Biometrika (2006), **93,** 3, *pp.* 627–640

Printed in Great Britain

Efficient estimation of semiparametric transformation models for counting processes

BY DONGLIN ZENG AND D. Y. LIN

Department of Biostatistics, CB# 7420, University of North Carolina, Chapel Hill, North Carolina 27599-7420, U.S.A.

dzeng@bios.unc.edu lin@bios.unc.edu

SUMMARY

A class of semiparametric transformation models is proposed to characterise the effects of possibly time-varying covariates on the intensity functions of counting processes. The class includes the proportional intensity model and linear transformation models as special cases. Nonparametric maximum likelihood estimators are developed for the regression parameters and cumulative intensity functions of these models based on censored data. The estimators are shown to be consistent and asymptotically normal. The limiting variances for the estimators of the regression parameters achieve the semiparametric efficient bounds and can be consistently estimated. The limiting variances for the estimators of smooth functionals of the cumulative intensity function can also be consistently estimated. Simulation studies reveal that the proposed inference procedures perform well in practical settings. Two medical studies are provided.

Some key words: Censoring; Intensity; Linear transformation model; Nonparametric likelihood; Proportional hazards; Proportional odds; Recurrent events; Semiparametric efficiency; Survival data; Time-varying covariate.

1. INTRODUCTION

Counting processes have been used extensively to describe event history data. For example, traditional survival data can be characterised as a counting process with a single jump at the survival time while recurrent events can be characterised as a counting process with jumps at recurrent event times. A number of statistical models formulate the effects of covariates on counting processes (Andersen et al., 1993), the most popular choice being the proportional intensity model (Andersen & Gill, 1982).

Let $N^*(t)$ be the counting process recording the number of events that have occurred by time t, and let $Z(t)$ be a vector of possibly time-varying covariates. The proportional intensity model specifies that the intensity function for $N^*(t)$ conditional on $Z(t)$ takes the form

$$\Lambda_Z(t) = \int_0^t Y^*(s) e^{\beta^{\mathrm{T}} Z(s)} d\Lambda(s), \tag{1}$$

where $Y^*(.)$ is a predictable process with values 0 and 1, $\Lambda(.)$ is an unspecified increasing function, and β is a vector of unknown regression parameters. For survival data, $Y^*(t) = I(T \geqslant t)$, where T is the survival time and $I(.)$ is the indicator function; for

recurrent events, $Y^*(.) = 1$. A large-sample estimation theory for this model based on the partial likelihood principle (Cox, 1972, 1975) has been established elegantly via the counting-process martingale theory (Andersen & Gill, 1982).

For survival data, model (1) is equivalent to the classical proportional hazards model (Cox, 1972). The proportional hazards assumption may be violated in certain applications, especially in long-term studies of chronic diseases. A useful alternative is the proportional odds model (Bennett, 1983; Pettitt, 1984), which constrains the ratio of the odds of survival associated with two sets of covariate values to be constant over time and consequently the ratio of the hazards to converge to unity as time increases. By contrast, the proportional hazards model constrains the hazard ratio to be constant while the odds ratio tends to zero or infinity. Physical and biological rationale for the proportional odds model was provided by Bennett (1983) among others. Maximum likelihood estimation for this model was studied by Murphy et al. (1997).

The proportional hazards and proportional odds models belong to the class of semi-parametric linear transformation models (Dabrowska & Doksum, 1988). General estimators for this class of models were proposed by Dabrowska & Doksum (1988), Cheng et al. (1995, 1997) and Chen et al. (2002) among others. None of these estimators is asymptotically efficient. For a subset of the models called generalised odds-rate models, maximum likelihood estimation was studied by Scharfstein et al. (1998). The class of linear transformation models is confined to survival data and does not allow time-varying covariates.

In the present paper, we consider a broad class of transformation models for general counting processes, which can accommodate time-varying covariates and recurrent events. The models incorporate a transformation into the right-hand side of equation (1):

$$\Lambda_Z(t) = G\left\{\int_0^t Y^*(s)e^{\beta^{\mathrm{T}}Z(s)}d\Lambda(s)\right\}, \tag{2}$$

where G is a thrice continuously differentiable and strictly increasing function with $G(0) = 0$, $G'(0) > 0$ and $G(\infty) = \infty$. Here and in the sequel, $f'(x) = df(x)/dx$. Of course, the choice of $G(x) = x$ yields model (1). When $N^*(.)$ has a single jump at survival time T, equation (2) implies that $G\{\int_0^t e^{\beta^{\mathrm{T}}Z(s)}d\Lambda(s)\}$ is a cumulative hazard function so that

$$\int_0^T e^{\beta^{\mathrm{T}}Z(s)}d\Lambda(s) = G^{-1}(-\log \varepsilon_0),$$

where ε_0 has a uniform distribution. If Z is time-invariant, then the above equation becomes the linear transformation model

$$\log \Lambda(T) = -\beta^{\mathrm{T}}Z + \log G^{-1}(-\log \varepsilon_0).$$

It is natural to consider the class of Box–Cox transformations, in which

$$G(x) = \{(1+x)^\rho - 1\}/\rho \quad (\rho \geqslant 0)$$

with $\rho = 0$ corresponding to $G(x) = \log(1+x)$. Chen et al. (2002) considered a class of logarithmic transformations:

$$G(x) = \log(1+rx)/r \quad (r \geqslant 0)$$

with $r = 0$ corresponding to $G(x) = x$. The choice of $\rho = 1$ or $r = 0$ yields the proportional hazards/intensity model while the choice of $\rho = 0$ or $r = 1$ yields the proportional odds

model. Figure 1 shows the patterns of covariate effects over time for these two classes of transformations. For the first class, covariate effects increase over time if $\rho > 1$ and decrease over time if $\rho < 1$. For the second class, covariate effects always decrease over time, the rate of decrease being higher for larger r.

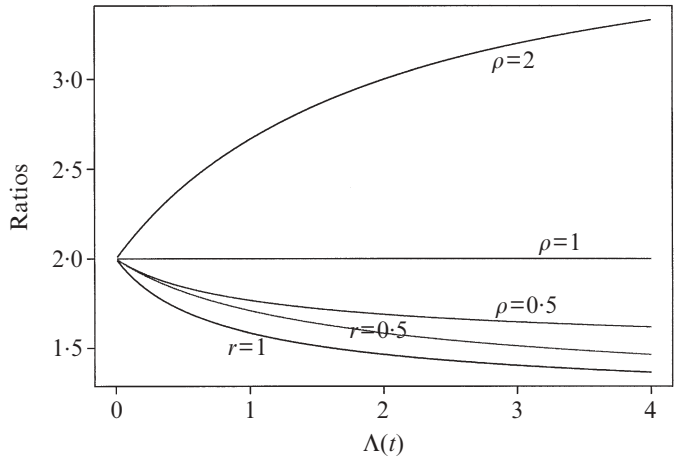

Fig. 1. Plots of the ratios $\Lambda_z(t)/\Lambda_{Z=0}(t)$ against $\Lambda(t)$ with $e^{\beta^{\mathrm{T}}z} = 2$ under the Box–Cox and logarithmic transformations.

We propose to estimate the parameters β and $\Lambda(.)$ in (2) by the nonparametric maximum likelihood method (Andersen et al., 1993, § IV.1.5; Bickel et al., 1993, pp. 339–44), the theoretical development relying on modern empirical process theory (van der Vaart & Wellner, 1996) and semiparametric efficiency theory (Bickel et al., 1993).

2. Inference procedures

Counting processes are commonly subject to right censoring. Let C denote the censoring time, which is assumed to be independent of $N^*(.)$ conditional on $Z(.)$. For a random sample of n subjects, the data consist of

$$\{N_i(t),\, Y_i(t),\, Z_i(t);\, t \in [0, \tau]\} \quad (i = 1, \ldots, n),$$

where $Y_i(t) = I(C_i \geqslant t)Y_i^*(t)$, $N_i(t) = N_i^*(t \wedge C_i)$, $a \wedge b = \min(a, b)$, and τ denotes the duration of the study. For general censoring/truncation patterns, we define $N_i(t)$ to be the number of events observed by time t on the ith subject, and $Y_i(t)$ the indicator of whether or not the ith subject is at risk at time t.

Under model (2), the intensity for $N_i(t)$ is $Y_i(t)e^{\beta^{\mathrm{T}}Z_i(t)}\lambda(t)G'\{\int_0^t Y_i(s)e^{\beta^{\mathrm{T}}Z_i(s)}d\Lambda(s)\}$, where $\lambda(t) = \Lambda'(t)$. Thus, the loglikelihood function concerning the parameters $\Lambda(.)$ and β can be written as

$$\sum_{i=1}^{n}\left[\int_0^\tau \log\lambda(t)dN_i(t) + \int_0^\tau \log G'\left\{\int_0^t Y_i(s)e^{\beta^{\mathrm{T}}Z_i(s)}d\Lambda(s)\right\}dN_i(t) + \int_0^\tau \beta^{\mathrm{T}}Z_i(t)dN_i(t) - G\left\{\int_0^\tau Y_i(t)e^{\beta^{\mathrm{T}}Z_i(t)}d\Lambda(t)\right\}\right]. \tag{3}$$

The maximum of this function does not exist if $\Lambda(.)$ is restricted to be absolutely continuous. Thus, we allow $\Lambda(.)$ to be discrete and replace $\lambda(t)$ in (3) with the jump size of Λ at time t, denoted by $\Lambda\{t\}$. The modified loglikelihood function takes the form

$$l_n(\Lambda, \beta) = \sum_{i=1}^{n} \left[\int_0^\tau \log \Lambda\{t\} dN_i(t) + \int_0^\tau \log G' \left\{ \int_0^t Y_i(s) e^{\beta^{\mathrm{T}} Z_i(s)} d\Lambda(s) \right\} dN_i(t) \right.$$
$$\left. + \int_0^\tau \beta^{\mathrm{T}} Z_i(t) dN_i(t) - G\left\{ \int_0^\tau Y_i(t) e^{\beta^{\mathrm{T}} Z_i(t)} d\Lambda(t) \right\} \right]. \quad (4)$$

We maximise (4) over $\Lambda(.)$ and β, restricting $\Lambda(.)$ to be a step function with jumps at the observed event times X_{ij} $(i = 1, \ldots, n; j = 1, \ldots, n_i)$, where n_i is the number of observed events on the ith subject. This is tantamount to maximising (4) over $\Lambda\{X_{ij}\}$ $(i = 1, \ldots, n; j = 1, \ldots, n_i)$ and β. The resulting estimators are referred to as the nonparametric maximum likelihood estimators. In the special case of the proportional hazards/intensity model, these estimators are identical to the maximum partial likelihood estimators (Andersen et al., 1993, pp. 481–3).

The proposed estimators can be obtained by the quasi-Newton method with an optimal search along the gradients of (4). The search algorithm is a subspace trust region procedure based on the interior-reflective Newton method of Coleman & Li (1994, 1996). In each iteration of the search, a large linear system is approximately solved by using the method of preconditioned conjugate gradients. The algorithm is deemed convergent when the search step size and the norms of the search gradients are smaller than certain thresholds. This algorithm has been implemented in MATLAB and other commercial software packages. Although the search does not guarantee the global maximum, our experience shows that the algorithm works very well provided that the starting values are not far from the maximisers. We recommend using $\beta = 0$ and $\Lambda\{X_{ij}\} = n^{-1}$ $(i = 1, \ldots, n; j = 1, \ldots, n_i)$ as the starting values.

Denote the true values of β and Λ by β_0 and Λ_0 and their nonparametric maximum likelihood estimators by $\hat{\beta}_n$ and $\hat{\Lambda}_n$. In the Appendix, we show that $\hat{\beta}_n$ is strongly consistent and $\hat{\Lambda}_n(.)$ uniformly converges to $\Lambda_0(.)$ with probability one. In addition, the random element $n^{\frac{1}{2}}\{\hat{\Lambda}_n(.) - \Lambda_0(.), \hat{\beta}_n - \beta_0\}$ converges weakly to a zero-mean Gaussian process, and $\hat{\beta}_n$ is an asymptotically efficient estimator for β_0.

We wish to estimate the covariance function of $n^{\frac{1}{2}}\{\hat{\Lambda}_n(.) - \Lambda_0(.), \hat{\beta}_n - \beta_0\}$. It suffices to obtain a variance estimator for the linear functional

$$n^{\frac{1}{2}} \int_0^\tau w(t) d\{\hat{\Lambda}_n(t) - \Lambda_0(t)\} + n^{\frac{1}{2}} b^{\mathrm{T}}(\hat{\beta}_n - \beta_0),$$

where w is a function with bounded total variation in $[0, \tau]$ and b is a real vector. Since $\Lambda_0(.)$ is estimated at the parametric convergence rate, we may regard $\Lambda\{X_{ij}\}$ $(i = 1, \ldots, n; j = 1, \ldots, n_i)$ and β as the parameters in (4). By parametric likelihood theory, the asymptotic covariance matrix for estimators of these parameters can be estimated by the inverse of the observed information matrix $n\mathscr{I}_n$, which is the negative Hessian matrix of $l_n(\Lambda, \beta)$ evaluated at $\hat{\Lambda}_n\{X_{ij}\}$ $(i = 1, \ldots, n; j = 1, \ldots, n_i)$ and $\hat{\beta}_n$. Thus, the asymptotic variance for $n^{\frac{1}{2}} \int_0^\tau w(t) d\{\hat{\Lambda}_n(t) - \Lambda_0(t)\} + n^{\frac{1}{2}} b^{\mathrm{T}}(\hat{\beta}_n - \beta_0)$ is equal to the asymptotic variance of $n^{\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n_i} w(X_{ij}) \hat{\Lambda}_n\{X_{ij}\} + n^{\frac{1}{2}} b^{\mathrm{T}}(\hat{\beta}_n - \beta_0)$, which can be estimated by

$$\hat{V}_n = (W^{\mathrm{T}}, b^{\mathrm{T}}) \mathscr{I}_n^{-1} \begin{pmatrix} W \\ b \end{pmatrix},$$

where W is the vector of $w(X_{ij})$ $(i = 1, \ldots, n; j = 1, \ldots, n_i)$.

Let $F(\Lambda, \beta)$ be a Hadamard differentiable functional, which is estimated by $F(\hat{\Lambda}_n, \hat{\beta}_n)$. By the functional delta-method (Andersen et al., 1993, § II.8), $n^{\frac{1}{2}}\{F(\hat{\Lambda}_n, \hat{\beta}_n) - F(\Lambda_0, \beta_0)\}$ converges weakly to a zero-mean Gaussian process whose variance can be estimated by $\hat{V}_n$ with (W, b) redefined as the gradients of $F(\hat{\Lambda}_n, \hat{\beta}_n)$ with respect to $\Lambda\{X_{ij}\}$ $(i = 1, \ldots, n; j = 1, \ldots, n_i)$ and β.

We may also use the profile likelihood method (Murphy & van der Vaart, 2000) to estimate the covariance matrix of $\hat{\beta}_n$. The estimator is the negative inverse of the second-order numerical differences of the profile loglikelihood function at $\hat{\beta}_n$. This approach avoids inverting a potentially large matrix; however, it does not provide a variance estimator for $\hat{\Lambda}_n(.)$. Variance estimation for $\hat{\Lambda}_n(.)$ is important for transformation models because the prediction of event history may be more interesting than the estimation of regression parameters for nonproportional hazards models.

3. Simulation studies

We generated a covariate Z_1 from the Bernoulli distribution with success probability 0·5. Conditional on Z_1, we generated covariate Z_2 as $Z_1 + \varepsilon I(|\varepsilon| \leqslant 3)$, where ε follows the standard normal distribution. We simulated recurrent event times from the counting process with cumulative intensity

$$\Lambda_Z(t) = [\{1 + \Lambda(t)e^{\beta_1 Z_1 + \beta_2 Z_2}\}^{\rho} - 1]/\rho, \tag{5}$$

where $\Lambda(t) = t$ and $\rho = 0{\cdot}5$ or 2. We also simulated recurrent event times from the model

$$\Lambda_Z(t) = \log[1 + r\{\Lambda(t)e^{\beta_1 Z_1 + \beta_2 Z_2}\}]/r, \tag{6}$$

where $\Lambda(t) = t$ and $r = 0{\cdot}5$, 1 or 2. In both (5) and (6), we set $\beta_1 = -1$ and $\beta_2 = 0{\cdot}2$. We simulated the censoring time C from the Un(1·5, 5) distribution and set $\tau = 3$. We considered $n = 100$ or 200. In the first set of studies, the average numbers of events per subject are 1·41 for $\rho = 0{\cdot}5$ and 4·38 for $\rho = 2$; in the second set, the average numbers of events per subject are 0·76, 1·05 and 1·32 for $r = 0{\cdot}5$, 1 and 2, respectively.

Table 1 summarises the results of these studies based on 1000 replicates. The proposed estimators for β_0 and $\Lambda_0(.)$ are virtually unbiased, the variance estimators accurately reflect the true variances, and the confidence intervals achieve proper coverages. It took less than three hours on an IBM BladeCenter HS20 machine to complete all the simulation studies. No convergence problem was encountered in any of the 10 000 simulated datasets, although there is no theoretical guarantee of convergence to the global maximum.

To compare our approach with that of Chen et al. (2002), we conducted a series of simulation studies with survival data. The cumulative hazard function for the survival time is in the form of (6) with $\Lambda(t) = 3t$, $\beta_1 = -1$ and $\beta_2 = 0{\cdot}2$. The censoring time is exponential with a hazard rate chosen to yield a desired level of censoring under $\tau = 6$. The results for sample size 100 and 1000 replicates are shown in Table 2. The asymptotic approximations appear to work well for both approaches. The Chen et al. estimators are considerably less efficient than the proposed estimators especially when r is large and censoring is low. Our algorithm always converged, whereas that of Chen et al. failed to converge in about 2% of the simulated datasets.

Table 1. *Simulation studies for recurrent event data*

		$n = 100$				$n = 200$			
Model	Parameter	Bias	SE	SEE	CP	Bias	SE	SEE	CP
$\rho = 0{\cdot}5$	β_1	−0·015	0·245	0·250	0·955	0·007	0·173	0·175	0·952
	β_2	0·005	0·109	0·111	0·959	−0·007	0·075	0·077	0·950
	$\Lambda(\tau/4)$	−0·004	0·132	0·136	0·961	−0·005	0·095	0·095	0·958
	$\Lambda(\tau/2)$	−0·007	0·219	0·229	0·966	−0·004	0·161	0·162	0·949
	$\Lambda(\tau)$	0·000	0·407	0·423	0·954	−0·009	0·299	0·297	0·950
$\rho = 2{\cdot}0$	β_1	−0·006	0·087	0·086	0·955	−0·002	0·060	0·060	0·955
	β_2	0·000	0·033	0·032	0·945	0·000	0·023	0·022	0·952
	$\Lambda(\tau/4)$	−0·005	0·073	0·071	0·943	−0·002	0·050	0·050	0·947
	$\Lambda(\tau/2)$	−0·003	0·084	0·084	0·956	−0·001	0·058	0·059	0·955
	$\Lambda(\tau)$	−0·003	0·113	0·110	0·946	0·000	0·079	0·077	0·945
$r = 0{\cdot}5$	β_1	−0·007	0·268	0·277	0·961	−0·006	0·190	0·195	0·951
	β_2	0·007	0·123	0·125	0·961	−0·001	0·086	0·087	0·945
	$\Lambda(\tau/4)$	−0·009	0·141	0·143	0·961	−0·004	0·101	0·101	0·958
	$\Lambda(\tau/2)$	−0·001	0·249	0·253	0·955	−0·004	0·175	0·179	0·962
	$\Lambda(\tau)$	−0·014	0·471	0·497	0·966	0·010	0·350	0·351	0·950
$r = 1$	β_1	−0·008	0·355	0·355	0·955	0·003	0·253	0·249	0·943
	β_2	0·005	0·161	0·161	0·948	−0·002	0·112	0·112	0·955
	$\Lambda(\tau/4)$	−0·007	0·177	0·175	0·948	−0·001	0·125	0·124	0·947
	$\Lambda(\tau/2)$	−0·004	0·338	0·331	0·952	0·002	0·237	0·234	0·950
	$\Lambda(\tau)$	0·023	0·710	0·682	0·941	0·009	0·483	0·477	0·940
$r = 2$	β_1	0·006	0·467	0·479	0·961	0·003	0·325	0·335	0·949
	β_2	−0·001	0·227	0·217	0·952	−0·008	0·151	0·151	0·946
	$\Lambda(\tau/4)$	−0·011	0·230	0·229	0·952	−0·002	0·171	0·163	0·945
	$\Lambda(\tau/2)$	0·005	0·459	0·462	0·955	0·007	0·341	0·326	0·947
	$\Lambda(\tau)$	0·044	0·975	0·977	0·950	0·026	0·707	0·683	0·951

SE, standard error; SEE, mean of standard error estimator; CP, coverage probability of 95% confidence interval.

Table 2. *Simulation studies for survival data*

			Proposed estimator				Chen et al. estimator			
Censoring	Model	Parameter	Bias	SE	SEE	CP	Bias	SE	SEE	CP
25%	$r = 0{\cdot}5$	β_1	−0·026	0·378	0·358	0·937	−0·035	0·393	0·366	0·947
		β_2	0·005	0·165	0·159	0·949	0·006	0·172	0·164	0·940
	$r = 1$	β_1	−0·022	0·440	0·420	0·941	−0·032	0·482	0·446	0·941
		β_2	0·005	0·193	0·187	0·956	0·007	0·210	0·203	0·949
	$r = 2$	β_1	−0·023	0·545	0·523	0·944	−0·029	0·655	0·602	0·949
		β_2	0·005	0·242	0·234	0·939	0·005	0·286	0·279	0·943
50%	$r = 0{\cdot}5$	β_1	−0·029	0·437	0·413	0·951	−0·051	0·444	0·410	0·945
		β_2	0·006	0·187	0·183	0·951	0·006	0·191	0·184	0·947
	$r = 1$	β_1	−0·031	0·488	0·463	0·944	−0·054	0·512	0·469	0·940
		β_2	0·007	0·213	0·207	0·955	0·008	0·225	0·214	0·948
	$r = 2$	β_1	−0·025	0·579	0·555	0·942	−0·045	0·644	0·588	0·938
		β_2	0·006	0·257	0·249	0·956	0·009	0·284	0·274	0·949

SE, standard error; SEE, mean of standard error estimator; CP, coverage probability of 95% confidence interval.

4. Examples

We first consider survival data from the Veterans' Administration lung cancer trial (Prentice, 1973). The subset of data for the 97 patients without prior therapy has been analysed by many authors, including Bennett (1983), Pettitt (1984), Cheng et al. (1995), Murphy et al. (1997) and Chen et al. (2002). Chen et al. related the survival time to the performance status and tumour type through linear transformation models with $G(x) = \log(1 + rx)/r$, where $r = 0$, 1, 1·5 and 2. For comparison, we fitted the same models and display the results in Table 3. These results differ appreciably from those of Chen et al. (2002). For $r = 0$, our numbers agree with the standard software output. For $r = 1$, our results are similar to those of Murphy et al. (1997).

Table 3. *Estimates of regression parameters for the Veteran's Administration lung cancer data, with standard error estimates shown in parentheses*

	$r = 0$	$r = 1$	$r = 1{\cdot}5$	$r = 2$
Performance status	−0·024 (0·006)	−0·053 (0·010)	−0·063 (0·012)	−0·072 (0·014)
Adeno vs large tumour	0·851 (0·348)	1·314 (0·554)	1·497 (0·636)	1·679 (0·712)
Small vs large tumour	0·547 (0·321)	1·383 (0·524)	1·605 (0·596)	1·814 (0·661)
Squam vs large tumour	−0·215 (0·347)	−0·181 (0·588)	−0·075 (0·675)	0·045 (0·749)

To determine which model best fits the data, we plot in Fig. 2 the observed values of the loglikelihood functions for the Box–Cox and logarithmic transformations. The likelihood is maximised at $r = 0{\cdot}83$. Since the likelihood at $r = 1$ is only slightly smaller, one would choose $r = 1$ to obtain the familiar proportional odds model. The prediction of the subject-specific survival experience under the proportional odds model is illustrated in Fig. 3.

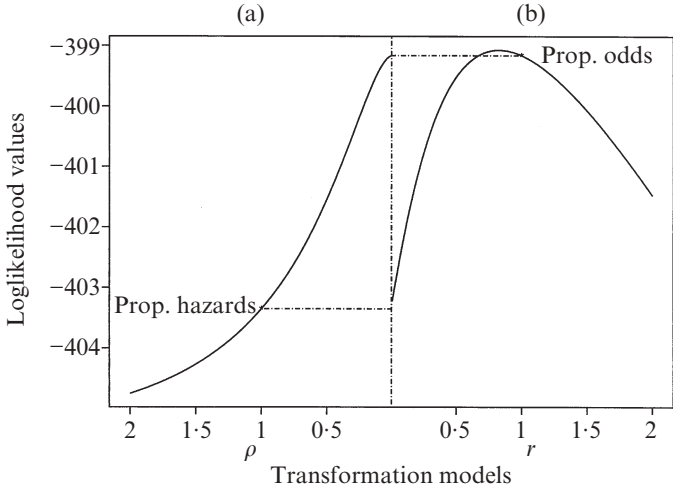

Fig. 2. The observed values of the loglikelihood functions for the lung cancer data: (a) pertains to the Box–Cox transformations $G(x) = \{(1 + x)^{\rho} - 1\}/\rho$; (b) pertains to the logarithmic transformations $G(x) = \log(1 + rx)/r$.

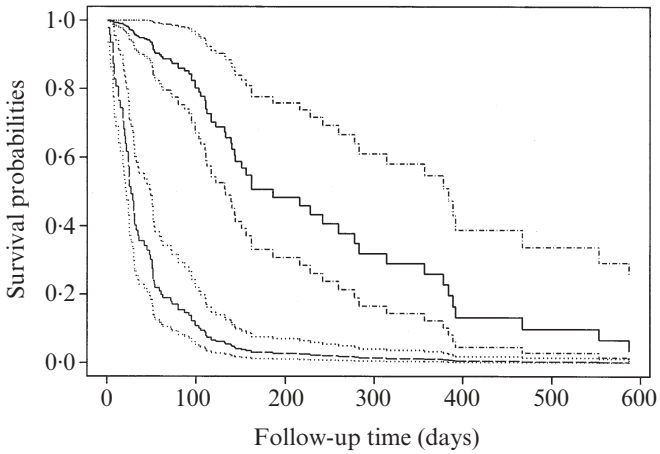

Fig. 3. Estimated survival curves for the lung cancer patients: the upper three curves pertain to the point estimate and 95% confidence limits for a patient with a large tumour and performance status of 80, and the lower three curves to those of a patient with a small tumour and performance status of 40.

As a second example, we consider the recurrent bladder tumour data from another Veterans' Administration study (Byar, 1980), which has been examined extensively in the literature of multivariate failure time data, including Wei et al. (1989) and Therneau & Grambsch (2000). The data contain information about tumour recurrence times, in months, for 86 patients who were on the placebo or thiotepa. Figure 4 displays the log-likelihood functions for the two classes of transformations, while Table 4 shows the estimates of regression parameters for selected transformations. The results under $\rho = 0$ are identical to those obtained from standard software. Figure 4 seems to suggest a model with a large value of ρ. It would suffice to choose $\rho = 2$ or even $\rho = 1$.

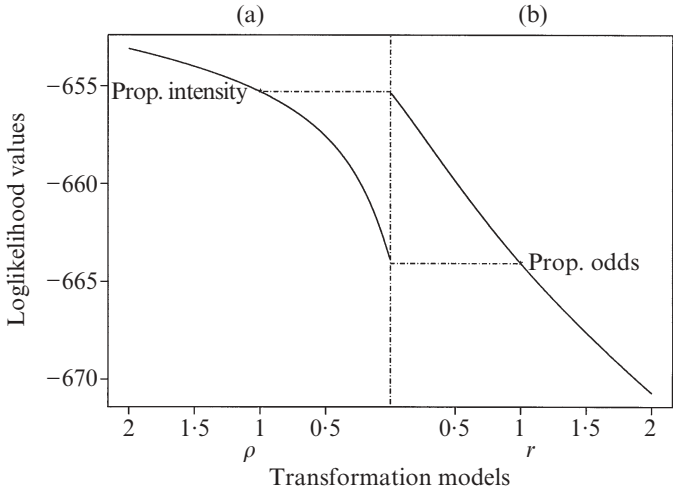

Fig. 4. The observed values of the loglikelihood functions for the bladder tumour data: (a) pertains to the Box–Cox transformations $G(x) = \{(1 + x)^{\rho} - 1\}/\rho$; (b) pertains to the logarithmic transformations $G(x) = \log(1 + rx)/r$.

Table 4. *Estimates of regression parameters for the bladder tumour data, with standard error estimates shown in parentheses*

	$\rho = 2$	$\rho = 1$	$\rho = 0{\cdot}5$	$r = 1$
Treatment	−0·369 (0·136)	−0·524 (0·187)	−0·701 (0·244)	−0·974 (0·358)
No. tumours	0·141 (0·030)	0·201 (0·044)	0·269 (0·061)	0·352 (0·101)
Tumour size	−0·035 (0·048)	−0·040 (0·065)	−0·041 (0·084)	−0·013 (0·123)

Suppose that we are interested in the conditional survival function of the second recurrence time X_2 given the first recurrence time x_1 for subjects with covariate values z. This probability function can be estimated by $\exp[G\{\hat{\Lambda}_n(x_1)e^{\hat{\beta}_n^{\mathrm{T}}z}\} - G\{\hat{\Lambda}_n(t)e^{\hat{\beta}_n^{\mathrm{T}}z}\}]$ for any $t > x_1$. Figure 5 shows the estimated curves for two sets of covariate values under $x_1 = 20$ and $\rho = 2$.

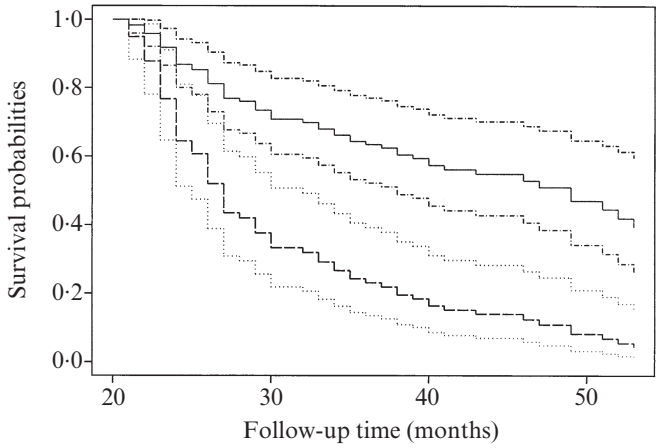

Fig. 5. Estimated conditional survival curves for the bladder tumour patients: the upper three curves correspond to the point estimate and 95% confidence limits for a thiotepa patient with one initial tumour, and the lower three curves to those of a placebo patient with four initial tumours.

5. Remarks

In the special case of linear transformation models for survival data, there exist a number of estimators based on ad hoc estimating equations. We may construct martingale-based estimating functions similar to those of Chen et al. (2002) for the general class of models given in (2). The estimators developed in the present paper have the advantage of being asymptotically efficient. Another attraction of this approach is that likelihood-based model selection methods, such as the Akaike (1985) information criterion, AIC, can be used. Figures 2 and 4 are special examples of AIC.

For recurrent events, intensity models impose a Poisson structure. This assumption may be relaxed by characterising the dependence of recurrent event times through appropriate time-varying covariates. Another approach, which we are currently pursuing, is to incorporate subject-specific random effects into the model. We are also exploring methods for handling clustered failure times, interval censoring and missing/mismeasured covariates.

For recurrent events with time-invariate covariates, Lin et al. (2001) studied the following class of transformation models for the mean frequency functions:

$$E\{N^*(t)|Z\} = g\{\mu_0(t)e^{\beta^{\mathrm{T}}Z}\},$$

where $g(.)$ is a known transformation, and $\mu_0(t)$ is an arbitrary increasing function. The estimators they proposed are inefficient. An attractive feature of this modelling approach is that the dependence structure of recurrent event times is unspecified. However, such models cannot be used for the kind of prediction shown in Fig. 5.

ACKNOWLEDGEMENT

The authors are grateful to two referees for helpful comments and to Dr Zhezhen Jin for providing a program for the Chen et al. estimators. This research was supported by the U.S. National Institutes of Health.

APPENDIX

Asymptotic properties

Let $\|.\|_{l^\infty[0,\tau]}$ denote the supremum norm in $[0, \tau]$, and $\|w\|_{BV[0,\tau]}$ the total variation of $w(t)$ in $[0, \tau]$. Also, define $\mathscr{Q} = \{w(t): \|w\|_{BV[0,\tau]} \leqslant 1\}$. Then $\hat{\Lambda}_n(t)$ can be regarded as a bounded linear functional in $l^\infty(\mathscr{Q})$, and $\{\hat{\Lambda}_n(.) - \Lambda_0(.), \hat{\beta}_n - \beta_0\}$ a random element in the metric space $l^\infty(\mathscr{Q}) \times \mathscr{R}^p$, where p is the dimension of β_0. We claim the following results: $\|\hat{\Lambda}_n(t) - \Lambda_0(t)\|_{l^\infty[0,\tau]} \to 0$ and $|\hat{\beta}_n - \beta_0| \to 0$ almost surely; the random element $n^{1/2}\{\hat{\Lambda}_n(.) - \Lambda_0(.), \hat{\beta}_n - \beta_0\}$ converges weakly to a zero-mean Gaussian process in the metric space $l^\infty(\mathscr{Q}) \times \mathscr{R}^p$; and the limiting variance of $n^{1/2}(\hat{\beta}_n - \beta_0)$ attains the semiparametric efficiency bound (Bickel et al., 1993, Ch. 3).

We shall establish the claims under the following conditions, although the results are expected to hold generally.

Condition 1. The function $\Lambda_0(t)$ is strictly increasing and continuously differentiable, and β_0 lies in the interior of a compact set $\mathscr{C}$.

Condition 2. With probability one, $Z(.)$ has bounded total variation in $[0, \tau]$. In addition, if there exists a vector γ and a deterministic function $\gamma_0(t)$ such that $\gamma_0(t) + \gamma^{\mathrm{T}} Z(t) = 0$ with probability one, then $\gamma = 0$ and $\gamma_0(t) = 0$.

Condition 3. With probability one, there exists a positive constant δ such that $\mathrm{pr}(C \geqslant \tau|Z) > \delta$ and $\mathrm{pr}(\bar{Y}^*(\tau) = 1|Z) > \delta$, where $\bar{Y}^*(\tau) = 1$ means that $Y^*(t) = 1$ for all $t \in [0, \tau]$.

Condition 4. For any positive c_0, $\limsup_{x\to\infty} \{G(c_0 x)\}^{-1} \log\{x \sup_{y\leqslant x} G'(y)\} = 0$. This condition is satisfied by $G(x) = \{(1 + x)^\rho - 1\}/\rho$ with $\rho > 0$.

Our proofs follow essentially the same steps as Murphy (1994, 1995), Parner (1998) and Scharfstein et al. (1998), although the technical details are different. We outline below our arguments. The complete proofs are given in a technical report posted on our website.

Consistency. The proof consists of three steps: first we show that the nonparametric maximum likelihood estimators exist or that the jump sizes of $\hat{\Lambda}_n$ are finite; next we show that $\hat{\Lambda}_n$ is bounded almost surely so that, along a subsequence, $\hat{\Lambda}_n \to \Lambda^*$ weakly and $\hat{\beta}_n \to \beta^*$; finally we show that $\Lambda^* = \Lambda_0$ and $\beta^* = \beta_0$.

Step 1. By Condition 2, $\sup_{\beta \in \mathscr{C}, t \in [0,\tau]} |\beta^{\mathrm{T}} Z(t)| \leqslant M$ almost surely, where M is a constant. Thus, the ith term in (4) is bounded above by

$$n_i G\{\Lambda(\tau \wedge C_i) e^M\} \left[\frac{\log\{\int_0^\tau Y_i(t) d\Lambda(t) e^M \sup_{y \leqslant \int_0^\tau Y_i(t) d\Lambda(t) e^M} G'(y)\}}{G\{\int_0^\tau Y_i(t) d\Lambda(t) e^{-M}\}} - n_i^{-1} \right].$$

Under Condition 4, the above quantity diverges to $-\infty$ if $\Lambda\{X_{ij}\}$ is infinite for some X_{ij}.

Step 2. Since $l_n(\Lambda, \beta)$ is maximised at $(\hat{\Lambda}_n, \hat{\beta}_n)$,

$$n^{-1}\{l_n(\xi_n \bar{\Lambda}_n, \hat{\beta}_n) - l_n(\bar{\Lambda}_n, \hat{\beta}_n)\} \geqslant 0, \tag{A1}$$

where $\xi_n = \hat{\Lambda}_n(\tau)$ and $\bar{\Lambda}_n = \hat{\Lambda}_n / \xi_n$. Suppose that $\xi_n \to \infty$ for some subsequence. Algebraic manipulations of (A1), together with the boundedness of $\bar{\Lambda}_n$, yield

$$n^{-1} \sum_{i=1}^n \int_0^{\tau \wedge C_i} dN_i(t) \log \xi_n \sup_{y \leqslant \xi_n e^M} G'(y) - n^{-1} \sum_{i=1}^n I\{\bar{Y}_i^*(\tau) = 1, C_i \geqslant \tau\} G(e^{-M} \xi_n) \geqslant O_p(1).$$

Condition 4 implies that $\log \xi_n \sup_{y \leqslant \xi_n e^M} G'(y) \leqslant \varepsilon G(\xi_n e^{-M})$ for any ε when n is sufficiently large. Thus,

$$\left[n^{-1} \varepsilon \sum_{i=1}^n N_i(\tau) - n^{-1} \sum_{i=1}^n I\{\bar{Y}_i^*(\tau) = 1, C_i \geqslant \tau\} \right] G(\xi_n e^{-M}) > -\infty.$$

The left-hand side diverges to $-\infty$ if we choose an ε such that $\varepsilon E\{N(\tau)\} \leqslant \mathrm{pr}\{\bar{Y}^*(\tau) = 1, C \geqslant \tau\}/2$. This is a contradiction. Thus, $\sup_n \hat{\Lambda}_n(\tau) < \infty$ almost surely. It then follows from Helly's selection theorem that there exists a convergent subsequence such that $\hat{\Lambda}_n \to \Lambda^*$ and $\hat{\beta}_n \to \beta^*$.

Step 3. By setting the derivatives of $l_n(\Lambda, \beta)$ with respect to the $\Lambda\{X_{ij}\}$ to zero, we obtain

$$\hat{\Lambda}_n(t) = n^{-1} \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{|\phi_n(s; \hat{\Lambda}_n, \hat{\beta}_n)|}, \tag{A2}$$

where

$$\begin{aligned} \phi_n(s; \hat{\Lambda}_n, \hat{\beta}_n) = {} & n^{-1} \sum_{k=1}^n G' \left\{ \int_0^\tau Y_k(t) e^{\hat{\beta}_n^{\mathrm{T}} Z_k(t)} d\hat{\Lambda}_n(t) \right\} e^{\hat{\beta}_n^{\mathrm{T}} Z_k(s)} Y_k(s) \\ & - n^{-1} \sum_{k=1}^n \int_0^\tau \frac{I(t \geqslant s) Y_k(s) e^{\hat{\beta}^{\mathrm{T}} Z_k(s)} G''\{\int_0^t Y_k(u) e^{\hat{\beta}^{\mathrm{T}} Z_k(u)} d\hat{\Lambda}_n(u)\}}{G'\{\int_0^t Y_k(u) e^{\hat{\beta}^{\mathrm{T}} Z_k(u)} d\hat{\Lambda}_n(u)\}} dN_k(t). \end{aligned}$$

By the Glivenko–Cantelli theorem, $\phi_n(t; \hat{\Lambda}_n, \hat{\beta}_n)$ converges uniformly to a continuously differentiable function $\phi^*(t; \Lambda^*, \beta^*)$. The quantity $\min_{t \in [0,\tau]} |\phi^*(t; \Lambda^*, \beta^*)|$ must be strictly positive; otherwise, the continuous differentiability of ϕ^* implies that

$$E\left[\int_0^\tau \frac{dN(t)}{|\phi^*(t; \Lambda^*, \beta^*)|} \right] = \infty,$$

which contradicts the boundedness of the limit of (A2). Thus, $|\phi_n(.; \hat{\Lambda}_n, \hat{\beta}_n)|$ is strictly positive for large n.

Define

$$\tilde{\Lambda}_n(t) = n^{-1} \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{|\phi_n(s; \Lambda_0, \beta_0)|}. \tag{A3}$$

By the Glivenko–Cantelli theorem, $\tilde{\Lambda}_n$ converges to Λ_0 uniformly. It follows from (A2), (A3) and the strict positivity of $|\phi_n|$ that $\hat{\Lambda}_n(t)$ is absolutely continuous with respect to $\tilde{\Lambda}_n(t)$ and that $d\hat{\Lambda}_n / d\tilde{\Lambda}_n$ converges to a bounded measurable function. Clearly, $n^{-1}\{l_n(\hat{\Lambda}_n, \hat{\beta}_n) - l_n(\tilde{\Lambda}_n, \beta_0)\} \geqslant 0$. By taking

the limits on both sides, we conclude that the Kullback–Leibler information between the density indexed by (Λ^*, β^*) and the true density is negative. Thus, with probability one,

$$\int_0^\tau \log\left[Y(t)\lambda^*(t)e^{\beta^{*\mathrm{T}}Z(t)} G'\left\{ \int_0^t Y(s)e^{\beta^{*\mathrm{T}}Z(s)} d\Lambda^*(s) \right\} \right] dN(t) - G\left\{ \int_0^\tau Y(t)e^{\beta^{*\mathrm{T}}Z(t)} d\Lambda(t) \right\}$$
$$= \int_0^\tau \log\left[Y(t)\lambda_0(t)e^{\beta_0^{\mathrm{T}}Z(t)} G'\left\{ \int_0^t Y(s)e^{\beta_0^{\mathrm{T}}Z(s)} d\Lambda_0(s) \right\} \right] dN(t) - G\left\{ \int_0^\tau Y(t)e^{\beta_0^{\mathrm{T}}Z(t)} d\Lambda_0(t) \right\}.$$

This equality holds when $\bar{Y}^*(\tau) = 1$, $N^*(\tau) = 0$ and $C \geqslant \tau$, and also holds when $\bar{Y}^*(\tau) = 1$, $N^*(\tau -) = 0$, $N^*(\tau) = 1$ and $C \geqslant \tau$. By taking the difference between the equalities in these two cases and applying Condition 2, we obtain $\beta^* = \beta_0$ and $\Lambda^* = \Lambda_0$. The convergence of $\hat{\Lambda}_n(t) \to \Lambda_0(t)$ can be strengthened to uniform convergence in $t \in [0, \tau]$ by the continuity of Λ_0.

Weak convergence. The proof entails the verification of the four conditions in Theorem 3.3.1 of van der Vaart & Wellner (1996). The random maps Ψ_n and Ψ in the theorem are defined as follows: for any $w(t) \in \mathscr{Q}$ and $b \in \mathscr{R}^p$ with $|b| \leqslant 1$,

$$\Psi_n(\Lambda, \beta)[w, b] = \mathscr{P}_n\left\{ l_\Lambda\left[\int w(t)d\Lambda \right] + l_\beta^{\mathrm{T}} b \right\}, \quad \Psi(\Lambda, \beta)[w, b] = \mathscr{P}\left\{ l_\Lambda\left[\int w(t)d\Lambda \right] + l_\beta^{\mathrm{T}} b \right\},$$

where $\mathscr{P}_n$ is the empirical measure, $\mathscr{P}$ is the probability measure, and $l_\Lambda[\int w(t)d\Lambda] + l_\beta^{\mathrm{T}} b$ is the score function along the path $(\Lambda + \varepsilon \int w(t)d\Lambda, \beta + \varepsilon b)$. We can show that the score functions for Λ and β are $\mathscr{P}$-Donsker so that the first two conditions of the theorem hold. Clearly, $\Psi(\Lambda_0, \beta_0) = \Psi_n(\hat{\Lambda}_n, \hat{\beta}_n) = 0$. Thus, it remains to show that the Fréchet derivative of Ψ at (Λ_0, β_0), denoted by $\dot{\Psi}$, is invertible. By direct calculations,

$$\dot{\Psi}(\Lambda - \Lambda_0, \beta - \beta_0)[w, b] = \int Q_1(w, b)d(\Lambda - \Lambda_0) + Q_2(w, b)^{\mathrm{T}}(\beta - \beta_0),$$

where (Q_1, Q_2) is a linear operator mapping $\mathscr{Q} \times \mathscr{R}^p$ into $BV[0, \tau] \times \mathscr{R}^p$ and $BV[0, \tau]$ is the space of functions with bounded total variation. In addition, (Q_1, Q_2) is the sum of an invertible operator and a compact operator. Therefore, $\dot{\Psi}$ is invertible if (Q_1, Q_2) is one-to-one. By the definition of $\dot{\Psi}$, $(Q_1[w, b], Q_2[w, b]) = 0$ implies that the score function along the path $(\Lambda_0 + \varepsilon \int w(t)d\Lambda_0, \beta_0 + \varepsilon b)$ is zero or equivalently that the information along this submodel is zero. Using Condition 2, we can further show that $b = 0$ and $w = 0$. Hence, $n^{1/2}(\hat{\Lambda}_n - \Lambda_0, \hat{\beta}_n - \beta_0)$ converges weakly to a zero-mean Gaussian process in $l^\infty(\mathscr{Q}) \times \mathscr{R}^p$.

Asymptotic efficiency. The above proof implies that $\hat{\beta}_n$ is an asymptotically linear estimator for β_0 and its influence function lies on the linear space spanned by the score functions. Thus, it follows from Proposition 1 of Bickel et al. (1993, p. 65) that $\hat{\beta}_n$ is asymptotically efficient in the semiparametric sense.

Consistency of variance estimators. This can be justified along the lines of Parner (1998). The key is to show that the linear operator constructed from the negative Hessian matrix of the loglikelihood function approximates the information operator.

Relaxing Condition 4. Condition 4 rules out the logarithmic transformations; however, this condition is only used in the first two steps of the consistency proof. Thus, if we can verify those two steps for $G(x) = \varrho \log(1 + rx)$, where ϱ and r are positive constants, then all the asymptotic results will hold for such transformations. The first step can be directly checked by using the explicit form of $G(x)$. To verify the second step, it suffices to show that $\hat{\Lambda}_n(\tau) < \infty$. By equation (A2) and the fact that $G'' < 0$,

$$\frac{1}{n\hat{\Lambda}_n\{X_{ij}\}} \geqslant n^{-1}\varrho r \sum_{k=1}^n \int_0^\tau \frac{I(t \geqslant X_{ij})Y_k(X_{ij})e^{-M}}{1 + re^M \int_0^t Y_k(s)d\hat{\Lambda}_n(s)} dN_k(t).$$

Thus, it follows from (A1) that

$$-n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\log\left\{n^{-1}\sum_{k=1}^{n}\int_0^{\tau}\frac{I(t\geqslant s)Y_k(s)e^{-M}}{1+re^{M}\int_0^{t}Y_k(u)d\hat{\Lambda}_n(u)}dN_k(t)\right\}dN_i(s)$$
$$-n^{-1}\varrho\sum_{i=1}^{n}\log\left\{1+re^{-M}\int_0^{\tau}Y_i(s)d\hat{\Lambda}_n(s)\right\}>-\infty. \tag{A4}$$

For simplicity, assume that $Y(.)$ is nonincreasing. We partition the time axis at

$$s_0=\tau>s_1>s_2>\ldots>s_{\varrho}=0.$$

Then the left-hand side of (A4) can be bounded by

$$\begin{aligned}
&-(2n)^{-1}\varrho\sum_{i=1}^{n}I\{Y_i(s_0)=1\}\log\{1+re^{-M}\hat{\Lambda}_n(\tau)\}\\
&\quad+\left[n^{-1}\sum_{i=1}^{n}I\{Y_i(s_0)=0,\ Y_i(s_1)=1\}N_i(\tau)\log\{1+re^{M}\hat{\Lambda}_n(\tau)\}\right.\\
&\qquad\left.-(2n)^{-1}\varrho\sum_{i=1}^{n}I\{Y_i(s_0=1)\}\log\{1+re^{-M}\hat{\Lambda}_n(\tau)\}\right]\\
&\quad+\sum_{q=1}^{\varrho-1}\left[n^{-1}\sum_{i=1}^{n}I\{Y_i(s_q)=0,\ Y_i(s_{q+1})=1\}N_i(\tau)\log\{1+re^{M}\hat{\Lambda}_n(s_q)\}\right.\\
&\qquad\left.-n^{-1}\varrho\sum_{i=1}^{n}I\{Y_i(s_{q-1})=0,\ Y_i(s_q)=1\}\log\{1+re^{-M}\hat{\Lambda}_n(s_q)\}\right]+O_p(1).
\end{aligned}\tag{A5}$$

With the choice of $(s_0, s_1, \ldots, s_{\varrho})$ such that

$$n^{-1}\sum_{i=1}^{n}I\{Y_i(s_0)=0,\ Y_i(s_1)=1\}N_i(\tau)<(2n)^{-1}\varrho\sum_{i=1}^{n}I\{Y_i(s_0=1)\},$$
$$n^{-1}\sum_{i=1}^{n}I\{Y_i(s_q)=0,\ Y_i(s_{q+1})=1\}N_i(\tau)<n^{-1}\varrho\sum_{i=1}^{n}I\{Y_i(s_{q-1})=0,\ Y_i(s_q)=1\},$$

the first term in (A5) diverges to $-\infty$ when $\hat{\Lambda}_n(\tau)\to\infty$ while the second and third terms do not diverge. Thus, the left-hand side of (A4) goes to $-\infty$, which is a contradiction. Such a sequence can be constructed along the lines of Murphy (1994).

It can be shown that the desired asymptotic results hold if Condition 4 is replaced by the following condition: for any sequence $0<x_1<\ldots<x_m\leqslant y$,

$$\prod_{l=1}^{m}\{(1+x_l)G'(x_l)\}\exp\{-G(y)\}\leqslant\mu_0^{m}(1+y)^{-\alpha_0},$$

where μ_0 and α_0 are positive constants. Both the class of Box–Cox transformations and the class of logarithmic transformations satisfy the above inequality.

References

Akaike, H. (1985). Prediction and entropy. In *A Celebration of Statistics*, Ed. A. C. Atkinson and S. E. Fienberg, pp. 1–24. New York: Springer-Verlag.

Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100–20.

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statist. Med.* **2**, 273–7.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. & WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.

BYAR, D. P. (1980). The Veterans Administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology*, Ed. M. Pavone-Macaluso, P. H. Smith and F. Edsmyn, pp. 363–70. New York: Plenum.

CHEN, K., JIN, Z. & YING, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89**, 659–68.

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.

COLEMAN, T. F. & LI, Y. (1994). On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds. *Math. Prog.* **67**, 189–224.

COLEMAN, T. F. & LI, Y. (1996). An interior, trust region approach for nonlinear minimization subject to bounds. *SIAM J. Optimiz.* **6**, 418–45.

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). Partial likelihood in transformation models with censored data. *Scand. J. Statist.* **18**, 1–23.

LIN, D. Y., WEI, L. J. & YING, Z. (2001). Semiparametric transformation models for point processes. *J. Am. Statist. Assoc.* **96**, 620–8.

MURPHY, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712–31.

MURPHY, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182–98.

MURPHY, S. A. & VAN DER VAART, A. W. (2000). On the profile likelihood. *J. Am. Statist. Assoc.* **95**, 449–65.

MURPHY, S. A., ROSSINI, A. J. & VAN DER VAART, A. W. (1997). Maximal likelihood estimation in the proportional odds model. *J. Am. Statist. Assoc.* **92**, 968–76.

PARNER, E. (1998). Asymptotic theory for the correlated gamma-frailty models. *Ann. Statist.* **26**, 183–214.

PETTITT, A. N. (1984). Proportional odds models 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**, 278–88.

SCHARFSTEIN, D. O., TSIATIS, A. A. & GILBERT, P. B. (1998). Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Anal.* **4**, 355–91.

THERNEAU, T. M. & GRAMBSCH, P. M. (2000). *Modeling Survival Data: Extending the Cox Model.* New York: Springer.

VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* New York: Springer-Varlag.

WEI, L. J., LIN, D. Y. & WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Am. Statist. Assoc.* **84**, 1065–73.

[*Received July* 2004. *Revised December* 2005]