Biometrika (2002), **89,** 3, *pp.* 659–668

Printed in Great Britain

Semiparametric analysis of transformation models with censored data

By KANI CHEN
Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
makchen@ust.hk

ZHEZHEN JIN
Department of Biostatistics, Mailman School of Public Health, Columbia University, 600 West 168th Street, New York, New York 10032, U.S.A.
zjin@biostat.columbia.edu

AND ZHILIANG YING
Department of Statistics, Columbia University, 2990 Broadway, New York, New York 10027, U.S.A.
zying@stat.columbia.edu

Summary

A unified estimation procedure is proposed for the analysis of censored data using linear transformation models, which include the proportional hazards model and the proportional odds model as special cases. This procedure is easily implemented numerically and its validity does not rely on the assumption of independence between the covariates and the censoring variable. The estimator is the same as the Cox partial likelihood estimator in the case of the proportional hazards model. Moreover, the asymptotic variance of the proposed estimator has a closed form and its variance estimator is easily obtained by plug-in rules. The method is illustrated by simulation and is applied to the Veterans' Administration lung cancer data.

Some key words: Estimating equation; Linear transformation model; Proportional hazards model; Proportional odds model.

1. Introduction

Since it was first introduced in Cox (1972), the proportional hazards model has been fully explored in theory and extensively used in practice (Cox, 1975; Tsiatis, 1981; Andersen & Gill, 1982). Another commonly-used model in survival analysis is the proportional odds model which has recently attracted considerable attention (Pettitt 1982, 1984; Bennett, 1983; Dabrowska & Doksum, 1988; Murphy et al., 1997). The two models are special cases of linear transformation models, which provide many useful alternatives. It is thus desirable to seek a unified estimation and inference procedure for linear transformation models. However, despite recent research progress, the problem has not been completely solved. The aim of the paper is to develop an estimation method, in line with

Cox's partial likelihood approach, that is easily implemented and has a reliable inference procedure for linear transformation models.

Let T be the failure time and Z the p-dimensional covariate. Without loss of generality, we assume throughout this paper that T is a positive continuous random variable. Linear transformation models assume that

$$H(T) = -\beta' Z + \varepsilon, \tag{1}$$

where H is an unknown monotone transformation function, ε is a random variable with a known distribution and is independent of Z, and β is an unknown p-dimensional regression parameter of interest. The proportional hazards model and the proportional odds model are special cases of (1) with ε following the extreme-value distribution and the standard logistic distribution, respectively. In the presence of censoring, we assume conditional independence of T and the censoring variable C given Z. Let $\tilde{T} = \min(T, C)$ be the event time and $\delta = I(T \leqslant C)$ be the failure/censoring index. The observations are then independent copies of $(\tilde{T}, \delta, Z)$.

Cheng et al. (1995) proposed and justified a general estimation method for linear transformation models with censored data. The method was further developed in Cheng et al. (1997), Fine et al. (1998) and Cai et al. (2000). A key step in their approach is the estimation of the survival function for the censoring variable by the Kaplan–Meier estimator. Its validity relies on the assumption that the censoring variable is independent of the covariates. Thus, unlike Cox's partial likelihood approach, this method of estimation fails when this assumption is violated. In practice, however, such an assumption is often too restrictive, even for randomised clinical trials.

The estimation procedure proposed in this paper is valid without the aforementioned independence assumption. In fact, our estimator reduces to the Cox partial likelihood estimator in the special case of the Cox model. It is easy to compute the estimator through an estimating function that resembles the Cox partial likelihood score. Furthermore, its asymptotic variance has a closed-form expression, so that variance estimation is simple and reliable through plug-in rules.

The next section derives the estimator of the regression parameter and establishes its consistency and asymptotic normality. Some algorithms and numerical studies are presented in § 3. Section 4 contains a brief discussion. Proofs are presented in the Appendix.

2. Estimating equations and asymptotic results

Let $(T_i, C_i, \tilde{T}_i, \delta_i, Z_i)$, for $i = 1, \ldots, n$, be independent copies of $(T, C, \tilde{T}, \delta, Z)$. The observations are $(\tilde{T}_i, \delta_i, Z_i)$, for $i = 1, \ldots, n$. Throughout this paper, $0 < t_1 < \ldots < t_K < \infty$ denote the observed K failure times among the n observations. Let $\lambda(.)$ and $\Lambda(.)$ be the known hazard and cumulative hazard functions of ε, respectively. Following the usual counting process notation, let

$$Y(t) = I(\tilde{T} \geqslant t), \quad N(t) = \delta I(\tilde{T} \leqslant t), \quad M(t) = N(t) - \int_0^t Y(s)\, d\Lambda\{\beta_0' Z + H_0(s)\},$$

where (β_0, H_0) are the true values of (β, H). Let $\{Y_i(t), N_i(t), M_i(t)\}$ be the sample analogues of $\{Y(t), N(t), M(t)\}$. Motivated by the fact that $M(t)$ is a martingale process, we consider estimating equations

$$U(\beta, H) \equiv \sum_{i=1}^{n} \int_0^\infty Z_i[dN_i(t) - Y_i(t)\, d\Lambda\{\beta' Z_i + H(t)\}] = 0, \tag{2}$$

$$\sum_{i=1}^{n} [dN_i(t) - Y_i(t)\, d\Lambda\{\beta' Z_i + H(t)\}] = 0 \quad (t \geqslant 0), \tag{3}$$

where H is a nondecreasing function satisfying $H(0) = -\infty$. The last requirement ensures that $\Lambda\{a + H(0)\} = 0$ for any finite a. Let $\mathscr{H}$ be the collection of all nondecreasing step functions on $[0, \infty)$ with $H(0) = -\infty$ and with jumps only at the observed failure times $t_1, \ldots, t_K$. We denote by $(\hat{\beta}, \hat{H})$ the solution of (2)–(3). It is then clear that $\hat{H} \in \mathscr{H}$.

Consider the special case of the Cox model, in which $\lambda(t) = \Lambda(t) = \exp(t)$. It then follows from (3) that $d[\exp\{H(t)\}] = \sum_{i=1}^{n} dN_i(t) / \sum_{i=1}^{n} \{Y_i(t) \exp(\beta' Z_i)\}$. If we plug this into (2), we obtain

$$\sum_{i=1}^{n} \int_0^\infty \left\{ Z_i - \frac{\sum_{j=1}^{n} Z_j Y_j(t) \exp(\beta' Z_j)}{\sum_{j=1}^{n} Y_j(t) \exp(\beta' Z_j)} \right\} dN_i(t) = 0,$$

which is precisely the Cox partial likelihood score equation.

There are some alternative versions of (3) that are simple for computational purposes. Note that (3) can be rewritten as

$$\begin{pmatrix} 1 - \sum_{i=1}^{n} Y_i(t_1)[\Lambda\{\beta' Z_i + H(t_1)\} - \Lambda\{\beta' Z_i + H(t_1 -)\}] \\ \vdots \\ 1 - \sum_{i=1}^{n} Y_i(t_K)[\Lambda\{\beta' Z_i + H(t_K)\} - \Lambda\{\beta' Z_i + H(t_K -)\}] \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \tag{4}$$

with $H \in \mathscr{H}$, implying $H(t_1 -) = -\infty$. Slightly differently from (4), one might also consider the following computationally simpler estimating equations:

$$\begin{pmatrix} 1 - \sum_{i=1}^{n} Y_i(t_1) \Lambda\{\beta' Z_i + H(t_1)\} \\ 1 - \sum_{i=1}^{n} Y_i(t_2) \lambda\{\beta' Z_i + H(t_2 -)\} \Delta H(t_2) \\ \vdots \\ 1 - \sum_{i=1}^{n} Y_i(t_K) \lambda\{\beta' Z_i + H(t_K -)\} \Delta H(t_K) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{5}$$

where $H \in \mathscr{H}$ and $\Delta H(t) = H(t) - H(t-)$. Algorithms based on (4) and (5) will be discussed in the next section. In the following proposition, we show the consistency and asymptotic normality of $\hat{\beta}$. It is not difficult to show that the solution of (2) with (5) and that of (2) with (4) are asymptotically equivalent.

Some notation and regularity conditions are needed. Let $\tau = \inf\{t : \mathrm{pr}(\tilde{T} > t) = 0\}$, and let $\psi(t) = (\partial/\partial t) \log \lambda(t) = \dot{\lambda}(t)/\lambda(t)$. The superscript dot always denotes derivatives. We assume the positivity of $\lambda(.)$, the continuity of $\psi(.)$ and that $\lim_{s \to -\infty} \lambda(s) = 0 = \lim_{s \to -\infty} \psi(s)$. Furthermore, we assume that $\mathrm{pr}(|Z| < m) = 1$ for some constant $m > 0$ and that H_0 has continuous and positive derivatives. For any $t, s \in (0, \tau]$, define

$$B(t, s) = \exp\left(\int_s^t \frac{E[\dot{\lambda}\{\beta_0' Z + H_0(x)\} Y(x)]}{E[\lambda\{\beta_0' Z + H_0(x)\} Y(x)]}\, dH_0(x) \right),$$

$$\mu_Z(t) = \frac{E[Z \lambda\{\beta_0' Z + H_0(\tilde{T})\} Y(t) B(t, \tilde{T})]}{E[\lambda\{\beta_0' Z + H_0(t)\} Y(t)]},$$

$$\Sigma^* = \int_0^\tau E[\{Z - \mu_Z(t)\}^{\otimes 2} \lambda\{\beta_0' Z + H_0(t)\} Y(t)]\, dH_0(t), \tag{6}$$

$$\Sigma_* = \int_0^\tau E[\{Z - \mu_Z(t)\}Z'\dot{\lambda}\{\beta_0'Z + H_0(t)\}Y(t)]\, dH_0(t), \tag{7}$$

where $b^{\otimes 2} = bb'$ for any vector b. Assume that Σ_* and Σ^* are finite and nondegenerate. Some additional technical conditions are presented in the Appendix.

PROPOSITION. *Under suitable regularity conditions, we have that*

$$n^{\frac{1}{2}}(\hat{\beta} - \beta_0) \to N\{0, \Sigma_*^{-1}\Sigma^*(\Sigma_*^{-1})'\} \tag{8}$$

in distribution, as $n \to \infty$. *Moreover,* Σ_* *and* Σ^* *can be consistently estimated by*

$$\hat{\Sigma}^* = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{Z_i - \bar{Z}(t)\}^{\otimes 2}\lambda\{\hat{\beta}'Z_i + \hat{H}(t)\}Y_i(t)\, d\hat{H}(t),$$

$$\hat{\Sigma}_* = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{Z_i - \bar{Z}(t)\}Z_i'\dot{\lambda}\{\hat{\beta}'Z_i + \hat{H}(t)\}Y_i(t)\, d\hat{H}(t),$$

respectively, where

$$\bar{Z}(t) = \frac{\sum_{i=1}^n Z_i\lambda\{\hat{\beta}'Z_i + \hat{H}(Y_i)\}Y_i(t)\hat{B}(t, \tilde{T}_i)}{\sum_{i=1}^n \lambda\{\hat{\beta}'Z_i + \hat{H}(t)\}Y_i(t)},$$

$$\hat{B}(t, s) = \exp\left(\int_s^t \frac{\sum_{i=1}^n \dot{\lambda}\{\hat{\beta}'Z_i + \hat{H}(x)\}Y_i(x)}{\sum_{i=1}^n \lambda\{\hat{\beta}'Z_i + \hat{H}(x)\}Y_i(x)}\, d\hat{H}(x)\right),$$

for $t, s \in [0, \tau]$.

In the special case of the proportional hazards model, it is easily verified that

$$B(t, s) = \exp\{H_0(t) - H_0(s)\}, \quad \mu_Z(t) = E(Z \mid \tilde{T} = t, \delta = 1),$$

$$\Sigma^* = \Sigma_* = \mathrm{var}\left[\int_0^\infty \{Z - \mu_Z(t)\}\, dM(t)\right],$$

which is the Fisher information matrix based on the Cox partial likelihood score.

It is proved in Step A5 of the Appendix that consistency of $\hat{H}$ holds in the sense that

$$\sup(|\exp\{\hat{H}(t)\} - \exp\{H_0(t)\}| : t \in [0, \tau]) \to 0$$

in probability as $n \to \infty$. This convergence is equivalent to

$$\sup(|\hat{H}(t) - H_0(t)| : t \in [a, \tau]) \to 0$$

in probability as $n \to \infty$ for any fixed $a \in (0, \tau]$.

3. COMPUTATIONAL ALGORITHMS AND NUMERICAL STUDIES

It is easy to show that, for every fixed value of β, the solution of (3), (4) or (5) is unique in $\mathscr{H}$. Equations (2) and (4) naturally suggest the following iterative algorithms for computing $(\hat{\beta}, \hat{H})$.

Step 0. Choose an initial value of β, denoted by $\beta^{(0)}$.

Step 1. Obtain $H^{(0)}$ as follows. First obtain $H^{(0)}(t_1)$ by solving

$$\sum_{i=1}^n Y_i(t_1)\Lambda\{\beta'Z_i + H(t_1)\} = 1,$$

with $\beta=\beta^{(0)}$. Then, obtain $H^{(0)}(t_k)$, for $k=2,\ldots,K$, one-by-one by solving the equation

$$\sum_{i=1}^{n} Y_i(t_k)\Lambda\{\beta' Z_i + H(t_k)\} = 1 + \sum_{i=1}^{n} Y_i(t_k)\Lambda\{\beta' Z_j + H(t_k-)\} \tag{9}$$

with $\beta=\beta^{(0)}$.

Step 2. Obtain new estimate of β by solving (2) with $H=H^{(0)}$.

Step 3. Set $\beta^{(0)}$ to be the estimate obtained in Step 2 and repeat Steps 1 and 2 until prescribed convergence criteria are met.

The above procedure is based on estimating equations (2) and (4); a similar procedure can be derived based on estimating equations (2) and (5), the only difference being that, instead of solving (9), one calculates

$$H(t_k) = H(t_k-) + \frac{1}{\sum_{i=1}^{n} Y_i(t_k)\lambda\{\beta' Z_i + H(t_k-)\}}.$$

Since this involves only direct calculation, it is much easier than solving (9).

We present two simulation examples. In both examples, we choose $H_0(t)=\log(t)$ and let the hazard function of ε be of the form $\lambda(t)=\exp(t)/\{1+r\exp(t)\}$, with $r=0$, 0·5, 1, 1·5 and 2 (Dabrowska & Doksum, 1988). Note that the proportional hazards and proportional odds models correspond to $r=0$ and $r=1$, respectively. Two covariates z_1 and z_2 are chosen to be independent of each other with z_1 following the standard normal distribution and z_2 taking values 0 or 1 with equal probability 0·5. We also set $\beta_0=(0, 1)$. Two different censoring schemes are considered. In Example 1, the censoring variable is independent of the covariates and follows the $\text{Un}(0, c)$ distribution. In Example 2, the censoring variable depends on the covariates and is set to be $-z_1 - z_2 + \text{Un}(0, c)$. In both examples, the constant c takes various values according to theoretically prespecified censoring proportions. The sample size n is set to be 100 and all simulations are based on 1000 replications. Table 1 lists the coverage probabilities for β_0, showing that the empirical coverage probabilities are very close to the nominal levels in nearly all cases.

We apply our estimation procedure to data from the Veterans' Administration lung cancer trial, which was also analysed in Prentice (1973), Bennett (1983), Pettitt (1984), Cheng et al. (1995) and Murphy et al. (1997). In our analysis, the subgroup of 97 patients with no prior therapy is used. The response is the survival time of each patient and the covariates are his/her tumour type, namely large, adeno, small or squamous, and performance status. As in the simulation examples, the hazard function of ε is of the form $\lambda(t)=\exp(t)/\{1+r\exp(t)\}$ with $r=0$, 1, 1·5 and 2. Table 2 presents the estimates of the regression parameters and their estimated standard deviations. We find that the results for the proportional odds model, $r=1$, are very similar to those reported in the literature, except for those corresponding to the covariate 'squamous versus large', which is regarded as relatively insignificant.

4. Remarks

In principle, a general efficient estimation procedure may be obtained through estimating equations that are similar to those of (2) and (3). However, such an improvement is at the cost of increasing computational complexity. Furthermore, the asymptotic variance of the efficient estimator does not in general have a closed form,

Table 1. *Empirical coverage probabilities of confidence intervals for regression coefficients*

(*a*) *Example* 1: *Covariate-independent censoring*

Censoring proportion (%)	Nominal level	$r=0$		$r=0{\cdot}5$		$r=1$		$r=1{\cdot}5$		$r=2$	
0	0·95	0·96	0·93	0·96	0·96	0·95	0·95	0·96	0·95	0·96	0·95
	0·90	0·92	0·88	0·91	0·90	0·89	0·90	0·91	0·90	0·91	0·89
	0·85	0·85	0·82	0·86	0·86	0·85	0·85	0·87	0·85	0·86	0·84
10	0·95	0·96	0·93	0·96	0·95	0·95	0·95	0·96	0·96	0·96	0·95
	0·90	0·91	0·88	0·91	0·91	0·90	0·90	0·90	0·89	0·90	0·88
	0·85	0·86	0·82	0·86	0·86	0·84	0·85	0·86	0·84	0·86	0·84
20	0·95	0·96	0·94	0·96	0·95	0·95	0·94	0·96	0·96	0·96	0·96
	0·90	0·91	0·88	0·92	0·91	0·90	0·90	0·91	0·90	0·90	0·89
	0·85	0·87	0·82	0·87	0·86	0·85	0·86	0·86	0·84	0·86	0·83
30	0·95	0·95	0·95	0·96	0·95	0·96	0·94	0·95	0·95	0·96	0·95
	0·90	0·90	0·88	0·91	0·90	0·91	0·90	0·90	0·90	0·90	0·89
	0·85	0·85	0·83	0·86	0·85	0·85	0·86	0·85	0·84	0·85	0·84

(*b*) *Example* 2: *Covariate-dependent censoring*

Censoring proportion (%)	Nominal level	$r=0$		$r=0{\cdot}5$		$r=1$		$r=1{\cdot}5$		$r=2$	
10	0·95	0·96	0·94	0·96	0·95	0·96	0·95	0·97	0·95	0·96	0·95
	0·90	0·92	0·88	0·91	0·90	0·91	0·90	0·92	0·90	0·92	0·89
	0·85	0·88	0·83	0·87	0·85	0·86	0·85	0·87	0·85	0·86	0·84
20	0·95	0·96	0·94	0·96	0·96	0·96	0·95	0·96	0·96	0·95	0·95
	0·90	0·92	0·88	0·91	0·90	0·92	0·90	0·91	0·89	0·91	0·89
	0·85	0·88	0·83	0·87	0·86	0·86	0·85	0·87	0·85	0·86	0·85
30	0·95	0·96	0·95	0·96	0·95	0·96	0·95	0·95	0·95	0·96	0·95
	0·90	0·90	0·89	0·92	0·92	0·90	0·91	0·91	0·90	0·90	0·90
	0·85	0·86	0·84	0·86	0·88	0·86	0·86	0·86	0·84	0·86	0·85

In each row, first entry is empirical coverage for the first component of β, second entry is empirical coverage for the second component of β.

Table 2. *Estimates of regression coefficients and, in brackets, their estimated standard deviations for the Veterans' Administration lung cancer data*

	$r=0$	$r=1$	$r=1{\cdot}5$	$r=2$
Performance status	−0·024 (0·007)	−0·044 (0·011)	−0·055 (0·014)	−0·065 (0·016)
Adeno versus large	0·851 (0·350)	1·503 (0·528)	1·829 (0·632)	2·164 (0·742)
Small versus large	0·548 (0·333)	1·230 (0·479)	1·531 (0·550)	1·828 (0·622)
Squamous versus large	−0·214 (0·361)	−0·469 (0·614)	−0·595 (0·760)	−0·716 (0·910)

so that a simple inference procedure is not available. A study of efficient estimation will be reported elsewhere.

Linear transformation models in (1) are equivalent to multiplicative transformation models with a log transformation. The estimation and inference procedure proposed in this paper can also be generalised in a straightforward fashion to slightly more general models with $H(t)=g(\beta'Z,\varepsilon)$, where g is a known smooth function.

Acknowledgement

The authors are grateful to the editor for his effort that led to numerous improvements in presentation. This research was supported in part by the Research Grants Council of Hong Kong, the New York City Council Speaker's Fund for Public Health Research and the U.S. National Institutes of Health, National Science Foundation and National Security Agency.

Appendix

Proof of the Proposition

Regularity conditions for ensuring the central limit theorem for counting process martingales such as those assumed in Fleming & Harrington (1991) are also assumed here without specific statement. In particular, we assume that τ is finite, $\text{pr}(T > \tau) > 0$ (Fleming & Harrington, 1991, pp. 289–90) and $\text{pr}(C = \tau) > 0$ (Murphy et al., 1997). This is to avoid a lengthy technical discussion about the tail behaviour. Some nonessential technical details in the proof are omitted for brevity of presentation. The proofs are divided into six steps.

Step A1. Here we show that $d\{\tilde{H}(., \beta_0) - H_0(.)\} \to 0$ almost surely, where $\hat{H}(., \beta) \in \mathscr{H}$ is the function implicitly defined as the unique solution of (3) for fixed β and $d(.,.)$ is a distance defined as follows. For any two nondecreasing functions H_1 and H_2 on $[0, \tau]$ such that $H_1(0) = H_2(0) = -\infty$, define

$$d(H_1, H_2) = \sup(|\exp\{H_1(t)\} - \exp\{H_2(t)\}| : t \in [0, \tau]).$$

Let A be a mapping on $\mathscr{H}$ defined by

$$A(H)(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t [dN_i(s) - Y_i(s)\, d\Lambda\{H(s) + \beta_0' Z_i\}].$$

For an arbitrary but fixed $\varepsilon_* > 0$, consider $H_1 \in \mathscr{H}$ and $H_2 \in \mathscr{H}$ such that $d(H_1, H_2) \geqslant \varepsilon_*$. There exists a $t_* \in [0, \tau]$ such that $|\exp\{H_1(t_*)\} - \exp\{H_2(t_*)\}| \geqslant \varepsilon_*/2$. Then

$$\begin{aligned}
\sup_{t\in[0,\tau]} |A(H_1)(t) - A(H_2)(t)|& \\
&= \frac{1}{n}\sup_{t\in[0,\tau]}\left|\sum_{i=1}^{n}[\Lambda\{H_1(t\wedge\tilde{T}_i) + \beta_0' Z_i\} - \Lambda\{H_2(t\wedge\tilde{T}_i) + \beta_0 Z_i\}]\right| \\
&\geqslant \frac{1}{n}\left|\sum_{i=1}^{n}\int_{H_2(t_*\wedge\tilde{T}_i)}^{H_1(t_*\wedge\tilde{T}_i)} \lambda(s + \beta_0' Z_i)\, ds\right| \\
&\geqslant \frac{1}{n}\left|\sum_{i=1}^{n} I(\tilde{T}_i = \tau)\int_{H_2(t_*)}^{H_1(t_*)} \lambda(s + \beta_0' Z_i)\, ds\right| \\
&\geqslant \frac{1}{n}\sum_{i=1}^{n} I(\tilde{T}_i = \tau)\inf\left\{\int_{\log a}^{\log b}\lambda(s + c)\, ds : 0 < a < b < m, b - a < \varepsilon_*/2, c < m\right\},
\end{aligned}$$

where m is a large but fixed constant. Since λ is a positive function, the above infimum is a positive number which does not depend on the choice of H_1 and H_2. The law of large numbers implies that

$$\sup(|A(H_1)(t) - A(H_2)(t)| : t \in [0, \tau]) > \tilde{\varepsilon}$$

for some fixed constant $\tilde{\varepsilon} > 0$, all $H_1 \in \mathscr{H}$ and $H_2 \in \mathscr{H}$ such that $d(H_1, H_2) \geqslant \varepsilon_*$ and all large n. Choose $H_0^* \in \mathscr{H}$ such that $H_0^*(t_i) = H_0(t_i)$ $(i = 1, \ldots, K)$. The law of large numbers and the continuity of H_0 imply that $\sup(|A(H_0^*)(t)| : t \in [0, \tau]) \to 0$ almost surely. It follows that

$$\sup(|A(H_0^*)(t) - A\{\hat{H}(., \beta_0)\}(t)| : t \in [0, \tau]) \to 0$$

almost surely because, by definition of $\hat{H}(.,\beta_0)$, $A\{\hat{H}(.,\beta_0)\}(t)=0$ for all $t\in[0,\tau]$. Then, with probability 1, $\hat{H}(.,\beta_0)$ is in the neighbourhood of H_0^* of radius ε_* under the metric $d(.,.)$. Therefore, $d\{\hat{H}(.,\beta_0), H_o\}\to 0$ almost surely, since $\varepsilon_*>0$ can be arbitrarily small and $\hat{H}(.,\beta_0)$ and H_0 are monotone.

Step A2. Here we construct a representative of $\Lambda^*\{\hat{H}(.,\beta_0)\}$. Let $a>0$ and b be fixed finite numbers and define

$$\lambda^*\{H_0(t)\}=B(t,a),\quad B_1(t)=\int_a^t E[\dot{\lambda}\{\beta_0'Z+H_0(s)\}Y(s)]\,dH_0(s),$$

$$B_2(t)=E[\lambda\{\beta_0'Z+H_0(t)\}Y(t)],\quad \Lambda^*(x)=\int_b^x \lambda^*(s)\,ds,$$

for $t>0$ and $x\in(-\infty,\infty)$. We choose finite $a>0$ and b as the lower limits of the integration to ensure that the integrals are finite. It is easy to see that $B(t,s)=\lambda^*\{H_0(t)\}/\lambda^*\{H_0(s)\}$. Observe that $d\lambda^*\{H_0(t)\}=[\lambda^*\{H_0(t)\}/B_2(t)]\,dB_1(t)$ and write

$$\begin{aligned}
-n^{-1}\sum_{i=1}^n M_i(t) &= -n^{-1}\sum_{i=1}^n\int_0^t Y_i(s)[d\Lambda\{\beta_0'Z_i+\hat{H}(s,\beta_0)\}-d\Lambda\{\beta_0'Z_i+H_0(s)\}]\\
&= -n^{-1}\sum_{i=1}^n\int_0^t Y_i(s)d\left(\frac{\lambda\{\beta_0'Z_i+H_0(s)\}}{\lambda^*\{H_0(s)\}}[\Lambda^*\{\hat{H}(s,\beta_0)\}-\Lambda^*\{H_0(s)\}]\right)+o_p(n^{-1/2})\\
&= -\int_0^t \frac{1}{n}\sum_{i=1}^n \frac{Y_i(s)\lambda\{\beta_0'Z_i+H_0(s)\}}{\lambda^*\{H_0(s)\}}[d\Lambda^*\{\hat{H}(s,\beta_0)\}-d\Lambda^*\{H_0(s)\}]\\
&\quad -\int_0^t [\Lambda^*\{\hat{H}(s,\beta_0)\}-\Lambda^*\{H_0(s)\}]\frac{1}{n}\sum_{i=1}^n Y_i(s)d\left[\frac{\lambda\{\beta_0'Z_i+H_0(s)\}}{\lambda^*\{H_0(s)\}}\right]+o_p(n^{-1/2})\\
&= -\int_0^t \frac{B_2(s)}{\lambda^*\{H_0(s)\}}[d\Lambda^*\{\hat{H}(s,\beta_0)\}-d\Lambda^*\{H_0(s)\}]\\
&\quad +\int_0^t [\Lambda^*\{\hat{H}(s,\beta_0)\}-\Lambda^*\{H_0(s)\}]\frac{\lambda^*\{H_0(s)\}\,dB_1(s)-B_2(s)\,d\lambda^*\{H_0(s)\}}{[\lambda^*\{H_0(s)\}]^2}+o_p(n^{-1/2})\\
&= -\int_0^t \frac{B_2(s)}{\lambda^*\{H_0(s)\}}[d\Lambda^*\{\hat{H}(s,\beta_0)\}-d\Lambda^*\{H_0(s)\}]+o_p(n^{-1/2}).
\end{aligned}$$

Therefore, for $t\in[0,\tau]$,

$$\Lambda^*\{\hat{H}(t,\beta_0)\}-\Lambda^*\{H_0(t)\}=\frac{1}{n}\sum_{i=1}^n\int_0^t\frac{\lambda^*\{H_0(s)\}}{B_2(s)}\,dM_i(s)+o_p(n^{-1/2}).$$

Step A3. Here we show that $(1/n)(\partial/\partial\beta)U\{\beta,\hat{H}(.,\beta)\}$ at $\beta=\beta_0$ converges to $-\Sigma_*$ in probability. Differentiating (3) with respect to β, we have that

$$\sum_{i=1}^n\int_0^t Y_i(s)d\left[\lambda\{\beta'Z_i+\hat{H}(s,\beta)\}\left\{Z_i+\frac{\partial}{\partial\beta}\hat{H}(s,\beta)\right\}\right]=0$$

for every $t\in[0,\tau]$. Similarly to Step A2, one can show that

$$\frac{\partial}{\partial\beta}\hat{H}(t,\beta)|_{\beta=\beta_0}=-\frac{1}{\lambda^*\{H_0(t)\}}\int_0^t\lambda^*\{H_0(s)\}\frac{E[Z\dot{\lambda}\{\beta_0'Z+H_0(s)\}Y(s)]}{B_2(s)}\,dH_0(s)+o_p(1).$$

It follows from the law of large numbers that

$$\frac{1}{n}\frac{\partial}{\partial\beta}U\{\beta,\hat{H}(.,\beta)\}|_{\beta=\beta_0} = -\frac{1}{n}\sum_{i=1}^{n} Z_i\lambda\{\beta_0'Z_i+\hat{H}(\tilde{T}_i,\beta_0)\}\left\{Z_i+\frac{\partial}{\partial\beta}\hat{H}(\tilde{T}_i,\beta)|_{\beta=\beta_0}\right\}'$$

$$= -\frac{1}{n}\sum_{i=1}^{n} Z_i\lambda\{\beta_0'Z_i+\hat{H}(\tilde{T}_i,\beta_0)\}$$

$$\times\left(Z_i' - \int_0^{\tilde{T}_i}\frac{\lambda_0^*\{H_0(s)\}E[Z'\dot{\lambda}\{\beta_0'Z+H_0(s)\}Y(s)]}{\lambda^*\{H_0(\tilde{T}_i)\}B_2(s)}\,dH_0(s)\right)+o_p(1)$$

$$= -\int_0^{\tau} E\left\{\left(Z-\frac{E[Z\lambda\{\beta_0'Z+H_0(\tilde{T})\}Y(s)/\lambda^*\{H_0(\tilde{T})\}]}{B_2(s)/\lambda^*\{H_0(s)\}}\right)\right.$$

$$\left.\times[Z'\dot{\lambda}\{\beta_0'Z+H_0(s)\}Y(s)]\right\}dH_0(s)+o_p(1)$$

$$=\Sigma_* + o_p(1),$$

where Σ_* is defined in (7).

Step A4. Here we show the asymptotic normality of $U\{\beta_0,\hat{H}(.,\beta_0)\}$. Using the results of Steps A1 and A2 and some empirical process approximation techniques, we can write

$$U\{\beta_0,\hat{H}(.,\beta_0)\}$$

$$=\sum_{i=1}^{n}\int_0^{\tau} Z_i[dN_i(t)-Y_i(t)\,d\Lambda\{\beta_0'Z_i+\hat{H}(t,\beta_0)\}]$$

$$=\sum_{i=1}^{n}\int_0^{\tau} Z_i\,dM_i(t)-\sum_{i=1}^{n} Z_i[\Lambda\{\beta_0'Z_i+\hat{H}(\tilde{T}_i,\beta_0)-\Lambda\{\beta_0'Z_i+H_0(\tilde{T}_i)\}]$$

$$=\sum_{i=1}^{n}\int_0^{\tau} Z_i\,dM_i(t)-\sum_{i=1}^{n}\frac{Z_i\lambda\{\beta_0'Z_i+H_0(\tilde{T}_i)\}}{\lambda^*\{H_0(\tilde{T}_i)\}}[\Lambda^*\{\hat{H}(\tilde{T}_i,\beta_0)\}-\Lambda^*\{H_0(\tilde{T}_i)\}]+o_p(n^{1/2})$$

$$=\sum_{i=1}^{n}\int_0^{\tau} Z_i\,dM_i(t)-\sum_{i=1}^{n}\frac{Z_i\lambda\{\beta_0'Z_i+H_0(\tilde{T}_i)\}}{\lambda^*\{H_0(\tilde{T}_i)\}}\frac{1}{n}\sum_{j=1}^{n}\int_0^{\tilde{T}_i}\frac{\lambda^*\{H_0(t)\}}{B_2(t)}\,dM_j(t)+o_p(n^{1/2})$$

$$=\sum_{i=1}^{n}\int_0^{\tau} Z_i\,dM_i(t)-\sum_{j=1}^{n}\int_0^{\tau} E\left[\frac{Z\lambda\{\beta_0'Z+H_0(\tilde{T})\}Y(t)}{\lambda^*\{H_0(\tilde{T})\}}\right]\frac{\lambda^*\{H_0(t)\}}{B_2(t)}\,dM_j(t)+o_p(n^{1/2})$$

$$=\sum_{i=1}^{n}\int_0^{\tau}\left(Z_i-E\left[\frac{Z\lambda\{\beta_0'Z+H_0(\tilde{T})\}Y(t)}{\lambda^*\{H_0(\tilde{T})\}}\right]\frac{\lambda^*\{H_0(t)\}}{B_2(t)}\right)dM_i(t)+o_p(n^{1/2})$$

$$=\sum_{i=1}^{n}\int_0^{\tau}\{Z_i-\mu_Z(t)\}\,dM_i(t)+o_p(n^{1/2}).$$

It then follows that $N^{-1/2}U\{\beta_0,\hat{H}(.,\beta_0)\}\rightarrow N(0,\Sigma^*)$, where Σ^* is defined in (6).

Step A5. Here we show consistency. Consider $(1/n)U\{\beta,\hat{H}(.,\beta)\}$ as a mapping from an arbitrarily small but fixed ball in R^p centred at β_0 to another open connected set in R^p, where p is the dimension of β. Observe that Σ_* is assumed to be a nondegenerate matrix. The proof of Step A3 can be slightly strengthened to show that convergence similar to the result of Step A3 also holds over this ball. Consequently, the probability that the mapping of this ball is homeomorphic tends to 1 as $n\rightarrow\infty$. The fact that $(1/n)U\{\beta_0,\hat{H}(.,\beta_0)\}\rightarrow 0$ in probability then implies that the probability that $\hat{\beta}$ is inside the ball tends to 1. This proves that $\hat{\beta}$ is consistent since this ball is centred at β_0 and can be arbitrarily small. With the consistency of $\hat{\beta}$, one can mimic the proof in Step A1 to show that $\hat{H}$ is also consistent under the metric $d(.,.)$ defined in Step A1.

Step A6. It follows from the Taylor expansion and the results of Steps A3 and A5 that (8) holds. Consistency of the variance estimators can be shown in a straightforward fashion and is omitted here. The proof is complete.

REFERENCES

ANDERSON, P. K. & GILL, R. (1982). Cox's regression model for counting process: a large sample study. *Ann. Statist.* **10**, 1100–20.
BENNETT, S. (1983). Analysis of survival data by the proportional odds model. *Statist. Med.* **2**, 273–7.
CAI, T., WEI, L. J. & WILCOX, M. (2000). Semiparametric regression analysis for clustered failure time data. *Biometrika* **87**, 867–78.
CHENG, S. C., WEI, L. J. & YING, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835–45.
CHENG, S. C., WEI, L. J. & YING, Z. (1997). Prediction of survival probabilities with semi-parametric transformation models. *J. Am. Statist. Assoc.* **92**, 227–35.
COX, D. R. (1972). Regression models and life-tables (with Discussion). *J. R. Statist. Soc.* B **34**, 187–220.
COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–76.
DABROWSKA, D. M. & DOKSUM, K. A. (1988). Estimation and testing in the two-sample generalized odds-rate model. *J. Am. Statist. Assoc.* **83**, 744–9.
FINE, J. P., YING, Z. & WEI, L. J. (1998). On the linear transformation model with censored data. *Biometrika* **85**, 980–6.
FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.
MURPHY, S. A., ROSSINI, A. J. & VAN DER VAART, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *J. Am. Statist. Assoc.* **92**, 968–76.
PETTITT, A. N. (1982). Inference for the linear model using a likelihood based on ranks. *J. R. Statist. Soc.* B **44**, 234–43.
PETTITT, A. N. (1984). Proportional odds model for survival data and estimates using ranks. *Appl. Statist.* **33**, 169–75.
PRENTICE, R. L. (1973). Exponential survivals with censoring and explanatory variables. *Biometrika* **60**, 279–88.
TSIATIS, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93–108.

[*Received March* 2001. *Revised February* 2002]