Biometrika (1995), **82**, 4, *pp.* 835–45
Printed in Great Britain

Analysis of transformation models with censored data

BY S. C. CHENG
Department of Biomathematics, M. D. Anderson Cancer Center, University of Texas, 1515 Holcombe Boulevard, Box 237, Houston, Texas 77030, U.S.A.

L. J. WEI
Department of Biostatistics, Harvard University, 677 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.

AND Z. YING
Department of Statistics, Rutgers University, New Brunswick, New Jersey 08903, U.S.A.

SUMMARY

In this paper we consider a class of semi-parametric transformation models, under which an unknown transformation of the survival time is linearly related to the covariates with various completely specified error distributions. This class of regression models includes the proportional hazards and proportional odds models. Inference procedures derived from a class of generalised estimating equations are proposed to examine the covariate effects with censored observations. Numerical studies are conducted to investigate the properties of our proposals for practical sample sizes. These transformation models, coupled with the new simple inference procedures, provide many useful alternatives to the Cox regression model in survival analysis.

Some key words: Generalised estimating equation; Martingale; Proportional hazards model; Proportional odds model; *U*-statistic.

1. INTRODUCTION

Let T be the 'failure time', the response variable, and Z a corresponding covariate vector. Suppose that we are interested in making inferences about the effect of Z on the response variable T. If there are censored observations in the data, one usually uses the Cox proportional hazards model to examine the covariate effect (Cox, 1972, 1975). The Cox model is semi-parametric, and its large sample inference properties have been demonstrated using martingale theory (Andersen & Gill, 1982). Moreover, practitioners have easy access to statistical software for this model. Therefore, there is a temptation to use the proportional hazards model to analyse failure time observations, even when the model does not fit the data well.

Let $S_Z(.)$ be the survival function of T given Z. The Cox model can be written as

$$\log[-\log\{S_Z(t)\}] = h(t) + Z^{\mathrm{T}}\beta, \qquad (1\cdot1)$$

where $h(t)$ is a completely unspecified strictly increasing function, which maps the positive half-line onto the whole real line, and β is a $p \times 1$ vector of unknown regression coefficients. Inference about β in $(1\cdot1)$ can be based on the partial likelihood function. An alternative is the proportional odds model:

$$-\operatorname{logit}\{S_Z(t)\} = h(t) + Z^{\mathrm{T}}\beta, \tag{1·2}$$

where $\operatorname{logit}(x) = \log\{x/(1-x)\}$ (Pettitt, 1982; Bennett, 1983). Although this model is appealing to practitioners, there is no theoretical justification for the large sample properties of inference procedures for β in the literature, except for the simple two-sample case (Bickel, 1986; Dabrowska & Doksum, 1988a).

A natural generalisation of (1·1) and (1·2) is

$$g\{S_Z(t)\} = h(t) + Z^{\mathrm{T}}\beta, \tag{1·3}$$

where $g(.)$ is a known decreasing function. The generalised odds-rate model studied by Dabrowska & Doksum (1988a) for the two-sample problem belongs to (1·3). It is easy to see that (1·3) is equivalent to the linear transformation model:

$$h(T) = -Z^{\mathrm{T}}\beta + \varepsilon, \tag{1·4}$$

where ε is a random error with distribution function $F = 1 - g^{-1}$. If F is the extreme value distribution $F(s) = 1 - \exp\{-\exp(s)\}$, (1·4) is the proportional hazards model, while if F is the standard logistic distribution, (1·4) is the proportional odds model. The parametric version of this transformation model, with h specified up to a finite-dimensional parameter vector, has been discussed extensively by Box & Cox (1964). For the case of h completely unspecified, methods for analysing failure time data with (1·4) have been proposed, for example, by Cuzick (1988) and P. J. Bickel and Ritov in an unpublished paper for the noncensored case, 'Local asymptotic normality of ranks and covariates in transformation models'. Cuzick (1988) suggested a way to extend his estimator to the censored case. Except for the proportional hazards model (1·1), however, the existing estimation procedures for β in (1·4) are either too complicated for practical use or have no rigorous justification of their large sample properties (Clayton & Cuzick, 1986; Dabrowska & Doksum, 1988b).

In this paper, we propose a class of simple estimating functions for β in the linear transformation model (1·4) with possibly censored observations. Under rather mild conditions, we show that the resulting estimators for β are consistent and asymptotically normal. Numerical comparisons are also made with Cox's estimator for the proportional hazards model and an estimator proposed by Dabrowska & Doksum (1988a) for the two-sample proportional odds model. With this simple new estimation procedure, model (1·4) provides useful alternatives to the Cox regression model in survival analysis.

2. Estimation for the linear transformation model

Let T_i be the failure time for the ith patient $(i = 1, \ldots, n)$. For T_i, one can only observe a bivariate vector (X_i, Δ_i), where $X_i = \min(T_i, C_i)$ and $\Delta_i = 1$ if $T_i = X_i$ and $\Delta_i = 0$ otherwise. The censoring variable C_i is assumed to be independent of T_i. Let Z_i, a $p \times 1$ vector, be the corresponding covariate vector for the ith patient. Furthermore, we assume that the 'survival' function $G(.)$ of C_i does not depend on Z_i. This assumption can easily be relaxed for the case when the covariate vector Z has a finite number of possible values.

Under the linear transformation model (1·4), h is a strictly increasing function. The rank configuration of $\{h(T_i), i = 1, \ldots, n\}$ is exactly the same as that of $\{T_i\}$. Therefore, it seems natural to use the marginal likelihood of ranks to make inferences about β. The corresponding maximum likelihood estimate and its variance, however, are difficult to obtain numerically. Moreover, the large sample properties of this estimator are not available for the censored case.

Consider the dichotomous variables $\{I(T_i \geqslant T_j), i \neq j = 1, \ldots, n\}$, where $I(.)$ denotes an indicator function. Then,

$$E\{I(T_i \geqslant T_j) | Z_i, Z_j\} = \mathrm{pr}\{h(T_i) \geqslant h(T_j) | Z_i, Z_j\},$$

which is

$$\xi(Z_{ij}^{\mathrm{T}}\beta_0) = \mathrm{pr}(\varepsilon_i - \varepsilon_j \geqslant Z_{ij}^{\mathrm{T}}\beta_0),$$

where β_0 is the true value for β, $Z_{ij} = Z_i - Z_j$,

$$\xi(s) = \int_{-\infty}^{\infty} \{1 - F(t+s)\}\, dF(t)$$

and F is the completely specified distribution function of ε. Although the dichotomous variables $\{I(T_i \geqslant T_j), i, j = 1, \ldots, n\}$ are dependent, one may make inferences about β_0 based on generalised estimating equations (Liang & Zeger, 1986). For example, if we assume that the dichotomous variables are independent, the resulting estimating function is

$$\tilde{U}(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\beta) Z_{ij} \{I(T_i \geqslant T_j) - \xi(Z_{ij}^{\mathrm{T}}\beta)\}, \qquad (2{\cdot}1)$$

where $w(.)$ is a weight function. Although those dichotomous variables $\{I(T_i \geqslant T_j)\}$ are dependent, $E\{\tilde{U}(\beta_0)\} = 0$. This suggests that a solution to $\tilde{U}(\beta) = 0$ is a reasonable estimator for β_0. To mimic the usual linear regression technique, one may set $w(.) = 1$; to mimic the quasi-likelihood approach for independent observations, we may take

$$w(.) = \frac{\xi'(.)}{v(.)}, \qquad (2{\cdot}2)$$

where $v(.) = \xi(.)\{1 - \xi(.)\}$.

When the failure times may be censored, the indicators $\{I(T_i \geqslant T_j)\}$ in (2·1) are not always observable. Since

$$E\left\{\frac{\Delta_j I(X_i \geqslant X_j)}{G^2(X_j)} \,\middle|\, Z_i, Z_j\right\} = E\left(E\left[\frac{I(T_i \geqslant T_j) I\{\min(C_i, C_j) \geqslant T_j\}}{G^2(T_j)} \,\middle|\, T_j, Z_i, Z_j\right]\right)$$
$$= E[I\{h(T_i) \geqslant h(T_j)\} | Z_i, Z_j] = \xi(Z_{ij}^{\mathrm{T}}\beta_0),$$

it seems natural to replace the dichotomous variable $I(T_i \geqslant T_j)$ in (2·1) with $\Delta_j I(X_i \geqslant X_j)\{\hat{G}(X_j)\}^{-2}$, where $\hat{G}$ is the Kaplan–Meier estimator for the 'survival' function G of the censoring variable. Let the resulting estimating function be denoted by

$$U(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\beta) Z_{ij} \left\{\frac{\Delta_j I(X_i \geqslant X_j)}{\hat{G}^2(X_j)} - \xi(Z_{ij}^{\mathrm{T}}\beta)\right\}. \qquad (2{\cdot}3)$$

In Appendix 1 we show that, if the weights $w(.)$ are positive, then the equation $U(\beta) = 0$ has, asymptotically, a unique solution $\hat{\beta}$. When $w = 1$ and the observed matrix $\sum\sum Z_{ij} Z_{ij}^{\mathrm{T}}$ is positive definite, which is trivially satisfied for most practical situations, the above equation has a unique solution. When F in (1·4) is the standard extreme value distribution, the weight function (2·2) becomes 1. In the next section, we show through examples that the estimation procedure with $w = 1$ works well for the proportional odds model and the model with standard normal error.

In Appendix 1, we also show that the distribution of $n^{-3/2}U(\beta_0)$ can be approximated by a normal distribution with mean 0 and variance-covariance matrx $\hat{\Gamma}$, where

$$\hat{\Gamma} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{\substack{k=1 \\ k \neq j}}^{n} \{w(Z_{ij}^{\mathrm{T}}\hat{\beta})\hat{e}_{ij}(\hat{\beta}) - w(Z_{ji}^{\mathrm{T}}\hat{\beta})\hat{e}_{ji}(\hat{\beta})\} \{w(Z_{ik}^{\mathrm{T}}\hat{\beta})\hat{e}_{ik}(\hat{\beta}) - w(Z_{ki}^{\mathrm{T}}\hat{\beta})\hat{e}_{ki}(\hat{\beta})\} Z_{ij} Z_{ik}^{\mathrm{T}}$$

$$- \frac{4}{n^3} \sum_{l=1}^{n} \frac{1-\Delta_l}{\{\sum_k I(X_k \geqslant X_l)\}^2} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\hat{\beta}) Z_{ij} \frac{\Delta_j I(X_i \geqslant X_j)}{\hat{G}^2(X_j)} I(X_j \geqslant X_l) \right\}^{\otimes 2},$$

$$\hat{e}_{ij}(\hat{\beta}) = \Delta_j I(X_i \geqslant X_j) \{\hat{G}(X_j)\}^{-2} - \xi(Z_{ij}^{\mathrm{T}}\hat{\beta})$$

and $v^{\otimes 2} = vv^{\mathrm{T}}$ for a vector v. It follows from the Taylor series expansion of $U(\hat{\beta})$ around β_0 that $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ is asymptotically equivalent to $n^{-3/2}\hat{\Lambda}U(\beta_0)$, where

$$\hat{\Lambda}^{-1} = \mathrm{n}^{-2} \sum \sum w(Z_{ij}^{\mathrm{T}}\hat{\beta}) \xi'(Z_{ij}^{\mathrm{T}}\hat{\beta}) Z_{ij}^{\otimes 2}.$$

Therefore, the distribution of $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ can be approximated by a normal distribution with mean 0 and covariance matrix $\hat{\Sigma} = \hat{\Lambda}\hat{\Gamma}\hat{\Lambda}$. Inferences for model (1·4) can then be made based on this large sample distribution of $\hat{\beta}$.

The above procedures are valid when the distribution of the censoring variable C is free of the covariate vector Z. This assumption may be strong for some observational studies, but is often satisfied in randomised controlled clinical trials. Now, suppose that one can discretise the covariate Z into K possible values. An analogue of the estimating function (2·3) that incorporates dependence between C and Z is

$$U^*(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\beta) Z_{ij} \left\{ \frac{\Delta_j I(X_i \geqslant X_j)}{\hat{G}_{Z_i}(X_j)\hat{G}_{Z_j}(X_j)} - \xi(Z_{ij}^{\mathrm{T}}\beta) \right\}, \qquad (2·4)$$

where $\hat{G}_Z(.)$ is the Kaplan–Meier estimator for the survival function of the censoring variable C based on those pairs $\{X_l, \Delta_l\}$ whose $Z_l = Z$ $(l = 1, \ldots, n)$.

In Appendix 2, we show that the distribution of $n^{-3/2}U^*(\beta_0)$ can be approximated by a normal distribution with mean 0 and variance-covariance matrix Γ^* given in (A2·1). It follows that the distribution of $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ can be approximated by a normal distribution with mean 0 and covariance matrix $\Sigma^* = \hat{\Lambda}\Gamma^*\hat{\Lambda}$.

If there is no obvious way to discretise the covariates, one may replace $\hat{G}_Z$ in (2·4) with a nonparametric functional estimate, for example a Kaplan–Meier estimate based on study subjects whose covariates are in a 'small neighbourhood' of Z. The corresponding estimator $\hat{\beta}$ is still consistent. If Z is univariate, we can choose the size of the neighbourhood to show that $\hat{\beta}$ is also asymptotically normal. More research, however, is needed for the multidimensional case.

3. EXAMPLES

The data in the first example are taken from Freireich (Cox, 1972). The observations are shown in Table 1. Censoring is heavy in Sample 2. The two-sample proportional hazards model fits the data well (Wei, 1984). Here, the group indicator is the only covariate; that is, $Z = 0$ if the observation is from the first sample and $Z = 1$ otherwise. Cox's maximum partial likelihood estimate for β_0 in (1·4) is $-1·51$ and the corresponding estimated standard error is 0·41. With the estimating function $U(\beta)$ in (2·3), the estimate for β_0 is

Table 1. *Times of remission (weeks) of leukaemia patients (Cox, 1972)*

Sample 1 (control)	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23
Sample 2 (drug 6-MP)	6, 6, 6, 6^+, 7, 9^+, 10, 10^+, 11^+, 13, 16, 17^+, 19^+, 20^+, 22, 23, 25^+, 32^+, 32^+, 34^+, 35^+

$^+$ Censored observations.

$-1{\cdot}74$. The estimated standard error based on $\hat{\Sigma}$ is 0·41. Since the censoring distributions for the two groups are obviously different, it is more appropriate to use the estimating function (2·4) to make inferences about β_0. The corresponding estimate is $-1{\cdot}64$ with an estimated standard error of 0·35. Thus, the results from the different approaches are very similar.

The second example is from the Veterans Administration lung cancer trial presented by Prentice (1973). Here, we will only use the subgroup of 97 patients with no prior therapy. The response variable is the patient's survival time and the covariates are tumour type, a factor with four levels (large, adeno, small, squamous), and performance status, a measure of general fitness on a scale from 0 to 100. Survival times range from 1 to 587 days and 6 of them are censored. Bennett (1983) and Pettitt (1984) used the proportional odds model to fit this set of data with various likelihood functions. In Table 2(a), we give estimates of β_0 using the estimating function U with weights $w=1$ and with the quasi-likelihood weights (2·2). Except for the only insignificant covariate, 'squamous versus large', the results from our procedures are similar to those from Bennett's nonparametric maximum likelihood (1983) and Pettitt's marginal likelihood methods (1984).

We also analysed the above data set using the proportional hazards model and the

Table 2. *Estimates (standard errors) of regression coefficients for lung cancer data (Prentice, 1973)*

(a) Proportional odds model

	New method with $w=1$	New method with w from (2·2)	Bennett (1983)	Pettitt (1984)
PS	−0·055 (0·010)	−0·055 (0·010)	−0·053 (0·010)	−0·055 (0·010)
Tumour type versus large				
adeno	1·556 (0·414)	1·559 (0·411)	1·314 (0·554)	1·302 (0·554)
small	1·496 (0·498)	1·494 (0·499)	1·383 (0·524)	1·438 (0·520)
squamous	−0·006 (0·572)	−0·004 (0·569)	−0·181 (0·588)	−0·177 (0·593)

	(b) Proportional hazards model		*(c) Model with $N(0,1)$ error*	
	New method	Cox	New method with $w=1$	New method with w from (2·2)
PS	−0·037 (0·007)	−0·024 (0·006)	−0·031 (0·006)	−0·031 (0·006)
Tumour type versus large				
adeno	1·061 (0·284)	0·851 (0·348)	0·894 (0·236)	0·897 (0·232)
small	1·020 (0·342)	0·548 (0·321)	0·860 (0·284)	0·855 (0·284)
squamous	−0·004 (0·391)	−0·214 (0·347)	−0·003 (0·328)	−0·004 (0·322)

PS, performance status

model with the standard normal error. The results are reported in Table 2(b), (c). Note that our estimates under the proportional hazards model are quite different from Cox's counterparts. This indicates that the proportional hazards model may not fit this data set well. For the case with the normal error, the results with weights 1 are almost identical to those with the 'optimal weights' (2·2).

4. NUMERICAL STUDIES

If the observed matrix $\sum\sum Z_{ij}Z_{ij}^{\mathrm{T}}$ is positive definite, the estimating function U in (2·3) with weights $w \equiv 1$ gives a unique estimate $\hat{\beta}$ of β_0. Moreover, $\hat{\beta}$ and its variance estimate can be easily obtained. Thus, if the procedure derived from this simple estimating function is reasonably efficient, it would be useful in practice. To this end, extensive empirical studies have been conducted to evaluate its efficiency. In one study, we considered a proportional hazards model (1·4) with two independent covariates, the first one from a uniform variable on (0, 1), and the second from a Bernoulli variable with 'success' probability 0·5. The survival time is obtained with h the natural logarithm function and ε having the standard extreme value distribution. Various uniform $U(0, c)$ censoring variables are considered, where c's are chosen with certain prespecified proportions of censoring. For each selected c, β_0 and sample size n, we simulate 500 realisations $\{(X_i, \Delta_i, Z_i)\}$ to estimate the ratio of the mean squared error of the Cox's maximum partial likelihood estimate to that of the new estimate. The results are reported in Table 3. Under the proportional hazards model, our new proposal is not expected to be as efficient as the Cox procedure; however, in the presence of moderate censoring it performs fairly well.

Table 3. *Estimated ratios of mean squared errors* (× 100): *Cox's versus new*

Censoring proportion (%)	$\beta_0^{\mathrm{T}}=(0,0)$ $n=100$		$n=200$		$\beta_0^{\mathrm{T}}=(-1,-1)$ $n=100$		$n=200$	
0	77*	78†	76*	79†	74*	80†	73*	75†
10	85	81	85	83	82	83	79	82
20	90	88	86	89	86	87	87	84
30	90	92	94	91	99	97	96	96

* Ratio for the first component of β_0.
† Ratio for the second component of β_0.

We also examined the performance of our simple procedure for the proportional odds model. Unfortunately, under this model the only inference procedure which has sound theoretical justification is for the simple two-sample problem. Bickel (1986) and Dabrowska & Doksum (1988a, p. 745) derived efficient estimates for β_0 under model (1·4) when Z is dichotomous. In our numerical comparisons, we let the error distribution be the standard logistic distribution, h be the identity function, and the censoring be various uniform variables $U(0, c)$, where c's are chosen with certain prespecified censoring proportions. For each n, c and β_0, we estimate the mean squared errors of our simple estimate and Dabrowska & Doksum's locally fully efficient estimate based on 500 simulated samples. The results are reported in Table 4. The new procedure appears to be as efficient as the optimal one proposed by Dabrowska & Doksum (1988a).

Empirical studies are also conducted to examine how sensitive the new procedure is with respect to the assumption that the censoring distribution G is free of the covariates.

Table 4. *Estimated ratios of mean squared errors* ($\times 100$): *Dabrowska & Doksum's* (1988*a*) *versus new*

Censoring proportion (%)	$n = 100$		$n = 200$	
	$\beta_0 = 0$	$\beta_0 = 1$	$\beta_0 = 0$	$\beta_0 = 1$
0	106	111	103	110
10	104	112	105	112
20	106	107	104	108
30	97	103	97	111

In general, we find that the inference procedure for β based on (2·3) is rather robust. For example, in one of our studies, we use the Stanford heart transplant data given by Miller & Halpern (1982) to check the adequacy of the new method when the censoring variable depends on a continuous covariate. For this particular study, it is well known that patient's censoring time depends on his or her entry age owing to the fact that the investigators tried to recruit younger patients during the later part of the study. In fact, if we use the Cox model to fit the censoring times with patient's age as the covariate, the estimate for the age effect is $-0{\cdot}018$ with an estimated standard error of 0·0136, indicating that an older patient tended to have a longer observation time than a younger patient did. To examine the age effect on patient's survival, the Cox proportional hazards model fits the data well with a quadratic age model (Lin, Wei & Ying, 1993). Based on the partial likelihood function, the point estimates for age and age^2 are $-0{\cdot}146$ and 0·00234, respectively. The corresponding estimated standard errors are 0·0554 and 0·00072. With the new procedure, the results are quite similar. The point estimates are $-0{\cdot}157$ and 0·00246 with estimated standard errors of 0·0581 and 0·00076, respectively.

We also simulate survival times from the above fitted Cox model with various types of covariate-dependent censorship to examine if the new confidence interval procedure based on (2·3) has correct coverage probabilities. The results are reported in Table 5. Each entry in the table is based on 500 random samples $\{(T_i, C_i, Z_i), i = 1, \ldots, 152\}$, where Z_i is the vector of the observed age and age^2 for the ith patient in the Stanford data, T_i is generated from the Cox model with parameters estimated from the Stanford data based on the partial likelihood, and the censoring variable C_i is generated from the Cox model with a linear age effect γ and a constant underlying hazard function η. The choice of γ reflects the degree of dependence between the censoring time and patient's entry age. The η is chosen with certain prespecified proportion of censoring. For each simulated sample, the Cox model with age and age^2 as the covariates is utilised to fit the data. The empirical coverage probabilities of the new interval procedure appear to be quite close to the nominal levels especially with moderate censoring.

Table 5. *Empirical coverage probabilities of interval procedures for the linear age effect* γ *with nominal level* 0·95

Censoring proportion (%)	$\gamma = -0{\cdot}02$		$\gamma = -0{\cdot}03$		$\gamma = -0{\cdot}04$		$\gamma = -0{\cdot}05$	
	New	Cox	New	Cox	New	Cox	New	Cox
10	0·95	0·95	0·92	0·93	0·95	0·96	0·95	0·97
20	0·93	0·96	0·95	0·97	0·93	0·95	0·95	0·97
30	0·92	0·95	0·94	0·95	0·92	0·93	0·93	0·94

γ, degree of dependence between censoring and age.

5. REMARKS

We have proposed a class of estimating functions for censored transformation models. The resulting estimation procedures for the regression parameters can easily be implemented and should be useful for analysing nonproportional hazards models.

When there is only one covariate in the model, Dabrowska & Doksum (1988b) find that using a wrong link function g in (1·3) has very little effect on the estimation of the parameter. However, this does not seem to be true for the case when there is more than one covariate in the model. Therefore, model-checking techniques are needed to examine the adequacy of the link function g or the distribution assumption for the error term and the deterministic portion of the fitted model.

ACKNOWLEDGEMENT

The authors are grateful to the Editor and two referees for helpful comments on the artifcle. This work was partially supported by the US National Institutes of Health and National Science Foundation.

APPENDIX 1

Asymptotic distribution of $\hat{\beta}$

First, we show that if the weights in $U(\beta)$ are positive, asymptotically there is a unique solution to the equation $U(\beta)=0$. To this end, let the distribution function of Z be denoted by H. Consider the quantity $n^{-2}U^{\mathrm{T}}(\beta)(\beta-\beta_0)$. With probability one, this converges to

$$\int_{z_1,z_2} w(z_{12}^{\mathrm{T}}\beta)(z_{12}^{\mathrm{T}}\beta - z_{12}^{\mathrm{T}}\beta_0)\{\xi(z_{12}^{\mathrm{T}}\beta_0)-\xi(z_{12}^{\mathrm{T}}\beta)\}dH(z_1)\,dH(z_2),$$

where $z_{12}=z_1-z_2$. Since $\xi(.)$ is a decreasing function, the above limit is nonnegative and is zero only when $\beta=\beta_0$. This implies that $\hat{\beta}$ is consistent.

It follows from a martingale integral representation for $(\hat{G}-G)/G$ (Gill, 1980, p. 37) that

$$\begin{aligned} n^{-3/2}U(\beta_0) &= n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}\left\{\frac{\Delta_j I(X_i\geqslant X_j)}{G^2(X_j)}-\xi(Z_{ij}^{\mathrm{T}}\beta_0)\right\} \\ &\quad + 2n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}\frac{\Delta_j I(X_i\geqslant X_j)}{G^2(X_j)}\,\frac{\{G(X_j)-\hat{G}(X_j)\}}{G(X_j)}+o_p(1) \\ &= n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}e_{ij}(\beta_0) + 2n^{-\frac{1}{2}}\sum_{k=1}^{n}\int_0^{\infty}\frac{q(t)}{\pi(t)}\,dM_k(t)+o_p(1), \qquad (\mathrm{A}1\cdot 1)\end{aligned}$$

where

$$e_{ij}(\beta_0)=\frac{\Delta_j I(X_i\geqslant X_j)}{G^2(X_j)}-\xi(Z_{ij}^{\mathrm{T}}\beta_0), \quad \pi(t)=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} I(X_i\geqslant t),$$

$$q(t)=\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}\frac{\Delta_j I(X_i\geqslant X_j)}{G^2(X_j)}I(X_j\geqslant t),$$

$$M_k(t)=I(X_k\leqslant t,\Delta_k=0)-\int_0^t I(X_k\geqslant u)\,d\Lambda_G(u),$$

and $\Lambda_G(.)$ is the common cumulative hazard function of C's. Using standard asymptotic theory of multivariate U-statistics (Wei & Johnson, 1985), one can show that the distribution of $n^{-3/2}U(\beta_0)$ is asymptotically normal with mean 0.

To calculate the corresponding limiting variance, note that the first term in (A1·1) is a U-statistic. Therefore, its variance can be approximated by

$$E\left[\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n}\{w(Z_{ij}^{\mathrm{T}}\beta_0)e_{ij}(\beta_0)-w(Z_{ji}^{\mathrm{T}}\beta_0)e_{ji}(\beta_0)\}\{w(Z_{ik}^{\mathrm{T}}\beta_0)e_{ik}(\beta_0)-w(Z_{ki}^{\mathrm{T}}\beta_0)e_{ki}(\beta_0)\}Z_{ij}Z_{ik}^{\mathrm{T}}\right].$$

For the second term in (A1·1), it follows from the standard variance calculation for a martingale that

$$\operatorname{var}\left\{2n^{-\frac{1}{2}}\sum_{k=1}^{n}\int_0^\infty \frac{q(t)}{\pi(t)}\,dM_k(t)\right\}\simeq 4\int_0^\infty \frac{q(t)q^{\mathrm{T}}(t)}{\pi(t)}\,d\Lambda_G(t).$$

To calculate the covariance between the first and second terms in (A1·1), note that, for $i\neq j$,

$$\begin{aligned}E\left\{\frac{\Delta_j I(X_i\geq X_j)}{G^2(X_j)}(1-\Delta_i)\frac{q^{\mathrm{T}}(X_i)}{\pi(X_i)}\right\} &= E\left[E\left\{\frac{\Delta_j I(C_i\geq X_j)}{G^2(X_j)}I(T_i>C_i)\frac{q^{\mathrm{T}}(C_i)}{\pi(C_i)}\,\middle|\, C_i, T_i, T_j\right\}\right]\\ &= E\left\{\int_0^\infty \frac{\Delta_j I(t\geq X_j)}{G^2(X_j)}I(T_i>t)G(t)\frac{q^{\mathrm{T}}(t)}{\pi(t)}\,d\Lambda_G(t)\right\}\\ &= E\left\{\int_0^\infty \frac{\Delta_j I(X_i\geq X_j)}{G^2(X_j)}I(X_j\leq t)I(X_i\geq t)\frac{q^{\mathrm{T}}(t)}{\pi(t)}\,d\Lambda_G(t)\right\}.\end{aligned}$$

This implies that

$$\begin{aligned}&\operatorname{cov}\left\{n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n}w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}e_{ij}(\beta_0),\ 2n^{-\frac{1}{2}}\sum_{k=1}^{n}\int_0^\infty\frac{q(t)}{\pi(t)}\,dM_k(t)\,\middle|\, Z_i, Z_j\right\}\\ &\simeq \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left[\int_0^\infty w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}\frac{\Delta_j I(X_i\geq X_j)}{G^2(X_j)}\frac{q^{\mathrm{T}}(t)}{\pi(t)}\,d\{M_i(t)+M_j(t)\}\,\middle|\, Z_i, Z_j\right]\\ &\simeq \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left[w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}\frac{\Delta_j I(X_i\geq X_j)}{G^2(X_j)}(1-\Delta_i)\frac{q^{\mathrm{T}}(X_i)}{\pi(X_i)}\right.\\ &\qquad\left.-\int_0^\infty w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}\frac{\Delta_j I(X_i\geq X_j)}{G^2(X_j)}\{I(X_i\geq t)+I(X_j\geq t)\}\frac{q^{\mathrm{T}}(t)}{\pi(t)}\,d\Lambda_G(t)\,\middle|\, Z_i, Z_j\right]\\ &\simeq -\frac{4}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left\{\int_0^\infty w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}\frac{\Delta_j I(X_i\geq X_j)}{G^2(X_j)}I(X_j\geq t)\frac{q^{\mathrm{T}}(t)}{\pi(t)}\,d\Lambda_G(t)\,\middle|\, Z_i, Z_j\right\}\\ &\simeq -4\int_0^\infty\frac{q(t)q^{\mathrm{T}}(t)}{\pi(t)}\,d\Lambda_G(t).\end{aligned}$$

Therefore, the limiting covariance matrix for $n^{-3/2}U(\beta_0)$ is

$$\begin{aligned}\lim_{n\to\infty}\Bigg[&\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n}\{w(Z_{ij}^{\mathrm{T}}\beta_0)e_{ij}(\beta_0)-w(Z_{ji}^{\mathrm{T}}\beta_0)e_{ji}(\beta_0)\}\\ &\times\{w(Z_{ik}^{\mathrm{T}}\beta_0)e_{ik}(\beta_0)-w(Z_{ki}^{\mathrm{T}}\beta_0)e_{ki}(\beta_0)\}Z_{ij}Z_{ik}^{\mathrm{T}}-4\int_0^\infty\frac{q(t)q^{\mathrm{T}}(t)}{\pi(t)}\,d\Lambda_G(t)\Bigg].\end{aligned}$$

A consistent estimator $\hat{\Gamma}$ for this matrix can be obtained by replacing β_0, G in e_{ij}, and Λ_G in the above, with $\hat{\beta}$, $\hat{G}$ and the Nelson estimate for the cumulative hazards function of C.

APPENDIX 2

Asymptotic distribution of $\hat{\beta}$ with discrete covariates

Since the covariate vector Z has a finite number of possible values, using the martingale integral representation for $\hat{G}_Z$, one can show that

$$n^{-3/2}U^*(\beta_0) \simeq n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}e_{ij}^*(\beta_0)$$

$$+ n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}\frac{\Delta_j I(X_i \geqslant X_j)}{G_{Z_i}(X_j)G_{Z_j}(X_j)}\left\{\frac{G_{Z_i}(X_j)-\hat{G}_{Z_i}(X_j)}{G_{Z_i}(X_j)} + \frac{G_{Z_j}(X_j)-\hat{G}_{Z_j}(X_j)}{G_{Z_j}(X_j)}\right\}$$

$$\simeq n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}e_{ij}^*(\beta_0) + 2n^{-\frac{1}{2}}\sum_{k=1}^{n}\int_0^\infty \frac{q^*(t)}{\pi_{Z_k}(t)}\,dM_{Z_k}(t),$$

where

$$e_{ij}^*(\beta_0) = \frac{\Delta_j I(X_i \geqslant X_j)}{G_{Z_i}(X_j)G_{Z_j}(X_j)} - \xi(Z_{ij}^{\mathrm{T}}\beta_0), \quad \pi_Z(t) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} I(X_i \geqslant t, Z_i = Z),$$

$$q^*(t) = \lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\beta_0)Z_{ij}\frac{\Delta_j I(X_i \geqslant X_j)}{G_{Z_i}(X_j)G_{Z_j}(X_j)} I(X_j \geqslant t),$$

$$M_Z(t) = I(X \leqslant t, \Delta = 0) - \int_0^t I(X \geqslant u)\,d\Lambda_{G_Z}(u),$$

and $\Lambda_{G_Z}(.)$ is the cumulative hazard function of the censoring variable C's whose covariate vector is Z. It follows that the distribution of $n^{-3/2}U^*(\beta_0)$ is asymptotically normal with mean 0.

Similar to the arguments in Appendix 1, the limiting covariance matrix for $n^{-3/2}U^*(\beta_0)$ is

$$\lim_{n\to\infty}\left[\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n}\{w(Z_{ij}^{\mathrm{T}}\beta_0)e_{ij}^*(\beta_0) - w(Z_{ji}^{\mathrm{T}}\beta_0)e_{ji}^*(\beta_0)\}\{w(Z_{ik}^{\mathrm{T}}\beta_0)e_{ik}^*(\beta_0) - w(Z_{ki}^{\mathrm{T}}\beta_0)e_{ki}^*(\beta_0)\}Z_{ij}Z_{ik}^{\mathrm{T}}\right.$$

$$\left. - \frac{4}{n}\sum_{k=1}^{n}\int_0^\infty \frac{q^*(t)q^{*\mathrm{T}}(t)}{\pi_{Z_k}^2(t)} I(X_k \geqslant t)\,d\Lambda_{G_{Z_k}}(t)\right].$$

Replacing all the theoretical quantities in the above with empirical ones, we obtain a consistent estimator Γ^* for the limiting covariance matrix, where

$$\Gamma^* = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n}\{w(Z_{ij}^{\mathrm{T}}\hat{\beta})\hat{e}_{ij}^*(\hat{\beta}) - w(Z_{ji}^{\mathrm{T}}\hat{\beta})\hat{e}_{ji}^*(\hat{\beta})\}\{w(Z_{ik}^{\mathrm{T}}\hat{\beta})\hat{e}_{ik}^*(\hat{\beta}) - w(Z_{ki}^{\mathrm{T}}\hat{\beta})\hat{e}_{ki}^*(\hat{\beta})\}Z_{ij}Z_{ik}^{\mathrm{T}}$$

$$- \frac{4}{n^3}\sum_{l=1}^{n}(1-\Delta_l)\left\{\sum_{k=1}^{n} I(X_k \geqslant X_l, Z_k = Z_l)\right\}^{-2}$$

$$\times\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} w(Z_{ij}^{\mathrm{T}}\hat{\beta})Z_{ij}\frac{\Delta_j I(X_i \geqslant X_j)}{\hat{G}_{Z_i}(X_j)\hat{G}_{Z_j}(X_j)} I(X_j \geqslant X_l)\right\}^{\otimes 2}, \qquad (\mathrm{A}2{\cdot}1)$$

$$\hat{e}_{ij}^*(\hat{\beta}) = \Delta_j I(X_i \geqslant X_j)\{\hat{G}_{Z_i}(X_j)\hat{G}_{Z_j}(X_j)\}^{-1} - \xi(Z_{ij}^{\mathrm{T}}\hat{\beta}).$$

REFERENCES

ANDERSEN, P. K. & GILL, R. D. (1982). Cox's regression model for counting process: A large sample study. *Ann. Statist.* **10**, 1100–20.

BICKEL, P. J. (1986). Efficient testing in a class of transformation models. In *Papers on Semiparametric Models at the ISI Centenary Session, Amsterdam*, Report MS-R8614, pp. 63–81. Amsterdam: Centrum voor Wiskunde en Informatica.

BENNETT, S. (1983). Analysis of survival data by the proportional odds model. *Statist. Med.* **2**, 273–7.
BOX, G. E. P. & COX, D. R. (1964). An analysis of transformations (with Discussion). *J. R. Statist. Soc.* B **26**, 211–52.
CLAYTON, D. G. & CUZICK, J. (1986). The semiparametrics Pareto model for regression analysis of survival times. In *Proc. Int. Statist. Inst.*, pp.23.3-1–18. Amsterdam: International Statistical Institute.
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.
CUZICK, J. (1988). Rank regression. *Ann. Statist.* **16**, 1369–89.
DABROWSKA, D. M. & DOKSUM, K. A. (1988a). Estimation and testing in a two-sample generalized odds-rate model. *J. Am. Statist. Assoc.* **83**, 744–9.
DABROWSKA, D. M. & DOKSUM, K. A. (1988b). Partial likelihood in transformation models with censored data. *Scand. J. Statist.* **15**, 1–23.
GILL, R. D. (1980). *Censoring and Stochastic Integrals*, Mathematical Centre Tract No. 124. Amsterdam: Mathematisch Centrum.
LIANG, K.-Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
LIN, D. Y., WEI, L. J. & YING, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–72.
MILLER, R. & HALPERN, J. (1982). Regression with censored data. *Biometrika* **69**, 521–31.
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.
WEI, L. J. (1984). Testing goodness of fit for proportional hazards model with censored observations. *J. Am. Statist. Assoc.* **79**, 649–52.
WEI, L. J. & JOHNSON, W. E. (1985). Combining dependent tests with incomplete repeated measurements. *Biometrika* **72**, 359–64.

[*Received December* 1994. *Revised May* 1995]