla_\gamma |_{\gamma = \gamma^*} \left[ \int_0^t \frac{d_u P(Y \leq u, R = 1 | (\beta' L, \gamma' L) = z)}{P(Y \geq u | (\beta' L, \gamma' L) = z)} (\hat{\gamma}_n - \gamma^*) \right. \]
\[+ O_p(a_n^2) + O_p(\|\hat{\beta}_n - \beta^*\|^2 + |\hat{\gamma}_n - \gamma^*|^2). \]

Furthermore, we define
\[ h^n_1(y, r, l; \beta, \gamma, t, z) = \frac{1}{a_n^2} K\left(\frac{\beta' L - z}{a_n}\right) I_{Y \leq \ell'} P_n \frac{1}{a_n^2} K\left(\frac{\beta' L - z}{a_n}\right) I_{Y \leq r} P_n \frac{1}{a_n} K\left(\frac{\beta' L - z}{a_n}\right) I_{Y \leq \gamma} P_n \frac{1}{a_n} K\left(\frac{\beta' L - z}{a_n}\right) I_{Y \geq z} P_n \frac{1}{a_n} K\left(\frac{\beta' L - z}{a_n}\right) I_{Y \geq y} P_n \frac{1}{a_n} K\left(\frac{\beta' L - z}{a_n}\right) I_{Y \geq z}. \]

and define \( h_2(y, l; \beta, \gamma, t, z) \) as
\[ \frac{1}{a_n^2} K\left(\frac{\beta' L - z}{a_n}\right) P_n \frac{1}{a_n} K\left(\frac{\beta' L - z}{a_n}\right) I_{Y \leq l} I_{Y \leq y} R \frac{1}{a_n} K\left(\frac{\beta' L - z}{a_n}\right) I_{Y \geq z} P_n \frac{1}{a_n} K\left(\frac{\beta' L - z}{a_n}\right) I_{Y \geq y} P_n \frac{1}{a_n} K\left(\frac{\beta' L - z}{a_n}\right) I_{Y \geq z}. \]

So we have that
\[ \hat{H}_{T|Z^*}(t|z) - H_{T|Z^*}(t|z) \]
\[= (P_n - P) h^n_1(Y, R, L; \hat{\beta}_n, \hat{\gamma}_n, t, z) \]
\[= -(P_n - P) h^n_2(Y, L; \hat{\beta}_n, \hat{\gamma}_n, t, z) \]
\[+ \nabla_\beta B(\beta^*, \gamma^*, z, t) (\hat{\beta}_n - \beta^*) \]
\[+ \nabla_\gamma B(\beta^*, \gamma^*, z, t) (\hat{\gamma}_n - \gamma^*) \]
\[+ O_p(a_n^2) + O_p\left(\frac{1}{n}\right) \]

uniformly in \( t \in [0, \tau] \) and \( z \in \mathcal{Q} \), where
\[ B(\beta, \gamma, z, t) = \int_0^t \frac{d_u P(Y \leq u, R = 1 | (\beta' L, \gamma' L) = z)}{P(Y \geq u | (\beta' L, \gamma' L) = z)}. \]

Hence,
\[ \hat{S}_{T|Z^*}(t|z) - S_{T|Z^*}(t|z) \] (3)
To further simplify the above expression, we first notice that by using the equation (C.3), the first two

\[
\sqrt{n}(\hat{S}_n(t) - S(t))
\]

\[= \sqrt{n}(P_n - P)[\hat{S}_{T|Z^*}(t|Z)] - P[\hat{S}_{T|Z^*}(t|Z)]
\]

\[= \sqrt{n}(P_n - P)[\hat{S}_{T|Z^*}(t|Z)] + \sqrt{n}P[\hat{S}_{T|Z^*}(t|Z) - S(t)]
\]

\[= \sqrt{n}(P_n - P)[\hat{S}_{T|Z^*}(t|Z)I_{Z^* \in Q}] + \sqrt{n}P[(\hat{S}_{T|Z^*}(t|Z) - S(t))I_{Z^* \in Q}]
\]

\[+ \sqrt{n}(P_n - P)[\hat{S}_{T|Z^*}(t|Z)I_{Z^* \notin Q}] + \sqrt{n}P[(\hat{S}_{T|Z^*}(t|Z) - S(t))I_{Z^* \notin Q}].
\]

uniformly in \(t \in [0, \tau]\) and \(z \in Q\).

For any \(\epsilon\), by the definition of the interior compact set, we can choose such a set \(Q\) of \(Z^*\) that

\[P(Z^* \notin Q) < \epsilon.
\]

Hence,

\[\sqrt{n}(\hat{S}_n(t) - S(t))
\]

\[= \sqrt{n}(P_n - P)[\hat{S}_{T|Z^*}(t|Z) - P[\hat{S}_{T|Z^*}(t|Z)]
\]

\[= \sqrt{n}(P_n - P)[\hat{S}_{T|Z^*}(t|Z)] + \sqrt{n}P[\hat{S}_{T|Z^*}(t|Z) - S(t)]
\]

\[= \sqrt{n}(P_n - P)[\hat{S}_{T|Z^*}(t|Z)I_{Z^* \in Q}] + \sqrt{n}P[(\hat{S}_{T|Z^*}(t|Z) - S(t))I_{Z^* \in Q}]
\]

\[+ \sqrt{n}(P_n - P)[\hat{S}_{T|Z^*}(t|Z)I_{Z^* \notin Q}] + \sqrt{n}P[(\hat{S}_{T|Z^*}(t|Z) - S(t))I_{Z^* \notin Q}].
\]

To further simplify the above expression, we first notice that by using the equation (C.3), the first two
terms in the above expression can be written as

\[\sqrt{n}(P_n - P)[\hat{S}_{T|Z^*}(t|Z)I_{Z^* \in Q}] + \sqrt{n}P[(\hat{S}_{T|Z^*}(t|Z) - S(t))I_{Z^* \in Q}]
\]

\[= \sqrt{n}(P_n - P)[\hat{S}_{T|Z^*}(t|Z)I_{Z^* \in Q}]
\]

\[+ \sqrt{n}(P_n - P)g_1^n(Y, R, L; \hat{\beta}_n, \hat{\gamma}_n, t, Q)
\]

\[- \sqrt{n}(P_n - P)g_2^n(Y, L; \hat{\beta}_n, \hat{\gamma}_n, t, Q)
\]

\[-P[I_{Z^* \in Q}S_{T|Z^*}(t|Z)] \int_0^t \frac{\hat{S}_{T|Z^*}(u - |Z)}{S_{T|Z^*}(u|Z)} d_u \nabla \beta B(\beta^*, \gamma^*, \hat{Z}, u)]\sqrt{n}(\beta_n - \beta^*)
\]

\[-P[I_{Z^* \notin Q}S_{T|Z^*}(t|Z)] \int_0^t \frac{\hat{S}_{T|Z^*}(u - |Z)}{S_{T|Z^*}(u|Z)} d_u \nabla \gamma B(\beta^*, \gamma^*, \hat{Z}, u)]\sqrt{n}(\gamma_n - \gamma^*)
\]

\[+ O_p(\sqrt{n}a_2^n) + o_p(1).
\]
where
\[
g_1^n(y, r, l; \hat{\beta}_n, \hat{\gamma}_n, t, Q) = -P[I_{Z^* \in Q}S_{T|Z^*}(t)\hat{Z} \int_0^t \frac{\hat{S}_{T|Z^*}(u - |\hat{Z}|)}{S_{T|Z^*}(u)} dh_1^n(y, r, l; \hat{\beta}_n, \hat{\gamma}_n, u, \hat{Z})]
\]
and
\[
g_2^n(y, l; \hat{\beta}_n, \hat{\gamma}_n, t, Q) = -P[I_{Z^* \in Q}S_{T|Z^*}(t)\hat{Z} \int_0^t \frac{\hat{S}_{T|Z^*}(u - |\hat{Z}|)}{S_{T|Z^*}(u)} dh_2^n(y, l; \hat{\beta}_n, \hat{\gamma}_n, u, \hat{Z})],
\]
So if define \(w_n(Y, R, L; \hat{\beta}_n, \hat{\gamma}_n, t, Q)\) as
\[
\hat{S}_{T|Z^*}(t)\hat{Z}I_{Z^* \in Q} - S(t)I_{Z^* \in Q} + g_1^n(Y, R, L; \hat{\beta}_n, \hat{\gamma}_n, t, Q) - g_2^n(Y, L; \hat{\beta}_n, \hat{\gamma}_n, t, Q),
\]
then since by Lemma C.2 and Lemma C.3,
\[
P[I_{Z^* \in Q}S_{T|Z^*}(t)\hat{Z} \int_0^t \frac{\hat{S}_{T|Z^*}(u - |\hat{Z}|)}{S_{T|Z^*}(u)} du \nabla_\beta B(\beta^*, \gamma^*, \hat{Z}, u)]
\]
\[
\rightarrow P[I_{Z^* \in Q}S_{T|Z^*}(t)Z^*] \nabla_\beta B(\beta^*, \gamma^*, Z^*, t),
\]
and
\[
P[I_{Z^* \in Q}S_{T|Z^*}(t)\hat{Z} \int_0^t \frac{\hat{S}_{T|Z^*}(u - |\hat{Z}|)}{S_{T|Z^*}(u)} du \nabla_\gamma B(\beta^*, \gamma^*, \hat{Z}, u)]
\]
\[
\rightarrow P[I_{Z^* \in Q}S_{T|Z^*}(t)Z^*] \nabla_\gamma B(\beta^*, \gamma^*, Z^*, t),
\]
we obtain that for any fixed positive constant \(\delta > 0\),
\[
P(\|\sqrt{n}(\hat{S}_n(t) - S(t))
\]
\[
- \sqrt{n}(P_n - P)w_n(Y, R, L; \hat{\beta}_n, \hat{\gamma}_n, t, Q)
\]
\[
- P[I_{Z^* \in Q}S_{T|Z^*}(t)Z^*] \nabla_\beta B(\beta^*, \gamma^*, Z^*, t)\sqrt{n}(\hat{\beta}_n - \beta^*)
\]
\[
- P[I_{Z^* \in Q}S_{T|Z^*}(t)Z^*] \nabla_\gamma B(\beta^*, \gamma^*, Z^*, t)\sqrt{n}(\hat{\gamma}_n - \gamma^*)|t \in [0, \tau] > \delta)
\]
\[
\leq 2P(Z^* \notin Q) \leq 2\epsilon
\]
as \(n\) is large enough.

Third, the Donsker theorem is applied to the above expression of \(\sqrt{n}(\hat{S}_n(t) - S(t))\) to obtain the asymptotic properties of \(\hat{S}_n(t)\).

For any large \(M\), we consider the empirical process
\[
\{\sqrt{n}(P_n - P)w_n(Y, R, L; \beta^* + \frac{\theta_1}{\sqrt{n}}, \gamma^* + \frac{\theta_2}{\sqrt{n}}, t, Q) : t \in [0, \tau], |\theta_1| \leq M, |\theta_2| \leq M\}
\]
46
which is indexed by \((t, \theta_1, \theta_2)\). First, we claim that uniformly in \(t\),

\[
    w_n(Y, R, L; \hat{\theta}_n, \hat{\gamma}_n, t, Q) \\
    \rightarrow I_{Z^* \in \Omega} [S_{T|Z^*}(t|Z^*) - S(t) - \frac{RI_{Y \leq t}S_{T|Z^*}(t|Z^*)}{P(Y \geq y'|Z^*)|y' = Y} \\
    + S_{T|Z^*}(t|Z^*) \int_0^{t \wedge Y} e^{H_{T|Z^*}(u|Z^*) + H_{C|Z^*}(u|Z^*)} dH_{T|Z^*}(u|Z^*) ]
\]

in probability. This is true because by using similar argument to the proof of Lemma C.2 and Lemma C.3, we have that in probability, uniformly in \(t \in [0, \tau]\),

\[
    g_1^n(y, r, l; \hat{\beta}_n, \hat{\gamma}_n, t, Q) \\
    \rightarrow -\frac{rI_{y \leq l}I_{(\beta^*, l, \gamma^*, l) \in Q} S_{T|Z^*}(t|Z^* = (\beta^* l, \gamma^* l))}{P(Y \geq y|Z^* = (\beta^* l, \gamma^* l))},
\]

\[
    g_2^n(y, l; \hat{\beta}_n, \hat{\gamma}_n, t, Q) \\
    \rightarrow -I_{(\beta^* l, \gamma^* l) \in Q} S_{T|Z^*}(t|Z^* = (\beta^* l, \gamma^* l)) \\
    \int_0^{t \wedge Y} e^{H_{T|Z^*}(u|Z^* = (\beta^* l, \gamma^* l)) + H_{C|Z^*}(u|Z^* = (\beta^* l, \gamma^* l))} d_u H_{T|Z^*}(u|Z^* = (\beta^* l, \gamma^* l)).
\]

Second, with careful checking, we can verify that each function \(w_n()\) indexed by \((t, \theta_1, \theta_2)\) belongs to \(BV[0, \tau]\) as a function of \(t\) and its variation on \([0, \tau]\) is uniformly bounded in \(\theta_1, \theta_2\). Moreover, each function is Lipschitz continuous with respect to \((\theta_1, \theta_2)\) and the Lipschitz coefficient is bounded by \(\frac{1}{\sqrt{n\alpha_n}}\) in probability. Thus we can apply the Donsker theorem for the changing class (Theorem 2.11.23, van der Vaart and Wellner (1996)) and obtain that, uniformly in \(t\),

\[
    \sqrt{n}(P_n - P)w_n(Y, R, L; \hat{\theta}_n, \hat{\gamma}_n, t, Q) \\
    = \sqrt{n}(P_n - P)w_n(Y, R, L; \hat{\theta}_n, \hat{\gamma}_n, t, Q)I_{|\hat{\beta}_n - \beta^*| \leq \frac{M}{\sqrt{n}}, |\hat{\gamma}_n - \gamma^*| \leq \frac{M}{\sqrt{n}} } \\
    + \sqrt{n}(P_n - P)w_n(Y, R, L; \hat{\theta}_n, \hat{\gamma}_n, t, Q)I_{|\hat{\beta}_n - \beta^*| \geq \frac{M}{\sqrt{n}} } \\
    + \sqrt{n}(P_n - P)w_n(Y, R, L; \hat{\theta}_n, \hat{\gamma}_n, t, Q)I_{|\hat{\gamma}_n - \gamma^*| \geq \frac{M}{\sqrt{n}} } \\
    = \sqrt{n}(P_n - P)\{I_{Z^* \in \Omega} [S_{T|Z^*}(t|Z^*) - S(t) - \frac{RI_{Y \leq t}S_{T|Z^*}(t|Z^*)}{P(Y \geq y'|Z^*)|y' = Y} \\
    + S_{T|Z^*}(t|Z^*) \int_0^{t \wedge Y} e^{H_{T|Z^*}(u|Z^*) + H_{C|Z^*}(u|Z^*)} d_u H_{T|Z^*}(u|Z^*) ]
    + S_{T|Z^*}(t|Z^*) \int_0^{t \wedge Y} e^{H_{T|Z^*}(u|Z^*) + H_{C|Z^*}(u|Z^*)} d_u H_{T|Z^*}(u|Z^*) ]
    + \sqrt{n}(P_n - P)w_n(Y, R, L; \hat{\theta}_n, \hat{\gamma}_n, t, Q)I_{|\hat{\gamma}_n - \gamma^*| \geq \frac{M}{\sqrt{n}} } \\
    + \sqrt{n}(P_n - P)w_n(Y, R, L; \hat{\theta}_n, \hat{\gamma}_n, t, Q)I_{|\hat{\gamma}_n - \gamma^*| \geq \frac{M}{\sqrt{n}} }.
\]

If we let \(M\) be large enough so that the probability that the last two terms are nonzero is smaller than

\[
P(|\hat{\beta}_n - \beta^*| \geq \frac{M}{\sqrt{n}}) + P(|\hat{\gamma}_n - \gamma^*| \geq \frac{M}{\sqrt{n}}) \leq 2\epsilon,
\]

47
then

\[ P(\|\sqrt{n}(\hat{S}_n(t) - S(t))\|) \]

\[ - \sqrt{n}(P_n - P)\{I_{\mathcal{Z}^* \in \mathcal{Q}}[S_T|\mathcal{Z}^*(t)|Z^*] - S(t) - \frac{RI_{Y \leq t}S_T|\mathcal{Z}^*(t)|Z^*}{P(Y \geq y'|Z^*)}_{y'=\mathcal{Y}} \]

\[ + S_T|\mathcal{Z}^*(t)|Z^* \int_0^{t\mathcal{Y}} e^{H_{T|\mathcal{Z}^*}(u'|\mathcal{Z}^*) + H_{C|\mathcal{Z}^*}(u'|\mathcal{Z}^*)} du H_{T|\mathcal{Z}^*}(u'|\mathcal{Z}^*) \} \]

\[ - P[I_{\mathcal{Z}^* \in \mathcal{Q}}S_T|\mathcal{Z}^*(t)|Z^*]|\mathcal{Y} B(\beta^*, \gamma^*, Z^*, t)] \sqrt{n}(\hat{\beta}_n - \beta^*) \]

\[ - P[I_{\mathcal{Z}^* \in \mathcal{Q}}S_T|\mathcal{Z}^*(t)|Z^*]|\mathcal{Y} B(\beta^*, \gamma^*, Z^*, t)] \sqrt{n}(\hat{\gamma}_n - \gamma^*) \|_{t\in[0, \tau]} > \delta \]

\[ \leq 4\epsilon. \]

By choosing \( Q \) close enough to the support of \( Z^* \), we finally obtain that uniformly in \( t \in [0, \tau] \),

\[ \sqrt{n}(\hat{S}_n(t) - S(t)) \]

\[ = \sqrt{n}(P_n - P)[S_T|\mathcal{Z}^*(t)|Z^*] - S(t) - \frac{RI_{Y \leq t}S_T|\mathcal{Z}^*(t)|Z^*}{P(Y \geq y'|Z^*)}_{y'=\mathcal{Y}} \]

\[ + S_T|\mathcal{Z}^*(t)|Z^* \int_0^{t\mathcal{Y}} e^{H_{T|\mathcal{Z}^*}(u'|\mathcal{Z}^*) + H_{C|\mathcal{Z}^*}(u'|\mathcal{Z}^*)} du H_{T|\mathcal{Z}^*}(u'|\mathcal{Z}^*) \] \[ - P[I_{\mathcal{Z}^* \in \mathcal{Q}}S_T|\mathcal{Z}^*(t)|Z^*]|\mathcal{Y} B(\beta^*, \gamma^*, Z^*, t)] \sqrt{n}(\hat{\beta}_n - \beta^*) \]

\[ - P[I_{\mathcal{Z}^* \in \mathcal{Q}}S_T|\mathcal{Z}^*(t)|Z^*]|\mathcal{Y} B(\beta^*, \gamma^*, Z^*, t)] \sqrt{n}(\hat{\gamma}_n - \gamma^*) + o_p(1) \]

\[ = \sqrt{n}(P_n - P)A(t; \gamma, R, L) + o_p(1), \]

where

\[ A(t; \gamma, R, L) \]

\[ = e^{-H_{T|\mathcal{Z}^*}(t)|Z^*]} - S(t) \]

\[ - RI_{Y \leq t}e^{H_{T|\mathcal{Z}^*}(Y)|Z^*} + H_{C|\mathcal{Z}^*}(Y)|Z^*} - H_{T|\mathcal{Z}^*}(t)|Z^*} \]

\[ + \int_0^{t\mathcal{Y}} e^{H_{T|\mathcal{Z}^*}(u'|\mathcal{Z}^*) + H_{C|\mathcal{Z}^*}(u'|\mathcal{Z}^*)} du H_{T|\mathcal{Z}^*}(u'|\mathcal{Z}^*) \]

\[ - E[e^{-H_{T|\mathcal{Z}^*}(t)|Z^*}] \frac{d}{d\gamma}|_{\gamma=\gamma^*} \int_0^{t\mathcal{Y}} du P(T \land C \leq u, R = 1|\gamma L, \beta^* L) \frac{S_\gamma(\gamma^*, L, Y, R)}{P(T \land C > u|\gamma L, \beta^* L)} \]

\[ - E[e^{-H_{T|\mathcal{Z}^*}(t)|Z^*}] \frac{d}{d\beta}|_{\beta=\beta^*} \int_0^{t\mathcal{Y}} du P(T \land C \leq u, R = 1|\gamma^* L, \beta L) \frac{S_\beta(\beta^*, L, Y, R)}{P(T \land C > u|\gamma^* L, \beta L)}. \]

We thus obtain the consistency and the asymptotic Gaussian property of \( \hat{S}_n(t) \).

The final step, we will show one of the last two terms is zero when one of the working models is correct. When the working model of \( T \) given \( L \) is correct, since for any \( \gamma \),

\[ \int_0^t du P(T \land C \leq u, R = 1|\gamma L, \beta^* L) \frac{P(T \land C > u|\gamma L, \beta^* L)}{P(T \land C > u|\gamma L, \beta^* L)} = H_{T|\beta^* L, \gamma L}(t|\beta^* L, \gamma L), \]

\[ = 48 \]
we have
\[ E[e^{-H_{T|Z^*}(t|Z^*)} \frac{d}{d\gamma}|_{\gamma=\gamma^*} \int_0^t d_u \frac{P(T \wedge C \leq u, R = 1|\beta'^L, \gamma' L)}{P(T \wedge C \geq u|\beta'^L, \gamma' L)}] \]
\[ = -\frac{d}{d\gamma}|_{\gamma=\gamma^*} E e^{-H_{T|\beta'^L, \gamma' L}(t|\beta'^L, \gamma' L)} \]
\[ = -\frac{d}{d\gamma}|_{\gamma=\gamma^*} S(t) = 0. \]

Hence, the first one in the last two terms of \( A() \) is zero. Similarly, the other one is zero if the working model for \( C \) given \( L \) is correct. •

**Acknowledgments.**

This work is part of my Ph.D. Dissertation advised by Susan Murphy at the University of Michigan. I owe many thanks to her for numerous discussions and helpful comments.

**References**


