University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2004 *Paper* 146

Semiparametric Regression Analysis of Mean Residual Life with Censored Survival Data

Ying Qing Chen* Su-Chun Cheng†

*Division of Biostatistics, School of Public Health, University of California, Berkeley, yqchen@stat.berkeley.edu

†Department of Epidemiology & Biostatistics, University of California, San Francisco, scheng@biostat.ucsf.edu

http://biostats.bepress.com/ucbbiostat/paper146

Semiparametric Regression Analysis of Mean Residual Life with Censored Survival Data

Ying Qing Chen and Su-Chun Cheng

Abstract

As a function of time t, mean residual life is the remaining life expectancy of a subject given survival up to t. The proportional mean residual life model, proposed by Oakes & Dasu (1990), provides an alternative to the Cox proportional hazards model to study the association between survival times and covariates. In the presence of censoring, we develop semiparametric inference procedures for the regression coefficients of the Oakes-Dasu model using martingale theory for counting processes. We also present simulation studies and an application to the Veterans' Administration lung cancer data.

1 Introduction

Duration or time-to-event data have been studied in various research areas for decades. In fields such as public health, industrial reliability, demography or actuarial science, it is often of interest to analyse mean residual life as a function of time to characterise the stochastic behaviour of survival over time. For a nonnegative survival time T with finite expectation, the mean residual life at time t is

$$m(t) = E(T - t \mid T > t) \quad \text{for} \quad t \geq 0.$$

To assess the covariate effects on mean residual life, we consider the proportional mean residual life model proposed by Oakes & Dasu (1990):

$$m(t \mid Z) = m_0(t) \exp(\beta^{\mathrm{T}} Z), \tag{1}$$

where $m(\cdot \mid Z)$ is the mean residual life corresponding to the p-vector covariate Z, $m_0(t)$ is some unknown baseline mean residual life when $Z = 0$, and β is the regression parameter. Here the superscript T denotes the transpose of a vector.

The proportional mean residual life model is closely related to the accelerated failure time model (Kalbfleisch & Prentice, 1980):

$$\log T = \beta^{\mathrm{T}} Z + e, \tag{2}$$

where e is the random error variable with an unspecified distribution. To see the relationship, let $t = 0$ in model (1), and then $\log E(T \mid Z) = \log m(0 \mid Z) = \log m_0(0) + \beta^{\mathrm{T}} Z$. Careful choices of $m_0(t)$ in (1) and the distribution of e in (2) will lead to identical models. For example, the two models coincide when the distribution of T is exponential, whereupon $m_0(t)$ is a constant and e follows an extreme value distribution.

In general, there is no straightforward relationship between the proportional mean residual life model and the Cox proportional hazards model (Cox, 1972):

$$\lambda(t \mid Z) = \lambda_0(t) \exp(\gamma^{\mathrm{T}} Z), \tag{3}$$

where $\lambda(\cdot \mid Z)$ is the hazard function of T, $\lambda_0(\cdot) = \lambda(\cdot \mid Z = 0)$, and γ is the regression parameter. However, as noted in Dasu (1991) and Maguluri & Zhang (1994), the hazard functions $\widetilde{\lambda}(\cdot \mid Z)$ and $\widetilde{\lambda}_0(\cdot)$ of the forward recurrence times (Cox, 1962) in the renewal processes formed by T's corresponding to Z and $Z = 0$, respectively, follow that

$$\widetilde{\lambda}(t \mid Z) = \widetilde{\lambda}_0(t) \exp(-\beta^{\mathrm{T}} Z).$$

Furthermore, Oakes & Dasu (1990, Theorem 2) showed that when a model satisfies both the proportional hazards and the proportional mean residual life assumptions, its underlying distributions then belong to the Hall-Wellner class of distributions with linear mean residual life functions (Hall & Wellner, 1981). Note that both models (1) and (3) hold for exponential distributions with $\gamma = -\beta$.

Previous work on mean residual life has focused on single-sample and two-sample cases. The methods for these cases are proposed in the work of Oakes & Dasu (2003) and the references therein. For regression analysis, Maguluri & Zhang (1994) used both the aforementioned model-relationships to develop estimation procedures for β in (1), but their approaches were essentially for uncensored survival data. To accommodate censoring, one straightforward approach is to apply the inverse-probability-of-censoring-weighted paradigm by Robins & Rotnitzky (1992) to the estimating equations built on complete event times; however, this would require estimating or modelling the distribution of censoring.

In this article, we employ martingale theory for counting processes to develop new inference procedures for the regression analysis of model (1) with censored data. Our semiparametric approach mimics the Cox partial score function and retains similar appealing properties. The resulting estimator for β resembles the maximum partial likelihood estimator if the two classes of models coincide. Moreover, our estimator for the baseline mean residual life function $m_0(\cdot)$ takes a closed form. In §2 of this article, we elaborate and study the proposed estimator for β and its efficiency. Numerical studies, including Monte Carlo simulations and an analysis of data from the VA lung cancer trial, are summarised in §3. Several relevant issues are discussed in §4. Technical proofs are collected in the Appendix.

2 Inference Procedures

The maximum likelihood estimation procedures can be applied to model (1) in situations when we have sufficient knowledge to specify the baseline mean residual life in some parametric forms, such as the Hall-Wellner family (Hall & Wellner, 1981). To gain modelling flexibility, it is often more desirable not to assume particular forms for the underlying mean residual life function. To that end, in this section we present semiparametric inference procedures to estimate the parameter of primary interest, β, in (1).

Let T and C be the failure time and potential censoring time, respectively. Conditional on the p-vector covariate Z, T and C are assumed to be independent. The observed data set consists of n independent triplets of (X_i, Δ_i, Z_i), where $i = 1, \ldots, n$, $X_i = \min(T_i, C_i)$, and $\Delta_i = I(T_i \leq C_i)$. Here, $I(\cdot)$ is the indicator function, which is 1 if the condition is satisfied and 0 otherwise. In addition, let $N_i(t) = I(X_i \leq t)\Delta_i$, $Y_i(t) = I(X_i \geq t)$, and $\Lambda_i(t)$ be the cumulative hazard function of T_i. It follows from Corollary 1.4.1 in Fleming & Harrington (1991) that

$$E\left\{dN_i(t) \mid \mathcal{F}_{t-}; \beta_*, m_*(\cdot)\right\} = Y_i(t)d\Lambda_i(t; \beta_*, m_*),$$

where $\mathcal{F}_t$ belongs to the right continuous filtration $\{\mathcal{F}_t : t \geq 0\}$ defined by

$$\mathcal{F}_t = \sigma\{N_i(u), Y_i(u+), Z_i : \ 0 \leq u \leq t, \ i = 1, \ldots, n\},$$

and β_* and $m_*(\cdot)$ are the true values of the parameters β and $m_0(\cdot)$ in (1), respectively. If we denote $M_i(t; \beta, m_0) = N_i(t) - \int_0^t Y_i(s)d\Lambda_i(s; \beta, m_0)$ for $i = 1, \ldots, n$, then $\{M_i(t; \beta_*, m_*)\}$ are zero-mean $\mathcal{F}_t-$martingales. Therefore it is natural to estimate β_* and m_* by estimating equations parallel to the partial score equations:

$$\begin{aligned} &\sum_{i=1}^{n} \left\{dN_i(t) - Y_i(t)d\Lambda_i(t; \beta, m_0)\right\} = 0, \qquad 0 \leq t \leq \tau, \\ &\sum_{i=1}^{n} \int_0^{\tau} Z_i \left\{dN_i(t) - Y_i(t)d\Lambda_i(t; \beta, m_0)\right\} = 0. \end{aligned} \tag{4}$$

To avoid a lengthy technical discussion on the tail behaviour of the limiting distribution, we assume $0 < \tau = \inf\{t : \text{pr}(X > t) = 0\} < \infty$, and it is used throughout the rest of the

article. If necessary, Ying's (1993) elegant treatment on the asymptotic properties beyond τ can be adapted to our method. Note that $m_0(\tau) = 0$ under this assumption.

It is well known that the survival function of T given Z is

$$S(t \mid Z) = \mathrm{pr}(T \geq t \mid Z) = \frac{m(0 \mid Z)}{m(t \mid Z)} \exp\left\{ - \int_0^t \frac{1}{m(u \mid Z)}\, du \right\},$$

and consequently that $m_0(t) d\Lambda_i(t) = \exp(-\beta^{\mathrm{T}} Z_i) dt + dm_0(t)$ under model (1). Then analogous to (4), the following estimating equations can be used to estimate (β_*, m_*):

$$\sum_{i=1}^{n} \left[m_0(t) dN_i(t) - Y_i(t) \left\{ \exp(-\beta^{\mathrm{T}} Z_i) dt + dm_0(t) \right\} \right] = 0, \qquad 0 \leq t \leq \tau, \tag{5}$$

$$\sum_{i=1}^{n} \int_0^{\tau} Z_i \left[m_0(t) dN_i(t) - Y_i(t) \left\{ \exp(-\beta^{\mathrm{T}} Z_i) dt + dm_0(t) \right\} \right] = 0. \tag{6}$$

As mentioned in §1, when a model satisfies both (1) and (3), its underlying distributions then belong to the Hall-Wellner family, where $m_0(t) = a\, t + b$, $a > -1$ and $b > 0$. In this case, the proposed estimating equations (6) are asymptotically equivalent to the partial score equations for β.

Following a simple algebraic manipulation, estimating equation (5) becomes

$$m_0(t) dQ_1(t) - dm_0(t) = Q_2(t; \beta) dt, \quad 0 \leq t \leq \tau, \tag{7}$$

where $dQ_1(t) = \sum_{i=1}^n dN_i(t) / \sum_{i=1}^n Y_i(t)$ and $Q_2(t; \beta) = \sum_{i=1}^n Y_i(t) \exp(-\beta^{\mathrm{T}} Z_i) / \sum_{i=1}^n Y_i(t)$. An intriguing feature of equation (7) is that it is a first-order linear ordinary differential equation in $m_0(t)$, and thus it has a closed-form solution

$$\widehat{m}_0(t; \beta) = \left[\exp\left\{ - \int_0^t dQ_1(u) \right\} \right]^{-1} \int_t^{\tau} \exp\left\{ - \int_0^u dQ_1(s) \right\} Q_2(u; \beta) du, \tag{8}$$

given that the marginal mean residual life function is continuously differentiable on $[0, \tau]$.

To obtain an estimator for β_*, we replace $m_0(t)$ and $dm_0(t)$ in (6) with $\widehat{m}_0(t; \beta)$ and $d\widehat{m}_0(t; \beta) = \widehat{m}_0(t; \beta) dQ_1(t) - Q_2(t; \beta) dt$, respectively. Then it is straightforward to show that the resulting equations (6) divided by n are algebraically equivalent to

$$U(\beta) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} \left\{ Z_i - \bar{Z}(t) \right\} \left\{ \widehat{m}_0(t; \beta) dN_i(t) - Y_i(t) \exp(-\beta^{\mathrm{T}} Z_i) dt \right\} = 0, \tag{9}$$

where $\bar{Z}(t) = \sum_{i=1}^n Y_i(t)Z_i / \sum_{i=1}^n Y_i(t)$. Let $\widehat{\beta}$ be the solution to equations (9). We show in the Appendix that, under the specified regularity conditions, the random vector $n^{1/2}(\widehat{\beta} - \beta_*)$ converges weakly to a p-vector normal variable with mean zero and covariance matrix $A^{-1}VA^{-1}$, where matrices A and V are given in (A·2) and (A·3), respectively. In addition, A and V can be consistently estimated by their empirical counterparts

$$\widehat{A} = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{Z_i - \bar{Z}(t)\right\}^{\otimes 2} Y_i(t)\exp(-\widehat{\beta}^{\mathrm{T}} Z_i)dt,$$

and

$$\widehat{V} = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{Z_i - \bar{Z}(t)\right\}^{\otimes 2} Y_i(t)\widehat{m}_0(t;\widehat{\beta})\left\{\exp(-\widehat{\beta}^{\mathrm{T}} Z_i)dt + d\widehat{m}_0(t;\widehat{\beta})\right\},$$

respectively, where $v^{\otimes 2}$ denotes vv^{T} for a vector v. Inferences for β_* can then be made through this large-sample distribution of $\widehat{\beta}$.

By the *ad hoc* nature of $U(\beta)$, the estimator $\widehat{\beta}$ is not necessarily efficient although it has properties such as consistency and asymptotic normality that can be used to make valid inferences about β_*. To gain efficiency, one common remedy is to use the weighted version of the estimating equations (5) and (6):

$$\sum_{i=1}^n W(t)\left[m_0(t)dN_i(t) - Y_i(t)\left\{\exp(-\beta^{\mathrm{T}} Z_i)dt + dm_0(t)\right\}\right] = 0, \qquad 0 \le t \le \tau,$$

$$\sum_{i=1}^n \int_0^\infty W(t)Z_i\left[m_0(t)dN_i(t) - Y_i(t)\left\{\exp(-\beta^{\mathrm{T}} Z_i)dt + dm_0(t)\right\}\right] = 0,$$

where $W(t)$ is an $\mathcal{F}_t$-measurable weight function which converges uniformly to some deterministic function $w(t)$ almost surely. Let $\widehat{\beta}_w$ denote the resulting weighted estimator for β_*. It is straightforward to derive the asymptotic variance of $\widehat{\beta}_w$ in the form of $A_w^{-1}V_wA_w^{-1}$, where matrices A_w and V_w are given in (A·4) and (A·5), respectively.

By applying the Cauchy-Schwarz inequality to $A_w^{-1}V_wA_w^{-1}$, it follows that the optimal weight is proportional to

$$w_*(t) = \frac{\exp(-\beta_*^{\mathrm{T}} Z)}{m_*(t)\{\exp(-\beta_*^{\mathrm{T}} Z) + m_*'(t)\}}, \tag{10}$$

which minimises the asymptotic variance of $\widehat{\beta}_w$ and reaches its Cramér-Rao bound, where $m'_*(t) = dm_*(t)/dt$. In fact, the partial derivative of the log full likelihood function with respect to β is

$$-\sum_{i=1}^{n}\int_0^\tau \frac{\exp(-\beta^{\mathrm{T}}Z_i)}{m_0(t)\{\exp(-\beta^{\mathrm{T}}Z_i)+m'_0(t)\}}\, Z_i\left[m_0(t)dN_i(t) - Y_i(t)\left\{\exp(-\beta^{\mathrm{T}}Z_i)+m'_0(t)\right\}dt\right]. \tag{11}$$

This is an interesting, yet not surprising, coincidence between the optimal weight function (10) and the coefficient in (11) because of the way that we mimic the partial likelihood score equation in constructing our estimating equations. In the special case of the proportional hazards model, under which $m_0(t)$ belongs to the Hall-Wellner family, $m_0(t)w_*(t)$ is constant and independent of t as in the partial likelihood score equation.

3 Numerical Studies

Simulation studies were conducted to assess the finite-sample properties of the proposed estimation procedures. We consider the sample size n being 100 or 200 with two covariates $Z = (Z_1, Z_2)^{\mathrm{T}}$ for each of n subjects. The covariate Z_1 is a Bernoulli random variable with success probability 0.5 and Z_2 is a uniform random variable on [0,1]. We choose $m_*(t) = t+1$, which corresponds to the Pareto distribution with the baseline survival function being $(1+t)^{-2}$. Failure times are generated according to model (1) where the true parameter is selected to be $(0,0)^{\mathrm{T}}$ or $(1,1)^{\mathrm{T}}$. Independent censoring times are generated from the uniform distribution on $[0,c]$, where the constant c is chosen to result in no censoring or, on average, about 10% or 30% censored observations.

The simulation results are summarised in Table 1, with each entry calculated based on 1000 data sets. The results show that the estimates of β_* are virtually unbiased and the nominal 95% confidence intervals have proper coverage probabilities. By examining the mean standard error estimates, we find that the weighted estimators tend to be much more efficient than the unweighted ones, especially for uncensored data. This advantage, however, is relatively small in the presence of censoring.

We now apply model (1) to a data set from the well-known Veterans' Administration lung cancer trial (Prentice, 1973). This data set has been analysed by several authors such as Pettitt (1984) and Cheng et al. (1995). To contrast with their results, we use the same subgroup of 97 patients with no prior therapy. The response variable is the patients' survival times, which range from 1 to 587 days with 6 censored observations. The first covariate is the performance status, on a scale from 0 to 100. The second one is the tumour type, which has 4 levels (large, adeno, small, squamous) and is treated as categorical with the large type being the reference group. In Table 2, we present estimates of the regression coefficients using the unweighted estimation function U in (9) and its weighted version with weights in the empirical form of (10). We also show the results using the proportional hazards model, the proportional odds model, and the accelerated failure time model in Table 2.

Under the Cox model and the proportional odds model, the only nonsignificant estimate is the one for comparing squamous and large tumour types in differentiating their associated hazards and failure odds, respectively. However, from the viewpoint of evaluating the effects of the tumour types on the mean residual life, using both the weighted and unweighted estimating functions for the proportional mean residual life model, none of the comparisons against the large type is significant . This finding is consistent with the results using the accelerated failure time model, which also yields nonsignificant tumour-type effects in comparing the log-transformed overall survival times.

It is interesting that the estimates using the Oakes-Dasu model (1) are close in magnitude to their counterparts under the Cox model (3), but have opposite signs. As noted in §1 and §2, this may suggest that models (1) and (3) coincide for this data set. Under this coincidence, the baseline mean residual life would be linear and belong to the Hall-Wellner family. Figure 1(a) plots the estimated baseline mean residual life and its lowess curve as of a function of time t. The curve suggests a change point before and after which the behaviour of the baseline mean residual life differs. The curve for the initial period $[0, 150]$ is noticeably flat, although the linear relationship appears strong for $t > 150$. Empirically we find that

the parametric maximum likelihood procedure with $m_0(t) = a\,t + b$, where $a > -1$ and $b > 0$, almost always yields unstable estimators toward the boundary of the space of the parameters a and b, except when $m_0(t)$ is predetermined.

Although the baseline mean residual life $m_0(t)$ may not be of primary interest in the semiparametric proportional mean residual life model, there is a natural constraint on $m_0(t)$ such that the mean conditional life $m_0(t) + t = E(T|T > t)$ should be monotonically non-decreasing. As shown in Figure 1(b), this constraint appears satisfied in this example. To check the adequacy of the proportionality in model (1), both the estimated marginal mean residual life function, without adjusting for any of the covariates, and the estimated baseline mean residual life are plotted in Figure 1(c). In log scale, their lowess curves appear to be parallel to each other and their difference function appears constant. This may suggest a reasonable goodness-of-fit of the proportionality assumption.

4 Discussion

There is an alternative way to derive $\widehat{m}_0(t;\beta)$ in (8). To this end, let $F_Z(\cdot)$ and $f_Z(\cdot)$ be the marginal distribution and density functions of Z, respectively, and $S(\cdot)$ be the marginal survival function of T. Under model (1),

$$m_0(t)S(t|z) = \exp(-\beta^{\mathrm{T}} z)m(t|z)S(t|z) = \exp(-\beta^{\mathrm{T}} z)\int_t^\tau S(u|z)du \tag{12}$$

for any possible $Z = z \in \mathrm{supp}\{z \in \mathbb{R}^p; F_Z\}$. Therefore, by the Bayes Theorem,

$$\begin{aligned} m_0(t) &= \frac{1}{S(t)}\int_z m_0(t)S(t \mid z)dF_Z(z) \\ &= \frac{1}{S(t)}\int_z \left\{\exp(-\beta^{\mathrm{T}} z)\int_t^\tau S(u \mid z)du\right\} dF_Z(z) \\ &= \frac{1}{S(t)}\int_t^\tau \left\{\int_z \frac{f_Z(z \mid T \geq u)S(u)}{f_Z(z)}\exp(-\beta^{\mathrm{T}} z)dF_Z(z)\right\} du \\ &= \frac{1}{S(t)}\int_t^\tau \left\{S(u)\int_z \exp(-\beta^{\mathrm{T}} z)dF_Z(z \mid T \geq u)\right\} du. \end{aligned} \tag{13}$$

Thus $\widehat{m}_0(t;\beta)$ in (8) can also result from substituting $\exp\{-\int_0^t dQ_1(u)\}$ and $Q_2(u;\beta)$ for the theoretical quantities $S(t)$ and $\int_z \exp(-\beta^{\mathrm{T}} z)dF_Z(z|T \geq u)$ in (13), respectively. It follows

that any corresponding consistent estimators for those two quantities may yield consistent estimators for $m_0(t)$, given appropriate regularity conditions. If we write $Q_1(t) = \int_0^t dQ_1(u)$, then it is the Nelson-Aalen estimator for the marginal cumulative hazard function at t and thus $\widehat{S}_{\mathrm{NA}}(t) = \exp\{-Q_1(t)\}$. As a result, (8) simply becomes

$$\widehat{m}_0(t;\beta) = \widehat{S}_{\mathrm{NA}}^{-1}(t) \int_t^\tau \widehat{S}_{\mathrm{NA}}^{-1}(u) Q_2(u;\beta) du.$$

In the single-sample setting, $Q_2(t;\beta)$ is 1. Consequently, $\widehat{m}_0(t;\beta)$ reduces to the plug-in estimator for $m_0(t) = \int_t^\tau S(u)du/S(t)$ if $S(t)$ is estimated by $\widehat{S}_{\mathrm{NA}}(t)$.

The proposed inference procedures in §2 are not examined within the framework of semiparametric efficiency bound calculation. To that end, one can follow the approach of Lai & Ying (1992) to study the parametric subfamilies in the form of

$$m(t \mid Z) = m_0(t)\{1 + \alpha \bar{m}_0(t)\} \exp(\beta^{\mathrm{T}} Z),$$

where $(\alpha, \beta)^{\mathrm{T}}$ are unknown parameters and $(m_0(\cdot), \bar{m}_0(\cdot))^{\mathrm{T}}$ are completely specified functions. Then, from the full likelihood of $(\alpha, \beta)^{\mathrm{T}}$, the Fisher information matrix can be calculated at $\alpha = 0$ and $\beta = \beta_*$, and it leads to the semiparametric information bound for estimating β_*. The complexity of such implementation is beyond the scope of this manuscript, and we intend to investigate this problem in future research.

There remain other issues with regard to model (1) that require further study. For example, as noted in §3, $m(t) + t$ should be nondecreasing in t. Although the asymptotic limit of $\widehat{m}(t, \widehat{\beta})$ is $m_*(t) \geq 0$ provided that $m_*(t)$ is proper, there is no guarantee that the finite-sample estimator $\widehat{m}(t, \widehat{\beta}) + t$ would maintain the necessary monotonicity, *per se*. Our future work includes implementing algorithms such as the pooled-adjacent-violators to obtain more reasonable estimates of $m_0(t)$. In addition, it is important to develop inference procedures for predicting individual mean residual life and conditional life expectancy. We also plan to develop objective analytical procedures for model checking or model selection between the proportional mean residual life model and other popular models, such as the accelerated failure time model and the proportional hazards model.

Appendix

Asymptotics of Estimators

To establish the onward asymptotic properties, we assume the necessary regularity conditions or analogous ones specified in Fleming & Harrington (1991, p. 289-90). In addition, we assume the following conditions:

1. $\inf \text{supp}(F) \leq \inf \text{supp}(G)$, where $F(\cdot)$ and $G(\cdot)$ are the distribution functions of T and C, respectively,
2. there exists some constant, $d_Z > 0$, such that $\text{pr}\{|Z| > d_Z\} = 0$,
3. $m_*(t)$ is continuously differentiable on $[0, \tau]$.

Consistency of $\widehat{m}_0(t; \beta_*)$: Consider the functional $D : m_0 \in \mathcal{M}_0 \mapsto D(m_0) \in \mathcal{D}$, where

$$D(m_0)(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \left[m_0(u)dN_i(u) - Y_i(u)\left\{\exp(-\beta_*^{\mathrm{T}}Z_i)du + dm_0(u)\right\}\right].$$

Here, $\mathcal{M}_0$ is the proper space for all the possible baseline mean residual life functions equipped with a norm defined as $\|m_1 - m_2\| = \sup_{t\in[0,\tau]}|m_1(t) - m_2(t)|$, $m_1, m_2 \in \mathcal{M}_0$, and $\mathcal{D} = D(\mathcal{M}_0)$. Let $\widehat{m}_0^*(t)$ denote $\widehat{m}(t; \beta_*)$, and $d_1 = D(m_0; \widehat{m}_0^* - m_0)(t)$ denote

$$D(\widehat{m}_0^*)(t) - D(m_0)(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \left[(\widehat{m}_0^* - m_0)(u)dN_i(u) - Y_i(u)\left\{d(\widehat{m}_0^* - m_0)(u)\right\}\right].$$

Then d_1 maps $\mathcal{M}_d = \{c(\widehat{m}_0^* - m_0) : \widehat{m}_0^*, m_0 \in \mathcal{M}_0, c \text{ real}\}$ to $\mathcal{D}_d = \{c(D(\widehat{m}_0^*) - D(m_0)) : D(\widehat{m}_0^*), D(m_0) \in \mathcal{D}, c \text{ real}\}$ (Serfling, 1980). For any fixed $\epsilon > 0$, if $\|m_1 - m_2\| > \epsilon$, then $\|D(m_1) - D(m_2)\|$ is

$$\begin{aligned}
&\sup_{t\in[0,\tau]}\left|\frac{1}{n}\sum_{i=1}^{n}\int_0^t \left[\left\{m_1(u) - m_2(u)\right\}dN_i(u) - Y_i(u)d\left\{m_1(u) - m_2(u)\right\}\right]\right| \\
= &\sup_{t\in[0,\tau]}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\left\{m_1(X_i) - m_2(X_i)\right\}\Delta_i I(X_i < t) - \left\{m_1(t\wedge X_i) - m_2(t\wedge X_i)\right\}\right]\right| \\
= &\sup_{t\in[0,\tau]}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{I(X_i > t) + I(X_i < t)(1-\Delta_i)\right\}\left\{m_1(t\wedge X_i) - m_2(t\wedge X_i)\right\}\right| \qquad \text{(A·1)} \\
> &c_\tau\epsilon,
\end{aligned}$$

for some constant c_τ. As a result, the inverse mapping of d_1,$d_1^{-1} : \mathcal{D}_d \to \mathcal{M}_d$, is continuous and hence bounded since ϵ is arbitrarily small. In addition, by the law of large numbers and the continuity of $m_*(t)$, we know that $\sup_{t\in[0,\tau]} |D(\widehat{m}_0^*)(t) - D(m_*)(t)| = \sup_{t\in[0,\tau]} | - D(m_*)(t)| \to 0$ almost surely. Therefore, $\widehat{m}_0^*$ is in an arbitrarily small neighbourhood of m_* in $\mathcal{M}_0$ as $n \to \infty$. It follows that $\widehat{m}_0(t;\beta_*)$ converges to $m_*(t)$ almost surely.

Consistency of $\partial U(\beta_*)/\partial\beta$: It follows from the law of large numbers and (12) that

$$\begin{aligned}
\frac{\partial \widehat{m}_0(t,\beta_*)}{\partial\beta} &= \frac{-1}{\widehat{S}_{\mathrm{NA}}(t)} \int_t^\tau \widehat{S}_{\mathrm{NA}}(u) \frac{\sum_{i=1}^n Z_i Y_i(u) \exp(-\beta_*^{\mathrm{T}} Z_i)}{\sum_{i=1}^n Y_i(u)} du \\
&= \frac{-1}{E\{S(t|Z)\}} \int_t^\tau E\{S(u|Z)\} \frac{\int_z z \exp(-\beta_*^{\mathrm{T}} z) S_*(u|z) dF_Z(z)}{E\{S_*(u|Z)\}} du + o_p(1) \\
&= \frac{-1}{E\{S(t|Z)\}} \int_z z \exp(-\beta_*^{\mathrm{T}} z) \left\{ \int_t^\tau S(u|z) du \right\} dF_Z(z) + o_p(1) \\
&= \frac{-1}{E\{S(t|Z)\}} \int_z z\, m_*(t) S(t|z) dF_Z(z) + o_p(1) \\
&= -m_*(t)\mu_z(t) + o_p(1),
\end{aligned}$$

where $F_Z(\cdot)$ and $\widehat{S}_{\mathrm{NA}}(\cdot)$ are defined in §4 respectively, $S_*(t|Z) = \mathrm{pr}(X \geq t|Z)$, and $\mu_z(t) = E\{ZS_*(t|Z)\}/E\{S_*(t|Z)\}$. Note that $\mu_z(t)$ is the limit of $\bar{Z}(t)$ as $n \to \infty$. Then, by differentiating (9) with respect to β, we have

$$\begin{aligned}
\frac{\partial U(\beta_*)}{\partial\beta} = & \quad \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{Z_i - \bar{Z}(t)\right\} \left\{ \frac{\partial \widehat{m}_0(t,\beta_*)}{\partial\beta} dN_i(t) + Y_i(t)\exp(-\beta_*^{\mathrm{T}} Z_i) Z_i dt \right\}^{\mathrm{T}} \\
= & \quad \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{Z_i - \bar{Z}(t)\right\} \left\{-m_*(t)\mu_z^{\mathrm{T}}(t)\right\} dM_i(t;\beta_*,m_*) \\
& + \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{Z_i - \bar{Z}(t)\right\} \left\{Z_i - \mu_z(t)\right\}^{\mathrm{T}} Y_i(t)\exp(-\beta_*^{\mathrm{T}} Z_i) dt + o_p(1) \\
= & \quad A + o_p(1),
\end{aligned}$$

where

$$A = \int_0^\tau E\left[\{Z - \mu_z(t)\}^{\otimes 2} S_*(t \mid Z) \exp(-\beta_*^{\mathrm{T}} Z)\right] dt. \qquad (\mathrm{A}\cdot 2)$$

Asymptotic normality of $n^{1/2}U(\beta_*)$: By the result in (A·1),

$$\begin{aligned} U(\beta_*) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left[\left\{Z_i - \bar{Z}(t)\right\}\left\{\widehat{m}_0(t;\beta_*)dN_i(t) - Y_i(t)\exp(-\beta_*^{\mathrm{T}}Z_i)dt\right\}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left\{Z_i - \bar{Z}(t)\right\} m_*(t)dM_i(t) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{Z_i - \bar{Z}(X_i)\right\}\left\{\widehat{m}_0(X_i;\beta_*) - m_*(X_i)\right\} I(X_i < t)\Delta_i + o_p(1) \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left\{Z_i - \bar{Z}(t)\right\} m_*(t)dM_i(t) - \frac{1}{n}\sum_{i=1}^{n}\frac{\left\{Z_i - \bar{Z}(X_i)\right\}}{\frac{1}{n}\sum_{j=1}^{n} I(X_j > t)}\frac{1}{n}\sum_{j=1}^{n}\int_0^t m_*(u)dM_j(u) + o_p(1) \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left(Z_i - \bar{Z}(t) - \frac{E[S(t|Z)\{Z-\mu_z(t)\}]}{E\{S(t|Z)\}}\right) m_*(t)dM_i(t) + o_p(1). \end{aligned}$$

Therefore $n^{1/2}U(\beta_*)$ is asymptotically normal with mean zero and variance-covariance matrix

$$V = \int_0^\tau E\left[\{Z-\mu_z(t)\}^{\otimes 2} S_*(t|Z)m_*(t)\left\{\exp(-\beta_*^{\mathrm{T}}Z)dt + dm_*(t)\right\}\right]. \tag{A·3}$$

Similar to (A·2) and (A·3), we derive

$$A_w = \int_0^\tau E\left[w(t)\{Z-\mu_z(t)\}^{\otimes 2}\ S_*(t|Z)\exp(-\beta_*^{\mathrm{T}}Z)\right]dt, \tag{A·4}$$

and

$$V_w = \int_0^\tau E\left(\left[w^2(t)\{Z-\mu_z(t)\}\right]^{\otimes 2} S_*(t|Z)m_*(t)\left\{\exp(-\beta_*^{\mathrm{T}}Z)dt + dm_*(t)\right\}\right). \tag{A·5}$$

Consistency of $\widehat{\beta}$, $\widehat{m}_0(t;\widehat{\beta})$, $\widehat{A}$ *and* $\widehat{V}$. For an arbitrarily small neighbourhood of $\beta_* \in \mathbb{R}^p$, denoted by $R(\beta_*)$, $U(\cdot)$ maps it to an open connected set in $\mathbb{R}^p$. In fact, since $U(\beta_*) \to 0$ can extended to any $\beta \in R(\beta_*)$ with stronger regularity conditions on uniform convergence, $\widehat{\beta}$ would fall in the same neighbourhood with probability one given A is nonsingular. Hence the consistency of $\widehat{\beta}$ is warranted. Using similar techniques, it is also true that $\widehat{m}_0(t;\widehat{\beta})$ is consistent. The consistency of $\widehat{A}$ and $\widehat{V}$ is straightforward following Taylor series expansions of $\widehat{A}(\widehat{\beta})$ and $\widehat{V}(\widehat{\beta})$ around β_*. Furthermore, the asymptotic distribution of $\widehat{\beta}$ follows from a Taylor series expansion of $U(\widehat{\beta})$ around β_*.

References

Cheng, S. C., Wei, L. J. & Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835-45.

Cox, D. R. (1962) *Renewal Theory*. New York: Wiley.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. R. Statist. Soc.* B **34**, 187-220.

Dasu, T. (1991). *The Proportional Mean Residual Life Model*. Ph.D. Dissertation. Rochester: University of Rochester.

Fleming, T. R. & Harrington, D. P. (1991) *Counting Process and Survival Analysis*. New York: Wiley.

Hall, W. J. & Wellner, J. A. (1981) Mean residual life. In *Proc. Int. Symp. Statistic and Related Topics* (M. Csörgő, et al., eds.), 169-81. Amsterdam: North-Holland.

Kalbfleisch, J. D. & Prentice, R. L. (1980) *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Lai, T. L. & Ying, Z. (1992). Asymptotically efficient estimation in censored and truncated regression models. *Statist. Sinica* **2**, 17-46.

Maguluri, G. & Zhang, C.-H. (1994) Estimation in the mean residual life regression model. *J. R. Statist. Soc.* B **56**, 477-89.

Oakes, D. & Dasu, T. (1990) A note on residual life. *Biometrika* **77**, 409-10.

Oakes, D. & Dasu, T. (2003) Inference for the Proportional Mean Residual Life Model. In *Crossing Boundaries: Statistical Essays in Honor of Jack Hall*, Kolassa, J.E. and Oakes, D. (eds.), IMS monographs, in press.

Pettitt, A. N. (1984). Proportional odds model for survival data and estimates using ranks. *Appl. Statist.* **33**, 169-75.

Robins, J. M. & Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology, Methodological Issues,* N. P. Jewell, K. Dietz and V. Farewell (eds). pp. 297-331. Boston: Birkhäuser.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

Ying, Z. (1993). A Large Sample Study of Rank Estimation for Censored Regression Data. *Ann. Statist.* **21**, 76-99.

Table 1: Summary of Simulation Studies

			$\beta_* = (0,0)^{\mathrm{T}}$						$\beta_* = (1,1)^{\mathrm{T}}$					
			Z_1			Z_2			Z_1			Z_2		
n	Cens. Per.	Wt.	Bias	Cov. Prob.	Mean $SE(\widehat{\beta})$	Bias	Cov. Prob.	Mean $SE(\widehat{\beta})$	Bias	Cov. Prob.	Mean $SE(\widehat{\beta})$	Bias	Cov. Prob.	Mean $SE(\widehat{\beta})$
100	0%	N	0.010	0.960	0.129	0.018	0.952	0.238	−0.010	0.944	0.388	0.032	0.954	0.633
100	0%	Y	0.015	0.951	0.054	0.013	0.941	0.107	−0.012	0.950	0.155	0.005	0.948	0.272
100	10%	N	0.027	0.951	0.134	−0.022	0.946	0.246	0.023	0.949	0.411	−0.032	0.952	0.759
100	10%	Y	0.014	0.955	0.970	0.013	0.943	0.193	0.018	0.947	0.302	0.037	0.939	0.663
100	30%	N	0.010	0.948	0.168	−0.021	0.953	0.333	−0.058	0.961	0.504	−0.039	0.944	0.817
100	30%	Y	−0.022	0.948	0.163	−0.022	0.946	0.326	0.060	0.948	0.450	0.029	0.970	0.625
200	0%	N	0.004	0.948	0.097	−0.013	0.955	0.146	−0.005	0.956	0.348	−0.007	0.946	0.571
200	0%	Y	−0.015	0.942	0.035	0.037	0.951	0.064	−0.026	0.951	0.138	0.020	0.961	0.161
200	10%	N	−0.037	0.958	0.101	0.024	0.944	0.158	−0.009	0.957	0.399	−0.016	0.952	0.589
200	10%	Y	−0.046	0.958	0.057	−0.017	0.956	0.089	−0.019	0.947	0.287	−0.026	0.964	0.475
200	30%	N	0.033	0.957	0.117	−0.025	0.961	0.183	0.020	0.950	0.427	−0.064	0.944	0.670
200	30%	Y	0.078	0.944	0.068	0.029	0.949	0.147	−0.044	0.948	0.388	0.027	0.953	0.561

Table 2: *Estimates of regression coefficients for the lung cancer trial (Prentice, 1973)*

Parameters	Oakes-Dasu model (1)		AFT† model (2)	Cox model (3)	Pettitt‡ model
	Unweighted	Weighted			
Performance status	0.021(0.008)*	0.030(0.006)	0.022(0.005)	−0.024(0.006)	−0.055(0.010)
Tumour type					
adeno vs large	−0.821(0.549)	−0.801(0.532)	−0.839(0.302)	0.851(0.348)	1.302(0.554)
small vs large	−0.556(0.544)	−0.499(0.522)	−0.521(0.284)	0.548(0.321)	1.438(0.520)
squamous vs large	0.143(0.721)	0.150(0.680)	0.175(0.307)	−0.214(0.347)	−0.177(0.593)

*Estimated standard errors are given in parentheses;

†AFT model is the accelerated failure time model with the log-linear life-testing method;

‡Pettitt model is the proportional odds model with the marginal likelihood method.

Figure 1: Estimated mean life functions from the lung cancer data