Sociedad de Estadística e Investigación Operativa
Test (2005) *Vol.* **14**, *No.* 2, pp. 489–514

Semiparametric Censorship Model with Covariates

Ming Yuan*
School of Industrial and Systems Engineering
Georgia Institute of Technology, USA

Abstract

In real applications, we may be confronted with the problem of informative censoring. Koziol-Green model is commonly used to model the possible information contained in the informative censoring. However the proportionality assumption cast by Koziol-Green model (see (1.2) below) is "too restrictive in that it limits the scope of the Cox model in practice" (see Subramanian, 2000). In this paper, we try to relax the proportionality condition of Koziol-Green model by modeling the censorship semiparametrically. It is shown that our suggested semiparametric censoring model is an applicable extension of the Koziol-Green model. Through a close connection with the logistic regression, our model assumptions are readily to be checked in practice. We also propose estimation for both the regression parameter and the cumulative baseline hazard function which can incorporate the additional information contained in the semiparametric censorship model. Simulations and the analysis of a real dataset confirm the applicability of the suggested model and estimation.

Key Words: Cox model, Koziol-Green model, maximum partial likelihood estimator, profiled likelihood, Breslow estimator.

AMS subject classification: Primary 62G05; Secondary 62E20.

1 Introduction

The Cox's proportional hazards model is the most commonly used model of regression analysis for censored data since its introduction (Cox, 1972). Suppose we observe the independent triples (X_i, δ_i, Z_i) for $i = 1, 2, ..., n$,

*Correspondence to: Ming Yuan. School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. E-mail: myuan@isye.gatech.edu

This research is partially supported by NSF Grant DMS-0772292

Received: September 2002; Accepted: March 2004

where X_i is the minimum of the survival time T_i of interest and the censoring time C_i, which is assumed to be independent of T_i when conditioned on covariates Z, and δ_i is the indicator of censoring, i.e., $\delta_i = 1$ if $T_i \leq C_i$ and 0 otherwise. We are interested in how the covariates affect the true survival time T given the observations. The Cox's proportional hazards model casts the following form of the conditional hazard of T.

$$\lambda_{T|Z}(t|z) = \exp(z'\beta)\,\lambda_0(t), \tag{1.1}$$

where λ_0 is the so-called baseline hazard and β is the coefficient which quantifies the relationship between T and Z.

A crucial assumption often made is the non-informativeness of censoring, but this is not always the case in real applications. In the absence of covariates, a commonly used technique to tackle the informative censoring is the Koziol-Green model (Koziol and Green, 1976). More specifically, this model assumes that the hazard of survival time T and the censoring time C satisfy

$$\lambda_C(x) = p\lambda_T(x) \tag{1.2}$$

for some constant $0 < p < 1$. Due to its success in applications, the Koziol-Green model was later extended to the situations when covariates are present. A common practice is to embed the Cox model in (1.2). Because of its ability to incorporate the informative censoring, this model has received some attention in the literature recently, *i.e.* de Uña Álvarez and González-Manteiga (1998, 1999); Subramanian (2000) and references therein. In this paper, we refer to the Koziol-Green model as the one with Cox model embedded in.

A major drawback of the Koziol-Green model is that (1.2) is "too restrictive in the sense that it limits the scope of Cox's model" (see Subramanian, 2000). In this article, we proposed an applicable extension of the Koziol-Green model with covariates – the proportionality factor is allowed to be a function known up to a finite–dimensional parameter. Our suggested model has at least the following merits:

(i) Although Koziol-Green model has its limitations, the proportionality function in the model is defined in a meaningful way. Our suggested model preserves the interpretability of the proportionality function. This also makes the parameters introduced by the semiparametric censoring model easy to interpret.

(ii) According to the rationale of the model formulation to be discussed in the next section, the proportionality function has a close connection with a binomial regression problem. Thus, by carefully narrowing down the parametric family of the proportionality functions, we can easily check the model assumption in practice.

(iii) Although Koziol-Green model with covariates is very restrictive, its efficient estimators can be easily implemented due to the work by Subramanian (2000). It is desirable to use it if the proportionality condition (1.2) does hold. But usually, this assumption is hardly found in real world. Our semiparametric model has Koziol-Green model as a special case and provides us a device to test the proportionality assumption.

Other proposals on generalizing the Koziol-Green model include the partial Koziol-Green model of Gather and Pawlitschko (1998) and the conditional proportional model introduced by Veraverbeke and Cadarso-Suárez (2000). The partial Koziol-Green model relies on the awareness of different mechanisms of censoring. It assumes that some censorings are informative and the others are not. The partial Koziol-Green model is very useful if we can identify the informative censorings. The conditional proportional model focuses more on the relationship between the censoring and the covariates. It assumes that the probability of being censored is solely determined by the covariates. It is a nice extension of the Koziol-Green model and can definitely find many applications in the real world.

Following the reasoning of Veraverbeke and Cadarso-Suárez (2000), in the next section, we proposed a semiparametric censoring model by examining the relationship between the proportionality function and the censoring. We also addressed how to use the standard statistical tools to test our model assumptions and the proportionality assumption. An issue accompanying the model is how to estimate the parameters. The information bound for estimating the parameters associated with the semiparametric censorship model was derived in Section 3. In Section 4, estimators which achieve these information bounds were provided. Technical proofs were deferred to the Appendix. Simulations presented in Section 5 gave us some idea how the new estimator incorporates the information contained in the informative censoring and how it compares with the maximum partial likelihood estimator if the model assumption is violated in certain way. A real dataset was also analyzed in Section 6 to demonstrate how to use our model in practice.

2 Semiparametric censoring model

As the first attempt to get around the limitations of the Koziol-Green model, the conditional proportional model was suggested. This model relaxes condition (1.2) by admitting p to be a function of Z. The conditional proportional model was introduced by Veraverbeke and Cadarso-Suárez (2000) as a generalization of the Koziol-Green model. The estimator of conditional cumulative hazard function of this model has also been addressed by Jensen and Wiedmann (2001).

The conditional proportional model was motived by a re-examination of a real dataset (see Veraverbeke and Cadarso-Suárez, 2000). In their paper, Veraverbeke and Cadarso-Suárez revealed that there is possible dependence between parameter p of the Koziol-Green model (see (1.2)) and the covariates Z and then they argued that the conditional proportional model should be the right way to incorporate this dependence. However, after a careful look at their argument, we found that although the conditional proportional model may make sense in some other applications, it is not the appropriate model for analyzing that specific dataset. The authors of that paper considered the following logistic regression between δ and (X, Z)

$$\mathrm{logit}(P(\delta = 1|X, Z)) = \beta_0 + \beta_1 X + \beta_2 Z + \beta XZ.$$

And they observed that the p-values for both β_1 and β_2 are less than 0.001 which strongly suggests the dependence of δ on *both* X *and* Z. By noting that

$$\begin{aligned} P(\delta = 1|X = t, Z = z) &= \frac{\lambda_{T|Z}(t|z)}{\lambda_{T|Z}(t|z) + \lambda_{C|Z}(t|z)} \\ &= \frac{1}{1 + \lambda_{C|Z}(t|z)/\lambda_{T|Z}(t|z)}, \end{aligned} \tag{2.1}$$

we conclude that it is not reasonable to assume that $\lambda_{C|Z}/\lambda_{T|Z}$ is free of X and is only related to the Z as the conditional proportional model assumes.

Rearranging (2.1), we may establish the following relationship between $\lambda_{T|Z}$ and $\lambda_{C|Z}$,

$$\lambda_{C|Z}(t|z) = \exp\{-\mathrm{logit}(P(\delta = 1|t, z))\}\, \lambda_{T|Z}(t|z). \tag{2.2}$$

This suggests a further extension of Koziol-Green model different from the conditional proportional model.

$$\lambda_{C|Z}(t|z) = \gamma(t, z; \theta)\lambda_{T|Z}(t|z), \tag{2.3}$$

where γ is a function known up to a finite-dimensional parameter, θ. Function γ models the possible information contained in the informative censoring.

In absence of covariates Z, the semiparametric censoring model (2.3) boils down to the semiparametric model proposed by Dikta (1998). A distribution estimator more efficient than Kaplan-Meier estimator was given in that paper. This result gives us a hint that function γ in (2.3) does contain some additional information provided by the informative censoring. The semiparametric censoring model without covariates is also studied by Dikta (2000) and Zhu et al. (2002).

A practical issue raised by Subramanian (2000) is how an assumption like (1.2) can be checked in real applications. Here, we have exactly the same problem for our semiparametric censoring model. Equation (2.2) gives us a possible solution. Through (2.2), we relate function γ with a logistic regression problem. More specifically,

$$\text{logit}(P(\delta = 1|t, z)) = -\ln(\gamma(t, z; \theta)). \tag{2.4}$$

Thus, to check condition (2.3) is equivalent to check (2.4). There are numerous results devoted to checking the goodness-of-fit of (2.4). The readers may refer to Hart (1997) or Aerts et al. (1999) for recent advances.

To use the semiparametric model (2.3), we must choose the parametric family for γ in the first place. In practice, we suggest to use a logistic regression as we shall do for a real example in Section 6. There are several reasons for us to do so. First, logistic regression has been proved to be one of the most efficient ways to model the binomial data. Second, due to the close relationship between the logit of $P(\delta = 1|t, z)$ and γ described in (2.4), it is much easier for us to interpret parameter θ by using logistic regression. Another advantage of using logistic model is the popular availability of softwares. As a commonly used statistical routine, logistic regression and the corresponding diagnostic procedures have been implemented in almost all the popular statistical packages. Thus, it is quite easy to do things like model checking and stepwise model selection for parsimonious parameterization γ.

By narrowing down the parametric family for γ, we are able to use the graphical or other statistical tools to test the proportionality assumption (1.2) against a more general alternative (2.3). Consider the binomial regression of δ on the covariates and the observed time. The proportionality condition assumes that the probability of $\delta = 0$ is a constant, which yields a nested model of a more general model assuming that γ lies in a large parametric family including constants. For example, if we consider the logistic regression. We have

$$\operatorname{logit}\left[P\left(\delta=0|X,Z\right)\right] = \left(1, X, Z'\right)\left(\theta_0, \theta_1, \theta_2'\right)'. \tag{2.5}$$

To test whether the proportionality condition (1.2) holds is equivalent to test whether $\theta_1 = 0$ and $\theta_2 = \mathbf{0}$. Usually, graphical and many other statistical tools are available for the generalized linear model (2.5) to conduct this test.

3 Information bound

Under general random censorship condition, it has been shown that the maximum partial likelihood estimator for the regression parameter β_0 and the Breslow estimator for the baseline cumulative hazard function

$$\Lambda_0(t) = \int_0^t \lambda_0(u)\,du$$

are efficient in the sense that they reach the information bound by assuming that the censoring is non-informative (Bickel et al., 1993). However, in our semiparametric censoring model, this assumption does not hold. The maximum partial likelihood estimator and the Breslow estimator are no longer optimal. There is still some room in which we can improve these estimators. For example, Subramanian (2000) derived information bounds different from that of the noninformative censoring model for estimating β_0 and Λ_0 of the Koziol-Green model. He also showed that both maximum partial likelihood estimator and Breslow estimator are not efficient for Koziol-Green model and thus can be improved. In this section, we establish the information bound associated with our semiparametric censoring model. Again, like for the Koziol-Green model, it is shown that the maximum partial likelihood estimator is not efficient for the semiparametric model.

We first write down the full likelihood of our semiparametric censoring model. Denote the density of Z by $k(\cdot)$. The full likelihood can be expressed as the following

$$\begin{aligned}
f(x,\delta,z) &= \left(f_{T|Z}(x|z)\, S_{C|Z}(x|z)\right)^{\delta} \left(f_{C|Z}(x|z)\, S_{T|Z}(x|z)\right)^{1-\delta} k(z) \\
&= \left[\left(f_{T|Z}(x|z)\right)^{\delta} \left(S_{T|Z}(x|z)\right)^{1-\delta}\right] \\
&\quad \cdot \left[\left(S_{C|Z}(x|z)\right)^{\delta} \left(f_{C|Z}(x|z)\right)^{1-\delta}\right] k(z) \\
&= \left[\left(\lambda_{T|Z}(x|z)\right)^{\delta} S_{T|Z}(x|z)\right] \left[S_{C|Z}(x|z) \left(\lambda_{C|Z}(x|z)\right)^{1-\delta}\right] k(z) \\
&\equiv f_1(x,\delta,z)\, f_2(x,\delta,z)\, k(z). \qquad (3.1)
\end{aligned}$$

The usual partial likelihood estimator can be obtained by maximizing the partial likelihood f_1. The Cox model assumes that

$$\begin{aligned}
\lambda_{T|Z}(x|z) &= \exp\left(z'\beta_0\right) \lambda_0(x), \\
S_{T|Z}(x|z) &= \exp\left(-\exp\left(z'\beta_0\right) \int_0^x \lambda_0(u)\, du\right).
\end{aligned}$$

Thus,

$$\begin{aligned}
L_1(\beta,\Lambda_0) &\equiv \sum_{i=1}^{n} \log f_1(X_i,\delta_i,Z_i) \qquad (3.2) \\
&= \sum_{i=1}^{n} \delta_i \left(\log \lambda_0(X_i) + Z_i'\beta_0\right) - \exp\left(Z_i'\beta_0\right) \int_0^{X_i} \lambda_0(u)\, du.
\end{aligned}$$

To estimate β_0 using L_1, the usual maximum partial likelihood estimator $\widehat{\beta}_p$ solves the following score equation

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \left(Z_i - \widetilde{Z}(\beta,t)\right) dN_i^u(t), \qquad (3.3)$$

where $N_i^u(t) = \delta_i N_i(t)$ with $N_i(t) = I(X_i \le t)$ and

$$\widetilde{Z}(\beta,t) = \frac{\sum_{i=1}^{n} Y_i(t) \exp\left(Z_i'\beta\right) Z_i}{\sum_{i=1}^{n} Y_i(t) \exp\left(Z_i'\beta\right)},$$

$Y_i(t) = I(X_i \geq t)$ is the indicator of risk set. In a pioneer article, Andersen and Gill (1982) showed that

$$\sqrt{n}\left(\widehat{\beta}_p - \beta_0\right) \rightarrow_d N\left(0, \Sigma_p^{-1}\right), \tag{3.4}$$

where $\Sigma_p^{-1} = -\lim \partial U(\beta_0)/\partial\beta$ is assumed to be positive definite. The Breslow estimator of the cumulative baseline hazards is given based on $\widehat{\beta}_p$:

$$\widehat{\Lambda}_b(x) = \sum_{i=1}^{n} \int_0^x \frac{dN_i^u(s)}{\sum_{i=1}^{n} Y_i(s)\exp\left(Z_i'\widehat{\beta}_p\right)}. \tag{3.5}$$

The original reason for only using partial likelihood L_1 to estimate β_0 and Λ_0 is because we do not have any idea how f_2 may look like and consequently, how it may contribute to estimating β_0 or Λ_0. However, for our semiparametric model, f_2 can be given explicitly. Simply ignoring f_2 may lead to suboptimal estimators. To assess the optimality of estimators, the information bounds were calculated in this section.

For expositional ease, let us start with a sample of size one. (2.3) implies that

$$\begin{aligned}
\lambda_{C|Z}(x|z) &= \exp\left(z'\beta_0\right)\gamma(x, z; \theta_0)\lambda_0(x), \\
S_{C|Z}(x|z) &= \exp\left(-\exp\left(z'\beta_0\right)\int_0^x \lambda_0(u)\gamma(u, z; \theta_0)\, du\right).
\end{aligned}$$

Thus, the full log likelihood, up to an additive term not depending on (β, θ, Λ), is

$$\begin{aligned}
L(\beta, \theta, \Lambda) &= Z'\beta + \log\lambda(X) + (1-\delta)\log\gamma(X, Z; \theta) \qquad (3.6) \\
&\quad - \exp\left(Z'\beta\right)\int_0^\infty I(X \geq u)\lambda(u)\left(1 + \gamma(u, Z; \theta)\right) du
\end{aligned}$$

Consider a semiparametric smooth submodel $\{\lambda_{(\eta)} : \eta \in R\}$ in which $\lambda_{(0)} = \lambda_0$ and

$$\left.\frac{\partial \log \lambda_{(\eta)}}{\partial \eta}\right|_{\eta=0} = a,$$

where $a \in L_2(P_X) \equiv \{a : E\left[a^2(X)\right] < \infty\}$. Let M^u be the usual counting process martingale associated with the Cox model:

$$M^u(t) = M^u(t|Z) = \delta I(X \leq t) - \int_0^t I(X \geq u)\exp(Z'\beta)d\Lambda_0(u).$$

Similarly, we define

$$\begin{aligned} M^c(t) &= M^c(t|Z) = (1-\delta)I(X \le t) \\ &\quad - \int_0^t I(X \ge u)\exp(Z'\beta)\gamma(u, Z;\theta)d\Lambda_0(u); \\ M(t) &= M(t|Z) = I(X \le t) \\ &\quad - \int_0^t I(X \ge u)\exp(Z'\beta)\left(1+\gamma(u, Z;\theta)\right)d\Lambda_0(u). \end{aligned}$$

The score operator for the hazard Λ and the score vectors of β and θ are the partial derivatives of the likelihood $L(\beta, \theta, \Lambda)$ with respect to η, β and θ evaluated at $\beta = \beta_0, \theta = \theta_0$ and $\eta = 0$:

$$\begin{aligned} L_\lambda a &= a(X) - \exp(Z'\beta_0)\int_0^\infty Y(t)a(t)\left(1+\gamma(t, Z;\theta)\right)d\Lambda_0(t) \\ &= \int_0^\infty a(t)dM(t), \qquad (3.7) \\ L_\beta &= Z - Z\exp(Z'\beta_0)\int_0^\infty Y(t)\left(1+\gamma(t, Z;\theta)\right)d\Lambda_0(t) \\ &= \int_0^\infty Z dM(t), \qquad (3.8) \\ L_\theta &= (1-\delta)\frac{\gamma_\theta(X, Z;\theta_0)}{\gamma(X, Z;\theta_0)} - \exp(Z'\beta_0)\int_0^\infty Y(t)\gamma_\theta(X, Z;\theta_0)d\Lambda_0(t) \\ &= \int_0^\infty \frac{\gamma_\theta(X, Z;\theta)}{\gamma(X, Z;\theta)}dM^c(t), \qquad (3.9) \end{aligned}$$

where $\gamma_\theta = \partial\gamma/\partial\theta$. To derive the efficient scores for β and θ, we first eliminate the hazard function by projecting L_β and L_θ onto the tangent space for the hazard. Let a_β^* and a_θ^* be the least favorable directions for β and θ respectively, i.e. $L_\beta - L_\lambda a_\beta^*$ and $L_\theta - L_\lambda a_\theta^*$ are orthogonal to the space $\{L_\lambda a : a \in L_2(P_X)\}$. That is, a_β^* and a_θ^* must satisfy

$$\begin{aligned} E\left[\left(L_\beta - L_\lambda a_\beta^*\right)L_\lambda a\right] &= 0, \qquad \forall a \in L_2(P_X), \\ E\left[\left(L_\theta - L_\lambda a_\theta^*\right)L_\lambda a\right] &= 0, \qquad \forall a \in L_2(P_X). \end{aligned}$$

By the martingale representations (3.7), (3.8) and (3.9), these two equations can be written as

$$\begin{aligned} E\left[\int_0^\infty \left(Z - a_\beta^*\right)dM \int_0^\infty a\,dM\right] &= 0, \forall a \in L_2(P_X), \\ E\left[\left(\int_0^\infty \frac{\gamma_\theta(X, Z;\theta)}{\gamma(X, Z;\theta)}dM^c - \int_0^\infty a_\theta^* dM\right)\int_0^\infty a\,dM\right] &= 0, \forall a \in L_2(P_X). \end{aligned}$$

To solve the above equations, we need the following result, which is a direct application of Lemma 1 from Sasieni (1992),

Lemma 3.1. *For any measurable functions f and g such that $E(f^2(X,Z))$, $E(g^2(X,Z)) < \infty$, we have*

$$
\begin{aligned}
E\left[\int_0^\infty f dM \int_0^\infty g\, dM\right] &= E[f(X,Z)g(X,Z)] && (3.10)\\
E\left[\int_0^\infty f dM^u \int_0^\infty g\, dM^u\right] &= E[\delta f(X,Z)g(X,Z)] && (3.11)\\
E\left[\int_0^\infty f dM^c \int_0^\infty g\, dM^c\right] &= E[(1-\delta)f(X,Z)g(X,Z)] && (3.12)\\
E\left[\int_0^\infty f dM^c \int_0^\infty g\, dM^u\right] &= 0 && (3.13)
\end{aligned}
$$

Using the above lemma, we can get the explicit formulae for a^*_β and a^*_θ

$$
\begin{aligned}
a^*_\beta(X) &= E(Z|X) && (3.14)\\
a^*_\theta(X) &= E\left(\frac{\gamma_\theta(X,Z;\theta)}{1+\gamma(X,Z;\theta)}\Big|X\right) && (3.15)
\end{aligned}
$$

Summing up, we have the following information bounds for estimating β and θ.

Theorem 3.1. *Under our model assumptions (1.1) and (2.3), provided that γ is twice differentiable with respect to θ and all the integrals and expectations exist, the efficient score vectors for β and θ are*

$$
\begin{aligned}
L^*_\beta &= \int_0^\infty (Z - E[Z|X=t])\, dM(t) && (3.16)\\
L^*_\theta &= \int_0^\infty \frac{\gamma_\theta(t,Z;\theta)}{\gamma(t,Z;\theta)} dM^c(t)\\
&\quad - \int_0^\infty E\left[\frac{\gamma_\theta(X,Z;\theta)}{1+\gamma(X,Z;\theta)}\Big|X=t\right] dM(t). && (3.17)
\end{aligned}
$$

Consequently, the information bound for estimating (β,θ) is

$$
I = \begin{pmatrix} I_{\beta\beta} & I_{\beta\theta} \\ I'_{\beta\theta} & I_{\theta\theta} \end{pmatrix}
$$

where

$$I_{\beta\beta} = E[\mathrm{Var}(Z|X)] \equiv E\left[Z^{\otimes 2} - (E(Z|X))^{\otimes 2}\right] \tag{3.18}$$

$$I_{\beta\theta} = E\left[(Z - E(Z|X))\left(\frac{\gamma_\theta(X,Z;\theta)}{1+\gamma(X,Z;\theta)}\right)'\right] \tag{3.19}$$

$$I_{\theta\theta} = E\left[\frac{\gamma_\theta^{\otimes 2}(X,Z;\theta)}{\gamma(X,Z;\theta)(1+\gamma(X,Z;\theta))} - \left(E\left(\frac{\gamma_\theta(X,Z;\theta)}{1+\gamma(X,Z;\theta)}|X\right)\right)^{\otimes 2}\right] \tag{3.20}$$

given that I is positive definite. Meanwhile, the inverse information covariance function for estimation of Λ_0 is

$$\begin{aligned} I_\Lambda^{-1}(s,t) = \int_0^{s\wedge t} \frac{\lambda_0(u)\,du}{E\left[I(X \geq u)\exp(Z'\beta_0)(1+\gamma(u,Z;\theta_0))\right]} \\ + \left(\int_0^s E(Z'|X)\lambda_0(u)du, \int_0^s E\left(\frac{\gamma_\theta'(X,Z;\theta_0)}{1+\gamma(X,Z;\theta_0)}|X\right)\lambda_0(u)du\right) I^{-1} \\ \left(\int_0^t E(Z|X)\lambda_0(u)du, \int_0^t E\left(\frac{\gamma_\theta(X,Z;\theta_0)}{1+\gamma(X,Z;\theta_0)}|X\right)\lambda_0(u)du\right) \end{aligned} \tag{3.21}$$

A consequence of Theorem 3.1 is that the usual maximum partial likelihood estimator $\widehat{\beta}_p$ and Breslow estimator $\widehat{\Lambda}_b$ are not efficient.

Corollary 3.1. *Under the assumption of Theorem 3.1, $\widehat{\beta}_p$ and $\widehat{\Lambda}_b$ are not efficient.*

The next section is thus devoted to finding an alternative estimator which is of full efficiency.

4 Estimation procedure

A general approach to derive efficient estimators for finite dimensional parameters in a semiparametric model is the profiled likelihood estimator. Given θ and β, the maximum likelihood estimate of Λ_0 is given by

$$\widehat{\Lambda}(x;\beta,\theta) = \sum_{i=1}^n \int_0^x \frac{dN_i(s)}{\sum_{j=1}^n Y_j(s)\exp(Z_j'\beta)(1+\gamma(s,Z_j;\theta))}. \tag{4.1}$$

Plugging this back to the full likelihood for observations (X_i, Z_i, δ_i) $,i = 1, ..., n$, the following log profiled likelihood is obtained:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{Z_i'\beta - \ln\left(\sum_{j=1}^{n} Y_j(X_i)\exp\left(Z_j'\beta\right)(1+\gamma(X_i, Z_j;\theta))\right)\right.$$
$$\left. + (1-\delta_i)\ln\gamma(X_i, Z_i;\theta)\right\}. \quad (4.2)$$

Denote the above log profiled likelihood by $PL(\beta, \theta)$. We define our estimate of β and θ as the maximizer of $PL(\beta, \theta)$:

$$\left(\widehat{\beta}_{new}, \widehat{\theta}\right) = \arg\max_{\beta,\theta} PL(\beta, \theta).$$

To maximize the profiled likelihood, we can iteratively solve the following score equations:

$$S_{n\beta}(\beta) \equiv \frac{1}{n}\frac{\partial PL(\beta,\theta)}{\partial\beta} = \frac{1}{n}\sum_{i=1}^{n}\int_0^\infty \left(Z_i - \bar{Z}(\beta, t)\right) dN_i(t) \quad (4.3)$$

$$S_{n\theta}(\theta) \equiv \frac{1}{n}\frac{\partial PL(\beta,\theta)}{\partial\theta} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(1-\delta_i)\gamma_\theta(X_i, Z_i;\theta)}{\gamma(X_i, Z_i;\theta)}\right)$$
$$-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{n} Y_j(X_i)\exp\left(Z_j'\beta\right)\gamma_\theta(X_i, Z_j;\theta)}{\sum_{j=1}^{n} Y_j(X_i)\exp\left(Z_j'\beta\right)[1+\gamma(X_i, Z_j;\theta)]}\right) \quad (4.4)$$

where

$$\bar{Z}(\beta, t) = \frac{\sum_{j=1}^{n} Y_j(t)\exp\left(Z_j'\beta\right)[1+\gamma(t, Z_j;\theta)]Z_j}{\sum_{j=1}^{n} Y_j(t)\exp\left(Z_j'\beta\right)[1+\gamma(t, Z_j;\theta)]}$$

The following theorem justifies the proposed estimator.

Theorem 4.1. *For the semiparametric model satisfying* (1.1) *and* (2.3), *if γ is twice differentiable with respect to θ and all the expectations and integrals in Theorem 3.1 are well defined and $I > 0$, then the following statements hold.*

(a) *(Existence) With probability going to 1, there exists a sequence of roots* $(\widehat{\beta}_{new}, \widehat{\theta})$ *to the score equations* (4.3) *and* (4.4) *such that*

$$\widehat{\beta}_{new} \to_p \beta_0 \quad \text{and} \quad \widehat{\theta} \to_p \theta_0 \qquad\qquad (n \to \infty).$$

(b) *(Uniqueness) If* $\{\widetilde{\beta}, \widetilde{\theta}\}$ *is also a sequence of consistent roots to* (4.3) *and* (4.4), *then* $\widetilde{\beta} = \widehat{\beta}_{new}$ *and* $\widetilde{\theta} = \widehat{\theta}$ *with probability going to one as* $n \to \infty$.

(c) *(Asymptotic Normality)* $\left(\widehat{\beta}_{new}, \widehat{\theta}\right)$ *is asymptotically normal, more specifically,*

$$\sqrt{n}\left(\begin{pmatrix} \widehat{\beta}_{new} \\ \widehat{\theta} \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \theta_0 \end{pmatrix}\right) \to_d N\left(0, I^{-1}\right) \qquad\qquad (n \to \infty). \quad (4.5)$$

In general, there might be multiple roots of the score equations (4.3) and (4.4), one must face a difficulty of choosing one root among them. To overcome this problem, we could choose a root from a vicinity of another consistent estimator of β, i.e., the maximum partial likelihood estimator $\widehat{\beta}_p$. Fortunately, for some specific parametric form of γ, this problem does not exist at all.

Corollary 4.1. *Under the assumptions of Theorem 4.1, if*

$$\gamma(x, z; \theta) = \exp\left(\theta_0 + x\theta_1 + z'\theta_2\right), \qquad (4.6)$$

then with probability tending to one, the log profiled likelihood $PL(\beta, \theta)$ *is convex and thus the score equations* (4.3) *and* (4.4) *have a unique solution.*

After getting an estimate of (β_0, θ_0), we can use it to construct an estimate of Λ_0. Replacing β and θ by their estimators in (4.1), the following estimator is obtained:

$$\widehat{\Lambda}_0(x) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} Y_j(s) \exp\left(Z_j' \widehat{\beta}_{new}\right)\left(1 + \gamma\left(s, Z_j; \widehat{\theta}\right)\right)}. \qquad (4.7)$$

In an attempt to improve the efficiency of Kaplan-Meier estimator for the semiparametric censoring model without covariates, Dikta (1998) suggested a modification of the Kaplan-Meier estimator which is similar to what we did here.

Theorem 4.2. *Under the conditions of Theorem 4.1,* $\sqrt{n}\left[\widehat{\Lambda}_0(x)-\Lambda_0(x)\right]$ *tends weakly to a Gaussian process B with the covariance structure given by*

$$\mathrm{Cov}\,(B(s),B(t))=I_\Lambda^{-1}(s,t).$$

Remark 4.1. *As pointed out by a referee, a benefit of using the semiparametric censorship model is that it allows a semiparametric estimator of the conditional cumulative hazard even if the Cox model* (1.1) *is not valid. i.e.*

$$\widehat{\Lambda}_{T|Z}(t|z)=\int_0^t \widehat{P}\,(\delta=1|X=x,Z=z)\,\widehat{\Lambda}_{X|Z}(dx|z), \qquad (4.8)$$

where $\widehat{P}$ is a parametric estimator to the binary regression (2.4) *and $\widehat{\Lambda}_{X|Z}$ denotes the empirical conditional cumulative hazard of the data (X_i,Z_i), $i=1,\ldots,n$. This estimator is expected to outperform its purely nonparametric counterparts (see Dabrowska, 1989).*

An advantage of the Koziol-Green model is that the censoring indicator δ and the observation X are independent of each other given the covariates Z. Thus, the information bound for estimating both θ and β can be further simplified. Actually, the scores for β,θ and λ_0 are independent for the Koziol-Green model (see Subramanian, 2000). Subramanian (2000) gave efficient estimators for both β and the baseline hazard based on this fact. It is not hard to show that his estimator is a special case of ours.

5 Simulations

To compare our estimator $\widehat{\beta}_{new}$ with the usual partial likelihood estimator $\widehat{\beta}_p$, we conducted the following simulation.

Example 5.1. *Consider the following model*

$$\lambda_0(t)=1,\quad \beta_0=0.8,\quad \gamma(t,Z;\theta)=\exp(\theta Z)-1. \qquad (5.1)$$

We sampled Z from a uniform distribution $U[0,1]$. To ensure the stability of the numerical results, replications are done for each combination of three different values of θ $(0.8,1.6,2.4)$ *and four different sample sizes* $(50,100,200,500)$.

Our interest here lies in how our estimator proposed in Section 4 compares with the partial likelihood estimator. We can easily control the level of censorship which plays an important role in the comparisons. The expected portion of censored observations is

$$P = E\left(\frac{\gamma}{1+\gamma}\right) = \int_0^1 [1 - \exp(-\theta z)]\, dz = 1 - \frac{1}{\theta}[1 - \exp(-\theta)]. \quad (5.2)$$

We investigated three different levels of censorship. When $\theta = 0.8$, the expected portion of censored observation is 30%, when $\theta = 1.6$, it is about 50% and $\theta = 2.4$ corresponds to 60%. To get conclusive numeric results, for each combination of sample size and censoring level, we repeated the simulation for $10,000$ times. The mean squared error of both estimators are reported in Table 1. For each sample size, the first row represents the new estimator proposed in Section 4 and the second row corresponds to the maximum partial likelihood estimator. In the simulation, we found that our estimator has a significantly better performance than the partial likelihood estimator in terms of the mean squared error. In fact, our estimator becomes more and more preferable to the maximum partial likelihood estimator in when the portion of censored observations increases. The reason is because we took care of the censored observations instead of discarding the censoring distribution in Formula (3.1) as the maximum partial likelihood estimator does.

Table 1: Comparison between our estimator and the partial likelihood estimator

Sample Size	$\theta = 0.8$	$\theta = 1.6$	$\theta = 2.4$
50	0.37206927	0.49515073	0.6710972
	0.48046879	0.82965151	1.4065614
100	0.17543832	0.23524427	0.3263588
	0.22443523	0.37684911	0.6451731
200	0.08547634	0.11696191	0.1631008
	0.11075901	0.18617300	0.3143165
500	0.03255874	0.04626697	0.0636069
	0.04227329	0.07336781	0.1217943

In the next example, we would like to check how sensitive our new estimate of β_0 is subject to the proportionality assumption (2.3). We adopted a similar setting as Example 5.1.

Example 5.2. *For each given sample size and θ, the survival time T was generated in the same way as that of Example 5.1. The censoring time C was generated according to*

$$\lambda_{C|T}(t|z) = \gamma(t, z; \theta + w)\lambda_{T|Z}(t|z), \tag{5.3}$$

where $\lambda_{T|Z}$ and γ were defined as (5.1) and random perturbation w was independently sampled from $U[-0.2, 0.2]$.

Table 2: Performance of both estimators for Example 5.2

Sample Size	$\theta = 0.8$	$\theta = 1.6$	$\theta = 2.4$
50	0.37148253	0.52326512	0.6744166
	0.48679919	0.87931921	1.4166742
100	0.17799426	0.23826997	0.3362547
	0.22789981	0.39437879	0.6594181
200	0.08430638	0.11623248	0.1612840
	0.10779201	0.18371902	0.3109602
500	0.03248861	0.04644066	0.0650155
	0.04165614	0.07371366	0.1229361

Table 2 summarizes the mean squared error of both estimators based on $10,000$ simulations. Again, the first row for each sample size corresponds to the new estimator and the second row for the maximum partial likelihood estimator. The new estimator still outperformed the maximum partial likelihood estimator in this simulation. Intuitively, although the proportionality condition (2.3) does not hold for this example, the censoring times were still distributed around it. This information was successfully captured by the proposed estimator. In fact, the best projection of the true model into the semiparametric censoring model shares the same regression coefficient β with the true model. According to the maximum likelihood theory for misspecified models (White, 1994), we can expect that our new estimator will still behave well. This example suggests that in certain scenario, our new estimator is preferable to the maximum partial likelihood estimator even if proportionality condition (2.3) does not hold.

6 Application

To illustrate the method developed before, we looked at a real example. The Stanford Heart Transplant data that were available in February 1980 has been studied by many researchers (Escobar and Meeker, 1992). The dataset contained 184 transplant cases with the following variables: Time, measured from the date of transplant in days; Status code (dead or alive); AGE, patient age at first transplant, in years; T5, mismatch score. Among 184 cases, 71 cases were censored. The previous studies suggested that T5 is not a useful explanatory variable for the dataset. Thus, we considered the following model:

$$\begin{aligned}\lambda_{T|Z}(t|z) &= \lambda_0(t)\exp(\beta_1(\text{AGE}-41.7)+\beta_2(\text{AGE}-41.7)^2);\\ \gamma(t,z;\theta) &= \exp(\theta_0+\theta_1\text{Time}+\theta_2(\text{AGE}-41.7)+\theta_3(\text{AGE}-41.7)^2).\end{aligned}$$

As indicated by Lin and Spiekerman (1996), although the Cox model is not quite satisfactory for survival time less that 100 days, overall, it still provides valid approximation to the dataset. To illustrate our method, we computed both the maximum partial likelihood estimator and the new efficient estimator for this dataset. In order to evaluate the variation of the estimators, we repeated both estimating procedures for 1000 boostrapped samples of the dataset. The following table summarizes the estimates and their associated boostrapped standard deviation. The first row corresponds to the efficient estimator proposed in Section 4 and the second row represents the maximum partial likelihood estimate.

Table 3: Estimates of the Stanford Heart Transplant data

Methods	β_1		β_2	
	Estimate	Std. Dev	Estimate	Std. Dev
New	0.043405908	0.0082490841	0.001845843	0.0006095718
Old	0.043397230	0.0105015275	0.001972247	0.0007603243

From Table 3, we see no essential difference between two estimates. But the efficient estimates have relatively smaller variation than the maximum partial likelihood estimate. The cumulative hazards estimated by (4.7) together with the Breslow estimator is also given in Figure 1. The solid line gives the Breslow estimator and the dashed line represents the new estimator.

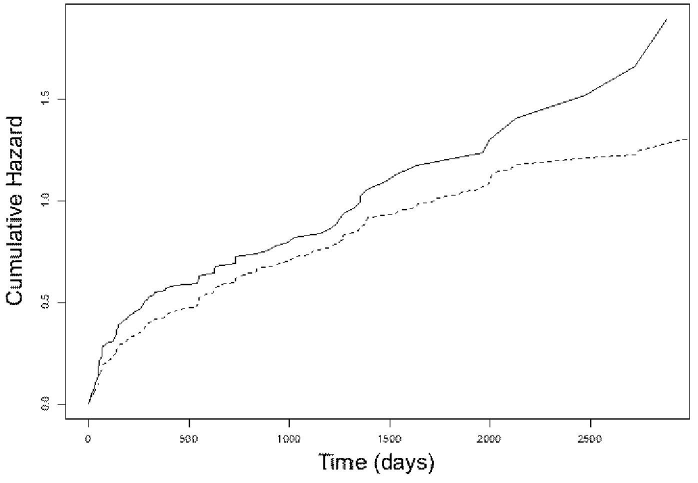

Figure 1: Estimated cumulative hazards

From the plot, we find that the new estimate is smaller than the Breslow estimator. To understand why this happens, let us look at the estimates of θ, which are reported in Table 4.

Table 4: Estimates of the Stanford Heart Transplant data

	θ_0	θ_1	θ_2	θ_3
Estimate	-1.248394665	0.001172283	-0.073850138	-0.001442896
Std. Dev.	0.252869910	0.000280947	0.022318961	0.001511038

From the estimates, we see that an observation is more likely to be censored when Time is larger. This agrees with our intuition. If a patient felt well during the study, he/she is more likely to lose interest in the followup later on. In other words, the uncensored subjects in the study may be more likely to die. As a consequence, Breslow estimate might overestimate the risk.

Appendix : Technical proofs

Proof of Lemma 3.1. (3.11) is from Sasieni (1992). Similarly, (3.12) holds because the symmetry between δ and $1-\delta$. (3.10) can be regarded as a special case of (3.11) when there is no censoring. Since $M = M^u + M^c$,

$$\begin{aligned} E\left[\int_0^\infty f\,dM \int_0^\infty g\,dM\right] &= E\left[\int_0^\infty f\,dM^u \int_0^\infty g\,dM^u\right] \\ &\quad + E\left[\int_0^\infty f\,dM^c \int_0^\infty g\,dM^c\right] + E\left[\int_0^\infty f\,dM^u \int_0^\infty g\,dM^c\right] \\ &\quad + E\left[\int_0^\infty f\,dM^c \int_0^\infty g\,dM^u\right]. \end{aligned}$$

By (3.10), (3.11) and (3.12) we know that

$$E\left[\int_0^\infty f\,dM^u \int_0^\infty g\,dM^c\right] + E\left[\int_0^\infty f\,dM^c \int_0^\infty g\,dM^u\right] = 0.$$

Now the proof of (3.13) is completed by the fact that

$$\begin{aligned} E\left[\int_0^\infty f\,dM^c \int_0^\infty g\,dM^u\right] &= E\left[\int_0^\infty f g\,d < M^u, M^c >\right] \\ &= E\left[\int_0^\infty f\,dM^u \int_0^\infty g\,dM^c\right]. \end{aligned}$$

□

Proof of Theorem 3.1. To prove that L^*_β and L^*_θ are efficient scores, it suffices to verify that they are orthogonal to L_λ. By Lemma 3.1, $\forall a \in L_2(P_X)$,

$$\begin{aligned} E\left[\int_0^\infty (Z - a^*_\beta)\,dM \int_0^\infty a\,dM\right] &= E\left[(Z - E(Z|X))\,a(X)\right] \\ &= E\left(E\left[(Z - E(Z|X))\,a(X)|X\right]\right) \\ &= 0 \end{aligned}$$

Similarly, since $M = M^u + M^c$,

$$
\begin{aligned}
&E\left[\left(\int_0^\infty \frac{\gamma_\theta(X,Z;\theta)}{\gamma(X,Z;\theta)}\, dM^c - \int_0^\infty a_\theta^*\, dM\right)\int_0^\infty a\, dM\right] \\
&= E\left[\int_0^\infty \frac{\gamma_\theta(X,Z;\theta)}{\gamma(X,Z;\theta)}\, dM^c \int_0^\infty a\, dM\right] - E\left[\int_0^\infty a_\theta^*\, dM \int_0^\infty a\, dM\right] \\
&= E\left[\int_0^\infty \frac{\gamma_\theta(X,Z;\theta)}{\gamma(X,Z;\theta)}\, dM^c \int_0^\infty a\, dM^c\right] - E\left[\int_0^\infty a_\theta^*\, dM \int_0^\infty a\, dM\right] \\
&= E\left[(1-\delta)\,\frac{\gamma_\theta(X,Z;\theta)}{\gamma(X,Z;\theta)} a(X)\right] - E\left[a_\theta^*(X)a(X)\right] \\
&= E\left(E\left[(1-\delta)\,\frac{\gamma_\theta(X,Z;\theta)}{\gamma(X,Z;\theta)} a(X)|X,Z\right]\right) - E\left[a_\theta^*(X)a(X)\right] \\
&= E\left[\frac{\gamma_\theta(X,Z;\theta)}{1+\gamma(X,Z;\theta)} a(X)\right] - E\left[a_\theta^*(X)a(X)\right] \\
&= E\left(E\left[\left(\frac{\gamma_\theta(X,Z;\theta)}{1+\gamma(X,Z;\theta)} - a_\theta^*(X)\right) a(X)|X\right]\right) \\
&= 0.
\end{aligned}
$$

Similar calculation leads to $E(L_\phi^* L_\psi^{*\prime}) = I_{\phi\psi}$ for $\phi, \psi = \beta$ or θ. The calculation of I_Λ^{-1} proceeds as in Bickel et al. (1993, pp. 293–294). □

***Proof of Corollary** 3.1.* Note that

$$
\begin{aligned}
I_{\beta\beta} &= E\left[(1-\delta)(Z^{\otimes 2} - (E(Z|X))^{\otimes 2})\right] + E\left[\delta(Z^{\otimes 2} - (E(Z|X))^{\otimes 2})\right] \\
&\geq E\left[(1-\delta)(Z^{\otimes 2} - (E(Z|X))^{\otimes 2})\right] \\
&\quad + E\left[\delta(Z^{\otimes 2} - (E(Z|X,\delta=1))^{\otimes 2})\right] \\
&\geq E\left[\frac{\gamma(X,Z;\theta)}{1+\gamma(X,Z;\theta)}(Z^{\otimes 2} - (E(Z|X))^{\otimes 2})\right] + \Sigma_p \\
&\equiv I_{\beta\beta}^c + \Sigma_p, \\
I_{\theta\theta} &\geq E\left[\frac{1+\gamma(X,Z;\theta)}{\gamma(X,Z;\theta)}\left(\left(\frac{\gamma_\theta(X,Z;\theta)}{1+\gamma(X,Z;\theta)}\right)^{\otimes 2}\right.\right. \\
&\quad \left.\left. - \left(E\left(\frac{\gamma_\theta(X,Z;\theta)}{1+\gamma(X,Z;\theta)}|X\right)\right)^{\otimes 2}\right)\right] \\
&\equiv I_{\theta\theta}^c.
\end{aligned}
$$

Now because

$$\begin{pmatrix} I^c_{\beta\beta} & I_{\beta\theta} \\ I'_{\beta\theta} & I^c_{\theta\theta} \end{pmatrix} \geq 0,$$

as a consequence of Theorem 3.1, $\widehat{\beta}_p$ is not efficient.

To show that Breslow estimator is not efficient, let us introduce another estimator of θ_0:

$$\widehat{\theta}_n = \arg\max_{\theta} \left(\prod_{k=1}^{n} \left[\gamma\left(X_k, Z_k; \theta\right)\right]^{1-\delta} \left[1 + \gamma\left(X_k, Z_k; \theta\right)\right]^{-1} \right).$$

A new estimator of Λ_0 could be defined as

$$\tilde{\Lambda}\left(t\right) = \sum_{i=1}^{n} \int_0^t \frac{dN_i\left(s\right)}{\left[\sum_{j=1}^{n} Y_j\left(s\right) \exp\left(\widehat{\beta_p}' Z_j\right)\right] \left[1 + \gamma\left(s, Z_i; \widehat{\theta}_n\right)\right]}.$$

Using the argument similar to that of Andersen and Gill (1982), we can show that $\sqrt{n}\left[\tilde{\Lambda}(t) - \Lambda_0(t)\right]$ converges to a Brownian motion with covariance structure

$$\begin{aligned} \tilde{I}_{\Lambda}^{-1}(s,t) = & \int_0^{s\wedge t} \frac{\lambda_0\left(u\right) du}{E\left[I\left(X \geq u\right) \exp\left(Z'\beta_0\right) \left(1 + \gamma\left(u, Z; \theta_0\right)\right)\right]} \\ & + \left(\int_0^s E(Z'|X)\lambda_0(u)du, \int_0^s E\left(\frac{\gamma'_\theta(X, Z; \theta_0)}{1 + \gamma(X, Z; \theta_0)} |X\right) \lambda_0(u)du \right) J^{-1} \\ & \left(\int_0^t E(Z|X)\lambda_0(u)du, \int_0^t E\left(\frac{\gamma_\theta(X, Z; \theta_0)}{1 + \gamma(X, Z; \theta_0)} |X\right) \lambda_0(u)du \right), \end{aligned}$$

where J is the covariance matrix of $n^{1/2}(\widehat{\beta}_p - \beta_0, \widehat{\theta}_n - \theta_0)$. Obviously, $J^{-1} \geq I^{-1}$ since $(\widehat{\beta}_p, \widehat{\theta}_n)$ is not an efficient estimator. Thus,

$$\tilde{I}_{\Lambda}^{-1}(t,t) \leq I_{\Lambda}^{-1}(t,t).$$

Denote $I_{\Lambda b}^{-1}(s,t)$ the covariance function of the Brownian motion which $n^{1/2}\left[\widehat{\Lambda}_b(t) - \Lambda_0(t)\right]$ converges to (*i.e.*, Andersen and Gill, 1982). Using the argument of Dikta (1998),

$$\tilde{I}_{\Lambda}^{-1}(t,t) \leq I_{\Lambda b}^{-1}(t,t).$$

Therefore,

$$I_{\Lambda}^{-1}(t,t) \leq I_{\Lambda b}^{-1}(t,t).$$

This implies that Breslow estimator is not efficient. □

Proof of Theorem 4.1. Different from the partial likelihood, $PL(\beta, \theta)$ is not necessarily a convex function its argument. This makes it difficult to apply the argument of Andersen and Gill (1982) here. However, we can rely on a more general technique developed for estimating equations in Foutz (1977). For convenience, denote $\nu = (\beta, \theta)$, $\nu_0 = (\beta_0, \theta_0)$ and $S_n(\nu) = (S'_{n\beta}, S'_{n\theta})'$. It suffices to verify the following conditions:

(i) $\partial S_n(\nu)/\partial\nu$ exists and is continuous for ν in a vicinity of ν_0.

(ii) There exists a neighborhood $N(\nu_0)$ of ν_0 and function h such that

$$\frac{\partial S_n(\nu)}{\partial \nu} \to_p h(\nu) \qquad uniformly.$$

(iii) $\partial S_n(\nu)/\partial\nu$ is negative definite at ν_0 with probability converging to 1 as $n \to \infty$.

(iv) $\partial S_n(\nu)/\partial\nu$ approaches 0 in probability at ν_0.

Conditions (i), (iii) and (iv) follow from straightforward calculations using the Law of Large Numbers. Now we are in the position to show (ii). For any $\epsilon > 0$, we could find a finite positive integer k and points $\nu_1, \ldots, \nu_k \in N(\nu_0)$ such that $N(\nu_0)$ can be covered by the union of open balls $B_\epsilon(\nu_1), \ldots, B_\epsilon(\nu_k)$, where $k = O(1/\epsilon^d)$ and d is the dimensionality of ν. Then,

$$\begin{aligned}
\sup_{\nu\in N(\nu_0)} \left| \frac{\partial S_n(\nu)}{\partial\nu} - E\left[\frac{\partial S_n(\nu)}{\partial\nu}\right]\right| \leq& \max_i \sup_{\nu\in B_\epsilon(\nu_i)} \left|\frac{\partial S_n(\nu)}{\partial\nu} - \frac{\partial S_n(\nu_i)}{\partial\nu}\right| \\
&+ \max_i \sup_{\nu\in B_\epsilon(\nu_i)} \left|E\left[\frac{\partial S_n(\nu)}{\partial\nu} - \frac{\partial S_n(\nu_i)}{\partial\nu}\right]\right| \\
&+ \max_i \left|\frac{\partial S_n(\nu_i)}{\partial\nu} - E\left[\frac{\partial S_n(\nu_i)}{\partial\nu}\right]\right| \\
\equiv& A_1 + A_2 + A_3.
\end{aligned}$$

Choose $N(\nu_0)$ such that

$$\sup_{\nu\in N(\nu_0)} \left|\partial^2 S_n(\nu)/\partial\nu^2\right| \leq C_1$$

for some positive constant C_1. Then the Mean Value Theorem yields that $A_1 + A_2 \leq 2C_1\epsilon$. On the other hand, applying Markov's inequality, we have

$$\begin{aligned} P(A_3 \geq \epsilon) &\leq \sum_i P\left(\left|\frac{\partial S_n(\nu_i)}{\partial \nu} - E\left[\frac{\partial S_n(\nu_i)}{\partial \nu}\right]\right| \geq \epsilon\right) \\ &\leq C_2 k/n\epsilon^2 \end{aligned}$$

for some constant $C_2 > 0$. Let ϵ tend to 0 in such a manner that $n\epsilon^{d+2} \to 0$. Then $k/n\epsilon^2 = O(1/n\epsilon^{d+2}) = o(1)$. Thus, as sample size n approaches ∞,

$$P(A_1 + A_2 + A_3 \geq (2C_1 + 1)\epsilon) \to 0.$$

This completes the proof of Part (a) and (b). To prove Part (c), we can use exactly the same argument as that for estimating equations and thus omitted here. $\square$

Proof of Theorem 4.2. Using the argument of Andersen and Gill (1982, p. 1104), we can decompose $n^{1/2}\left[\widehat{\Lambda}(t) - \Lambda_0(t)\right]$ into

$$n^{1/2}\left(\sum_{i=1}^{n}\int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} Y_j(s)\exp(Z_j\beta_0)(1+\gamma(s,Z_j;\theta_0))} - \Lambda_0^*(t)\right)$$

$$+ n^{1/2}\sum_{i=1}^{n}\int_0^t \frac{\sum_{j=1}^{n} Z_jY_j(s)\exp\left(Z_j\widehat{\beta}\right)(1+\gamma(s,Z_j;\theta_0))\,dN_i(s)}{\left[\sum_{j=1}^{n} Y_j(s)\exp(Z_j\beta_0)(1+\gamma(s,Z_j;\theta_0))\right]^2}(\beta_0 - \widehat{\beta}_{new})$$

$$+ n^{1/2}\sum_{i=1}^{n}\int_0^t \frac{\sum_{j=1}^{n} Y_j(s)\exp(Z_j\beta_0)\gamma_\theta(s,Z_j;\theta_0)\,dN_i(s)}{\left[\sum_{j=1}^{n} Y_j(s)\exp(Z_j\beta_0)(1+\gamma(s,Z_j;\theta_0))\right]^2}(\theta_0 - \widehat{\theta})$$

$+ o_p(1)$,

where

$$\Lambda_0^* = \int_0^t \lambda_0(u) I\left(\sum_{i=1}^n Y_i(u) > 0\right) du.$$

The covariance matrix of the first term could be given as

$$E\left[\int_0^{s\wedge t} \frac{\lambda_0(u)\,du}{E\left[I(X \geq u)\exp(\beta_0' Z)(1+\gamma(u, Z; \theta_0))\right]}\right].$$

Now the proof is completed since the first term is independent of $S_{n\beta}, S_{n\theta}$.

$\square$

Acknowledgement

The author is grateful to two anonymous referees whose comments helped him greatly improve both the content and the presentation of this paper.

References

Aerts, M., Claeskens, G., and Hart, J. D. (1999). Testing the fit of a parametric function. *Journal of the American Statistical Association*, 94:869–879.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10:1100–1120.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. The John Hopkins University Press, Baltimore and London.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society. Series B*, 34:187–220.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *The Annals of Statistics*, 17(3):1157–1167.

de Uña Álvarez, J. and González-Manteiga, W. (1998). Distributional convergence under proportional censorship when covariates are present. *Statistics & Probability Letters*, 39:305–315.

de Uña Álvarez, J. and González-Manteiga, W. (1999). Strong consistency under proportional censorship when covariates are present. *Statistics & Probability Letters*, 42:283–292.

Dikta, G. (1998). On semiparametric random censorship models. *Journal of Statistical Planning and Inference*, 66:253–279.

Dikta, G. (2000). The strong law under semiparametric random censorship models. *Journal of Statistical Planning and Inference*, 83(1):1–10.

Escobar, L. A. and Meeker, W. Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics*, 48:507–528.

Foutz, R. V. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association*, 72:147–148.

Gather, U. and Pawlitschko, J. (1998). Estimating the survival function under a generalized Koziol-Green model with partially informative censoring. *Metrika*, 48:189–207.

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-fit Tests*. Springer-Verlag, New York.

Jensen, U. and Wiedmann, J. (2001). Estimation of a survival curve with randomly censored data in the presence of a covariate. *Metrika*, 53(3):223–244.

Koziol, J. A. and Green, S. B. (1976). A Cramer-von Mises statistic for randomly censored data. *Biometrika*, 63:465–474.

Lin, D. Y. and Spiekerman, C. F. (1996). Model checking techniques for parametric regression with censored data. *Scandinavian Journal of Statistics. Theory and Applications*, 23(2):157–177.

Sasieni, P. D. (1992). Non–orthogonal projections and their application to calculating the information in a partly linear Cox model. *Scandinavian Journal of Statistics. Theory and Applications*, 19(3):215–233.

Subramanian, S. (2000). Efficient estimation of regression coefficients and baseline hazard under proportionality of conditional hazard. *Journal of Statistical Planning and Inference*, 84(1–2):81–94.

Veraverbeke, N. and Cadarso-Suárez, C. (2000). Estimation of the conditional distribution in a conditional Koziol-Green model. *Test*, 9(1):97–122.

White, H. (1994). *Estimation, Inference and Specification Analysis*. No. 22 in Econometric Society Monographs. Cambridge University Press, Cambridge.

Zhu, L. X., Yuen, K. C., and Tang, N. Y. (2002). Resampling methods for testing a semiparametric random censorship model. *Scandinavian Journal of Statistics. Theory and Applications*, 29(1):111–123.