Year 2004 *Paper* 154

Linear Life Expectancy Regression with Censored 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 and Biostatistics, University of California, San Francisco, scheng@biostat.ucsf.edu

Linear Life Expectancy Regression with Censored Data

Ying Qing Chen and Su-Chun Cheng

Abstract

Life expectancy, i.e., mean residual life function, has been of important practical and scientific interests to characterise the distribution of residual life. Regression models are often needed to model the association between life expectancy and its covariates. In this article, we consider a linear mean residual life model and further developed some inference procedures in presence of censoring. The new model and proposed inference procedure will be demonstrated by numerical examples and application to the well-known Stanford heart transplant data. Additional semiparametric efficiency calculation and information bound are also considered.

1 Introduction

Suppose that failure time T is nonnegative random variable on a probability space $\{\Omega, \mathscr{F}, \mathcal{P}\}$. Its mean residual life function is defined as $m(t) = E(T - t \mid T > t)$, for $t \geq 0$. When covariates are present, the proportional mean residual life model by Oakes & Dasu (1990) can be used to study the association between $m(t)$ and the covariates, Z, say,

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

where $m(\cdot)$ are mean residual life functions, Z are $p-$vector covariates and β are associated parameters. In its semiparametric version, $m_0(\cdot)$ is unspecified. Compared with the widely used Cox proportional hazards model, it directly models the mean residual life functions and thus has appealing interpretation in terms of life expectancy. Nevertheless, β can be also interpreted in terms of the reciprocal of hazard function projected onto $\mathscr{F}_{\mathrm{T}}(t) = \sigma\{T > t\}$, due to the fact that $m(t) = E\left\{\lambda(T)^{-1} \mid T > t\right\}$, where $\lambda(\cdot)$ denotes the associated hazards function. To estimate β in (1), Maguluri & Zhang (1994) and Oakes & Dasu (2003) studied some estimation procedures when there is no censoring. Recently, Chen & Cheng (2004) developed quasi partial score estimating equations when censoring presents.

Since $m(t) = E(T \mid T > t) - t$, for $t \geq 0$, the shape of $m(t)$ has an embedded constraint, i.e., $m(t) + t$ is monotonically nondecreasing. In the Oakes-Dasu model, however, $m(t \mid Z) + t = m_0(t) \exp(\beta^{\mathrm{T}} Z) + t$ may not satisfy this constraint for an arbitrary $\beta \in \Re^p$, unless $m_0(\cdot)$ itself is monotonically nondecreasing, as pointed out in Oakes & Dasu (1990). A monotonically nondecreasing $m_0(\cdot)$, although plausible mathematically, may not be always consistent with the aging process, for example, of human life. To cope with this constraint, we instead consider a linear mean residual life model,

$$m(t \mid Z) = m_0(t) + \beta^{\mathrm{T}} Z. \tag{2}$$

It is apparent the additive structure in model (2) complies with the embedded constraint. In practice, the parameter β in the linear model can be interpreted as average difference in life expectancy per unit change in covariate. When Z is binary of 0 or 1 as treatment indicator in a randomised clinical trial, for example, β can be considered as treatment effect measure in life expectancy due to different treatment assignments. Furthermore, notice that $E(T \mid Z) = m_0(0) + \beta^{\mathrm{T}} Z$ is implied by the linear model (2). It thus closely relates to the standard linear model in Miller (1976) and Buckley & James (1979), which assume that $T - \gamma_0 - \gamma_1^{\mathrm{T}} Z$ have common zero-mean distribution with $E(T \mid Z) = \gamma_0 + \gamma_1^{\mathrm{T}} Z$.

To apply the new linear model in real applications, it is desirable of $m_0(\cdot)$ to be unspecified without restrictive parametric assumptions. More challenge occurs as the survival outcomes are often censored. The rest of this article aims to developing and studying some inference procedures for the new linear model. The proposed methodologies are demonstrated by numerical examples. Additional semiparametric model efficiency and information bound are also considered.

2 Inference procedures

Suppose that there are n independent subjects in a data set. Let T_i and C_i be the failure and censoring times, respectively, for $i = 1, 2, \ldots, n$. The data set consists of $\{(X_i, \Delta_i, Z_i); i = 1, 2, \ldots, n\}$, where $X_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \le C_i)$ and Z_i are covariates, respectively. Given Z_i, T_i and C_i are assumed independent. Their actual observed values are denoted by corresponding lower cases. Without confusion in notations, subscripts may be occasionally suppressed. Denote $N(t) = I(X \le t, \Delta = 1)$ and $Y(t) = I(X \ge t)$.

2.1 *Estimation of baseline function*

Since the baseline residual life function in model (2) is preferred to be unspecified, we first develop an estimator for $m_0(t)$ as if β is known. Consider the filtration defined by $\mathscr{F}_t = \sigma\{N_i(t), Y_i(t), Z_i, i = 1, 2, \ldots, n\}$. Then $E\{dN(t) \mid \mathscr{F}_{t-}\} = Y(t) d\Lambda(t \mid Z)$, where $\Lambda(\cdot)$ denotes the cumulative hazard function. Applying the inversion formula in Oakes & Dasu

(1990) to the linear model, we know that the survival function of T with covariate Z is,

$$S(t \mid Z) = \frac{m_0(0) + \beta^{\mathrm{T}} Z}{m_0(t) + \beta^{\mathrm{T}} Z} \exp\left\{ - \int_0^t \frac{du}{m_0(u) + \beta^{\mathrm{T}} Z} \right\}.$$

As a result, $\{m_0(t) + \beta^{\mathrm{T}} Z\} d\Lambda(t \mid Z) = \{1 + m_0'(t)\} dt$. Thus, the following equation can be used to estimate $m_0(\cdot)$,

$$\sum_{i=1}^{n} \{m_0(t) + \beta^{\mathrm{T}} z_i\} dN_i(t) = \sum_{i=1}^{n} Y_i(t) \{1 + m_0'(t)\} dt,$$

which is in fact

$$- m_0(t) \frac{\sum_i dN_i(t)}{\sum_i Y_i(t)} + dm_0(t) = \frac{\sum_i \{\beta^{\mathrm{T}} z_i dN_i(t) - Y_i(t) dt\}}{\sum_i Y_i(t)}. \tag{3}$$

Let $\widehat{S}_{\mathrm{NA}}(t) = \exp\{ - \int_0^t \sum_i dN_i(u) / \sum_i Y_i(u)\}$ and $dQ(t; \beta) = Y(t) dt - \beta^{\mathrm{T}} Z dN(t)$, respectively. Here, $\widehat{S}_{\mathrm{NA}}(t)$ would reduce to the usual Nelson-Aalen estimator for the survival function if Z are identical. Then the equation (3) is indeed a first-order ordinary differential equation which yields a closed-form solution of

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

As a result, $d\widehat{m}_0(t) = \widehat{m}_0(t) \sum_i dN_i(t) / \sum_i Y_i(t) - \sum_i \{Y_i(t) - \beta^{\mathrm{T}} z_i dN_i(t)\} / \sum_i Y_i(t)$. When $\beta = \beta_0$ is the true value, $\widehat{m}_0(t)$ is consistent of $m_0(t)$, similar as in Chen & Cheng (2004). Here $\tau = \sup\{t : \mathrm{pr}(X > t) > 0\} < \infty$ to avoid technical discussion on the right-hand tail of censored data, and Δ would be redefined as 1 for the last observation to ensure meaningful $\widehat{m}_0(\cdot)$ by the convention in Reid (1981) and James (1986).

2.2 *Extended Buckley-James estimation*

As noted previously, the new linear model is closely related to the standard linear regression model. We first consider a Buckley-James estimation procedure (Buckley & James, 1979), as it has been demonstrated to be a reliable estimation procedure in linear regression models with censored data (Miller & Halpern, 1982; Lin & Wei, 1992). Let $\alpha(T, Z; \beta) = Z\{T -$

$m_0(0) - \beta^{\mathrm{T}} Z\}$ and $\xi(X, \Delta, Z; \beta) = \Delta\alpha(X, Z; \beta) + (1-\Delta)E\{\alpha(T, Z; \beta) \mid T > X, Z\}$. Then $E(\xi \mid Z) = E(\alpha \mid Z) = 0$, which leads to the unbiased estimating equations of

$$g_1(\beta) = \sum_{i=1}^{n} \xi(x_i, \delta_i, z_i; \beta) = 0, \tag{4}$$

as in the Buckley-James procedure for standard linear regression model. However, both $\alpha(T, Z)$ and $E\{\alpha(T, Z) \mid T > X\}$ depends on the unknown $m_0(\cdot)$ in (4). In the linear regression model, Buckley & James (1979) required that the residuals of $T - m_0(0) - \beta^{\mathrm{T}} Z$ would share an identical distribution, which was estimated by the Kaplan-Meier product-limit estimator. However, this is not necessarily true in the linear mean residual life model. It is thus not appropriate to use the self-consistency representation of the Kaplan-Meier estimator by Efron (1967) to simplify $g_1(\beta)$. Instead we extend the Buckley-James procedure in model (2) with the proposed baseline estimator in the preceding section.

Consider a natural estimator of $\alpha(T, Z; \beta)$ by defining $\widehat{\alpha}(T, Z; \beta) = Z\{T - \widehat{m}_0(0) - \beta^{\mathrm{T}} Z\}$, and $E\{\alpha(T, Z; \beta) \mid T > X, Z\}$ by

$$\widehat{E}\{\alpha(T, Z; \beta) \mid T > X, Z\} = -\widehat{S}(X \mid Z)^{-1} \int_X^{\tau} \widehat{\alpha}(u, Z; \beta) d\widehat{S}(u \mid Z),$$

respectively. Here, $\widehat{S}(t \mid Z) = \{\widehat{m}_0(0) + \beta^{\mathrm{T}} Z\}\{\widehat{m}_0(t) + \beta^{\mathrm{T}} Z\}^{-1} \exp[-\int_0^t \{\widehat{m}_0(u) + \beta^{\mathrm{T}} Z\}^{-1} du]$. Thus the actual estimating functions of β are:

$$\widehat{g}_1(\beta) = \sum_{i=1}^{n} \widehat{\xi}(x_i, \delta_i, z_i; \beta) = \sum_{i=1}^{n} \left[\delta_i \widehat{\alpha}(x_i, z_i; \beta) + (1-\delta_i)\widehat{E}\left\{\alpha(T, z_i; \beta) \mid T > x_i\right\}\right]. \tag{5}$$

Denote $\widehat{\beta}_{\mathrm{BJ}}$ the solution to $\widehat{g}_1(\beta) = 0$. With some straightforward algebra,

$$\frac{\partial \widehat{m}_0(t; \beta_0)}{\partial \beta} = \widehat{S}_{\mathrm{NA}}(t)^{-1} \int_t^{\tau} \widehat{S}_{\mathrm{NA}}(u) \frac{\sum_j Z_j dN_j(u)}{\sum_j Y_j(u)} = -\mu_Z(t) + o_p(1),$$

where $\mu_Z(t) = E\left\{ZS(t \mid Z)\right\}/E\left\{S(t \mid Z)\right\}$. Thus, $\partial\widehat{\alpha}(T, Z; \beta_0)/\partial\beta = -Z\{Z - \mu_Z(0)\}^{\mathrm{T}} + o_p(1)$. Denote $\widehat{\mu}_Z(\cdot)$ be the empirical estimator of $\mu_Z(\cdot)$. Then, $\partial\widehat{S}(t \mid Z; \beta)/\partial\beta$ can be consistently estimated by

$$\widehat{S}(t \mid Z; \beta) Z \left[\frac{Z - \widehat{\mu}_Z(0)}{\widehat{m}_0(0) + \beta^{\mathrm{T}} Z} - \frac{Z - \widehat{\mu}_Z(t)}{\widehat{m}_0(t) + \beta^{\mathrm{T}} Z} - \int_0^t \frac{\{Z - \widehat{\mu}_Z(u)\} du}{\widehat{m}_0(u) + \beta^{\mathrm{T}} Z}\right]^{\mathrm{T}},$$

and $\partial\widehat{E}\{\alpha(T,Z;\beta)\mid T>X,Z\}/\partial\beta$ estimated by

$$\widehat{S}(t\mid Z;\beta)^{-1}\int_X^\tau Z\{Z-\mu_Z(u)\}^{\mathrm{T}}d\widehat{S}(u\mid Z;\beta)-\widehat{S}(t\mid Z;\beta)^{-1}\int_X^\tau \widehat{\alpha}(u,Z;\beta)d\left\{\frac{\partial\widehat{S}(u,Z;\beta)}{\partial\beta}\right\}$$

$$+\widehat{S}(t\mid Z;\beta)^{-1}Z\left[\frac{Z-\widehat{\mu}_Z(0)}{\widehat{m}_0(0)+\beta^{\mathrm{T}}Z}-\frac{Z-\widehat{\mu}_Z(t)}{\widehat{m}_0(t)+\beta^{\mathrm{T}}Z}-\int_0^t\frac{\{Z-\widehat{\mu}_Z(u)\}du}{\widehat{m}_0(u)+\beta^{\mathrm{T}}Z}\right]^{\mathrm{T}}\int_X^\tau\widehat{\alpha}(u,Z;\beta)d\widehat{S}(u\mid Z;\beta).$$

Thus $-n^{-1}\partial\widehat{g}_1(\beta_0)/\partial\beta$ goes to

$$D_1=-E\left[\Delta\partial\widehat{\alpha}(X,Z;\beta_0)/\partial\beta+(1-\Delta)\partial\widehat{E}\left\{\alpha(T,Z;\beta_0)\mid T>X,Z\right\}/\partial\beta\right],$$

as $n\to\infty$. In addition, by a Multivariate Central Limit Theorem, $n^{-1/2}\widehat{g}_1(\beta_0)$ approaches to a zero-mean normal distribution with variance-covariance of $V_1=E\left\{\Delta\widehat{\alpha}(X,Z;\beta_0)^{\otimes 2}\right\}+E[(1-\Delta)\widehat{E}\left\{\alpha(T,Z;\beta_0)\mid T>X,Z\right\}^{\otimes 2}]$, asymptotically. Following an application of Taylor expansion, $\widehat{\beta}_{\mathrm{BJ}}$ has asymptotic normality as

$$n^{1/2}(\widehat{\beta}_{\mathrm{BJ}}-\beta_0)\xrightarrow{\mathscr{D}}N(0,D_1^{-1}V_1D_1^{-1}),$$

given D_1 is nonsingular, as $n\to\infty$. Here D_1 and V_1 can be estimated by their empirical estimators of

$$\widehat{D}_1=n^{-1}\sum_{i=1}^n\left[\delta_i\frac{\partial\widehat{\alpha}(x_i,z_i;\widehat{\beta}_{\mathrm{BJ}})}{\partial\beta}+(1-\delta_i)\frac{\partial\widehat{E}\left\{\alpha(T,z_i;\widehat{\beta}_{\mathrm{BJ}})\mid T>x_i,z_i\right\}}{\partial\beta}\right],\ \text{and}$$

$$\widehat{V}_1=n^{-1}\sum_{i=1}^n\left[\delta_i\widehat{\alpha}(x_i,z_i;\widehat{\beta}_{\mathrm{BJ}})^{\otimes 2}+(1-\delta_i)\widehat{E}\left\{\alpha(T,z_i;\widehat{\beta}_{\mathrm{BJ}})\mid T>x_i,z_i\right\}^{\otimes 2}\right].$$

Apparently, $\widehat{\alpha}(\cdot)$ is an ad hoc choice of the estimating functions, the estimators obtained are thus not necessarily efficient. In fact, one possibly more efficient choice might be

$$\widehat{\alpha}(T,Z;\beta)=\frac{\partial\log f(T\mid Z;\beta)}{\partial\beta}=-Z\left[\frac{1}{\widehat{m}_0(t)+\beta^{\mathrm{T}}Z}-\int_0^t\frac{\{1+\widehat{m}_0'(u)\}du}{\{\widehat{m}_0(u)+\beta^{\mathrm{T}}Z\}^2}\right],\qquad(6)$$

where $f(\cdot)$ is density functions of failure time T. The efficient estimation would be considered more in §2.4 where the semiparametric information bound of model (2) is further studied.

2.3 Quasi partial score estimation

An alternative estimation procedure is by way of constructing estimating functions similar to those of partial score functions of the Cox proportional hazards model. Notice that the estimating functions of

$$g_2(\beta) = \sum_{i=1}^{n} \int_0^{\tau} z_i \left[dN_i(t) - \frac{Y_i(t)\{1 + m_0'(t)dt\}}{m_0(t) + \beta^{\mathrm{T}} z_i} \right]$$

are unbiased at $\beta = \beta_0$. Thus it is natural to use the following estimating functions by plugging in the estimator of $\widehat{m}_0(\cdot)$:

$$\widehat{g}_2(\beta) = \sum_{i=1}^{n} \int_0^{\tau} z_i \left[dN_i(t) - \frac{Y_i(t)\{1 + \widehat{m}_0'(t)dt\}}{\widehat{m}_0(t) + \beta^{\mathrm{T}} z_i} \right].$$

Straightforward algebra shows the above functions are indeed

$$\sum_{i=1}^{n} \int_0^{\tau} \{z_i - \overline{z}(t)\} \left\{ dN_i(t) - \frac{Y_i(t)dt}{\widehat{m}_0(t) + \beta^{\mathrm{T}} z_i} \right\},$$

where $\overline{z}(t) = \sum_i \left[Y_i(t)z_i/\{\widehat{m}_0(t) + \beta^{\mathrm{T}} z_i\}\right] / \sum \left[Y_i(t)/\{\widehat{m}_0(t) + \beta^{\mathrm{T}} z_i\}\right]$. Let $\widehat{\beta}_{\mathrm{QP}}$ be the solution such that $\widehat{g}_2(\widehat{\beta}_{\mathrm{QP}}) = 0$. Then by standard counting processes arguments, $n^{-1/2}\widehat{g}_2(\beta_0)$ converges to a zero-mean normal with the variance-covariance matrix that can be consistently estimated by

$$\widehat{V}_2 = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \frac{Y_i(t)\{z_i - \overline{z}(t)\}^{\otimes 2}\{1 + \widehat{m}_0'(t)\}}{\widehat{m}_0(t) + \widehat{\beta}_{\mathrm{QP}}^{\mathrm{T}} z_i} dt.$$

Furthermore, $-n^{-1}\partial\widehat{g}_2(\beta_0)/\partial\beta$ goes to the matrix that can be consistently estimated by

$$\widehat{D}_2 = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} Y_i(t)\{z_i - \overline{z}(t)\}^{\otimes 2}\{1 + \widehat{m}_0'(t)\}dt.$$

Thus the inference on $\widehat{\beta}_{\mathrm{QP}}$ can be made by the fact that $n^{1/2}(\widehat{\beta}_{\mathrm{QP}} - \beta_0)$ converges weakly to a zero-mean normal with the variance-covariance matrix estimated by $\widehat{D}_2^{-1}\widehat{V}_2\widehat{D}_2^{-1}$, due to a Taylor expansion.

To improve the efficiency for estimators of quasi partial score estimating equations, a common approach is to include weight function as in

$$\sum_{i=1}^{n} \int_0^{\tau} W(t) \{z_i - \overline{z}(t)\} \left\{ dN_i(t) - \frac{Y_i(t)dt}{\widehat{m}_0(t) + \beta^{\mathrm{T}} z_i} \right\} = 0, \tag{7}$$

where $W(\cdot)$ converges to a deterministic function of $w(\cdot)$. Thus the variance-covariance of the estimator obtained by solving (7) can be estimated by $\widehat{D}_{2w}^{-1}\widehat{V}_{2w}\widehat{D}_{2w}^{-1}$, where

$$\widehat{V}_{2w} = n^{-1}\sum_{i=1}^{n}\int_{0}^{\tau}\frac{Y_i(t)W(t)^2\{z_i-\overline{z}(t)\}^{\otimes 2}\{1+\widehat{m}_0'(t)\}}{\widehat{m}_0(t)+\widehat{\beta}_{\mathrm{QP}}^{\mathrm{T}}z_i}dt, \text{ and}$$

$$\widehat{D}_{2w} = n^{-1}\sum_{i=1}^{n}\int_{0}^{\tau}Y_i(t)W(t)^2\{z_i-\overline{z}(t)\}^{\otimes 2}\{1+\widehat{m}_0'(t)\}dt.$$

By applying a Cauchy-Schwarz inequality, the optimal weight function should be then in the form of $1/\{m_0(t)+\beta^{\mathrm{T}}Z\}$ to minimize the variance of the weighted estimator of β.

2.4 Semiparametric information and efficient estimation

Although efficient estimation of model (2) has been conjectured in previous sections, its semiparametric information bound can be alternatively calculated by considering the parametric submodels,

$$m(t \mid Z) = m_0(t) + \theta^{\mathrm{T}}\eta(t) + \beta^{\mathrm{T}}Z, \tag{8}$$

where θ and β are $p-$vector parameters, and $m_0(\cdot)$ and $\eta(t)$ are fixed functions, as in Lai & Ying (1992) and Lin & Ying (1994). Then its associated loglikelihood function of $(\beta^{\mathrm{T}},\theta^{\mathrm{T}})^{\mathrm{T}}$ is

$$\begin{aligned} l(\beta,\theta) &= \sum_{i=1}^{n}\left\{\int_{0}^{\tau}\log\lambda(t\mid Z_i)dN_i(t) - Y_i(t)\lambda(t\mid Z_i)dt\right\} \\ &= \sum_{i=1}^{n}\left[\int_{0}^{\tau}\log\left\{\frac{m_0'(t)+\theta^{\mathrm{T}}\eta'(t)}{m_0(t)+\theta^{\mathrm{T}}\eta(t)+\beta^{\mathrm{T}}Z_i}\right\}dN_i(t) - Y_i(t)\left\{\frac{1+m_0'(t)+\theta^{\mathrm{T}}\eta'(t)}{m_0(t)+\theta^{\mathrm{T}}\eta(t)+\beta^{\mathrm{T}}Z_i}\right\}dt\right]. \end{aligned}$$

Then

$$\left.\frac{\partial l(\beta,\theta)}{\partial\beta}\right|_{\beta=\beta_0,\theta=0} = -\sum_{i=1}^{n}\int_{0}^{\tau}\frac{Z_i}{m_0(t)+\beta_0^{\mathrm{T}}Z_i}\left[dN_i(t)-\frac{Y_i(t)\{1+m_0'(t)\}dt}{m_0(t)+\beta_0^{\mathrm{T}}Z_i}\right]$$

$$\left.\frac{\partial l(\beta,\theta)}{\partial\theta}\right|_{\beta=\beta_0,\theta=0} = \sum_{i=1}^{n}\int_{0}^{\tau}\left\{\frac{\eta'(t)}{1+m_0'(t)}-\frac{\eta(t)}{m_0(t)+\beta_0^{\mathrm{T}}Z_i}\right\}\left[dN_i(t)-\frac{Y_i(t)\{1+m_0'(t)\}dt}{m_0(t)+\beta_0^{\mathrm{T}}Z_i}\right]$$

Consider the Fisher information at β_0 and $\theta = 0$ as a function of fixed η, denoted by matrix

$$I(\eta) = \begin{pmatrix} I_{\beta\beta}(\eta) & I_{\beta\theta}(\eta) \\ I_{\theta\beta}(\eta) & I_{\theta\theta}(\eta) \end{pmatrix},$$

with $I_{\beta\beta} = E(\partial^2 l/\partial\beta^2)$, $I_{\beta\theta} = E(\partial^2 l/\partial\beta\partial\theta)$ and $I_{\beta\beta} = E(\partial^2 l/\partial\theta^2)$, respectively. Then by an application of the Cauchy-Schwarz inequality, the variance-covariance matrix of any regular semiparametric estimator $\widetilde{\beta}$ in the linear model, if $n^{1/2}(\widetilde{\beta} - \beta_0)$ converges to a zero-mean normal, would be larger than $(I_{\beta\beta} - I_{\beta\theta} I_{\theta\theta}^{-1} I_{\beta\theta}^{\mathrm{T}})^{-1}$ for any η. Here matrix A is 'larger' than matrix B if $A - B$ is nonnegative definite. Since

$$I_{\beta\beta}(\eta) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_0^{\tau} E\left[\frac{Y_i(t)\{1 + m_0'(t)\}Z_i^{\otimes 2}}{\{m_0(t) + \beta^{\mathrm{T}} Z_i\}^3}\right] dt,$$

$$I_{\beta\theta}(\eta) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_0^{\tau} E\left[\frac{Y_i(t)\{1 + m_0'(t)\}Z_i}{\{m_0(t) + \beta^{\mathrm{T}} Z_i\}^2} \left\{\frac{\eta'(t)}{1 + m_0'(t)} - \frac{\eta(t)}{m_0(t) + \beta^{\mathrm{T}} Z_i}\right\}^{\mathrm{T}}\right] dt, \text{ and}$$

$$I_{\theta\theta}(\eta) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_0^{\tau} E\left[\frac{Y_i(t)\{1 + m_0'(t)\}}{\{m_0(t) + \beta^{\mathrm{T}} Z_i\}} \left\{\frac{\eta'(t)}{1 + m_0'(t)} - \frac{\eta(t)}{m_0(t) + \beta^{\mathrm{T}} Z_i}\right\}^{\otimes 2}\right] dt,$$

$(I_{\beta\beta} - I_{\beta\theta} I_{\theta\theta}^{-1} I_{\beta\theta}^{\mathrm{T}})^{-1}$ thus reaches its maximum at $\eta(t) = \eta_0(t)$ such that

$$\eta_0'(t) E\left\{\frac{Y(t)}{1 + m_0'(t)}\right\} - \eta_0(t) E\left\{\frac{Y(t)}{m_0(t) + \beta^{\mathrm{T}} Z}\right\} = E\left\{\frac{Y(t) Z}{m_0(t) + \beta^{\mathrm{T}} Z}\right\},$$

which has a closed form solution in $\eta_0(\cdot)$:

$$\eta_0(t) = \overline{P}(t)^{-1} \int_t^{\tau} \overline{P}(u) \overline{Q}(u) du,$$

where

$$\overline{P}(t) = \exp\left[-\int_0^t E\left\{\frac{Y(u)}{m_0(u) + \beta^{\mathrm{T}} Z}\right\} \Big/ E\left\{\frac{Y(u)}{1 + m_0'(u)}\right\} du\right]$$

and $\overline{Q}(t) = E\left[Y(t)Z/\{m_0(u) + \beta^{\mathrm{T}} Z\}\right] / E\left\{Y(t)/\{1 + m_0'(t)\}\right\}$, respectively. Therefore, the semiparametric information bound for β at β_0 is the supremum parametric information bound at β_0 given any choice of $\eta(\cdot)$, which is

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_0^{\tau} E\left[\frac{Y_i(t)\{Z_i - \overline{z}_0(t)\}^{\otimes 2}}{\{m_0(t) + \beta^{\mathrm{T}} Z_i\}^2}\right] dt.$$

Here

$$\overline{z}_0(t) = \lim_{n\to\infty} \overline{Z}(t) = \lim_{n\to\infty} \frac{E\left[\sum_i Y_i(t)Z_i/\{m_0(t)+\beta^{\mathrm{T}}Z_i\}\right]}{E\left[\sum_i Y_i(t)/\{m_0(t)+\beta^{\mathrm{T}}Z_i\}\right]}.$$

Therefore, the optimal estimating function for β in the linear model is

$$g_{\mathrm{opt}}(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \frac{Z_i - \overline{Z}(t)}{m_0(t)+\beta^{\mathrm{T}}Z_i} \left[dN_i(t) - \frac{Y_i(t)\{1+m_0'(t)\}dt}{m_0(t)+\beta^{\mathrm{T}}Z_i}\right].$$

For those subjects with $\Delta_i = 1$, the individual terms reduce to the $\widehat{\alpha}$ conjectured earlier in (6). When $m_0(t)$ is constant, i.e., underlying distributions are exponential, it is seen that the extended Buckley-James estimation procedures is fully efficient. While comparing $g_{\mathrm{opt}}(\cdot)$ with the weighted quasi partial score estimating equations, it is straightforward to see that the optimal weight leads to semiparametric efficient score functions and hence results in the most efficient estimator.

3 Examples

To better understand the proposed linear model in (2), we illustrate it with some special examples when the underlying distribution is of Hall-Wellner class, i.e., the baseline mean residual life function is linear as $m_0(t) = \phi_0 + \phi_1 t$, where ϕ_0 and ϕ_1 are parameters such that $\phi_0 > 0$ and $\phi_1 > -1$. In Fig. 1, $\phi_0 = 1.5$ and $\phi_1 = -0.1$, 0 and 0.1, respectively. Assume that $\beta = 0.5$ for binary Z of 0 and 1 in model $m(t \mid Z) = m_0(t) + \beta Z$. Their mean residual life functions are plotted along with the corresponding hazard functions. The hazard ratios are also plotted. It is interesting to see that the constant additivity in mean residual life functions does not imply constant proportionality in hazard functions, except when the underlying distribution is exponential. The ratio of hazard functions tends decreasing in the graphs of decreasing mean residual life functions, while increasing in those of increasing mean residual life functions. In fact, when $m_0(t) = 1/(1+t)$, the hazard functions would be identical at $t = 0$ under the linear model, as shown in Fig. 2, which apparently the usual Cox proportional hazards model with constant proportionality may not apply. Compared with the proportional hazards model with time-dependent covariates, the linear mean residual life model with time-independent covariates may have advantage in summarizing difference in covariate effect and hence simpler parameter interpretation.

[Figure 1 about here]

We apply the proposed methodology to the well-known Stanford Heart Transplantation data in Miller & Halpern (1982). The time-to-event outcome being considered is the survival time since first heart transplantation between October 1967 and February 1980. Two covariates were originally considered: age at the time of first transplant and T5 mismatch score which measures the degree of tissue incompatability between the initial donor and recipient hearts with respect to HLA antigens. To contrast with their results, we consider the linear mean residual life model for the base 10 log-transformed survival time against age and T5 mismatch scores, as in Miller & Halpern (1982) and Lin & Wei (1992). The results are tabulated in Table 1. along with those from the partial likelihood estimation of the Cox proportional hazards model and the Buckley-James estimation of the linear regression models. As shown in the table, all the estimates for age and T5 mismatch scores are negatively associated with life expectancy, although none of them is significant for the T5 mismatch scores. That is, the covariate age is not only significant predictor for the patients' hazard and their average survival lifetimes, but also for their life expectancy throughout time. For different estimation procedures of the same linear mean residual life model, it is not surprising to see that the efficient estimation procedure yields the smallest variance.

[Table 1 about here]

In addition, as demonstrated in Miller & Halpern (1982), a quadratic pattern in the covariate of age might appear for both the Cox proportional hazards model and the linear regression models. Thus we fitted the linear mean residual life model as well to compare with their results, with the T5 mismatch scores omitted due to their insignificance shown in Table 1. As shown in Table 2, both age and the squared age are significant predictors of the life expectancy. However, their impact are of different directions: the age itself positively predicts the life expectancy while the quadratic age negatively predicts the life expectancy, which demonstrates similar patterns as those shown by the Cox model and the linear regression model. Again, the efficient estimation procedure yields smaller variances compared with other ad hoc procedures.

[Table 2 about here]

4 Discussion

There are some fundamental challenges to develop inference procedures for any statistical methods of estimation, hypothesis testing or regression based on mean residual life functions in presence of censoring. One of the challenges comes from the tail behaviour of underlying distributions of failure times. In reality, the underlying failure times may be heavily right skewed and early censored, such as for long-term survivors on cancer treatment or subjects in HIV/AIDS prevention/vaccine trials, it is impossible to estimate the mean residual life function on the whole positive real line without extra assumptions, although some techniques such as in Koul, Susarla & Van Ryzin (1981), Gill (1983) and Ying (1993) can be extended. In general, it is difficult to determine a reasonable upper limit τ without the robustness of the proposed methodologies being compromised.

To deal with the situation, two possible approaches may be adapted. The first approach is modify the fully unspecified $m_0(\cdot)$ by including parametric component in the tail. For instance, let $\widetilde{\tau}$ be a prespecified truncation time. Then it is assumed that

$$\widetilde{m}_0(t) = m_0(t)I(t < \widetilde{\tau}) + m_r I(t > \widetilde{\tau}),$$

where m_r is some positive constant. This means, the baseline mean residual life function is unspecified up to the truncation time $\widetilde{\tau}$, while it becomes exponential after $\widetilde{\tau}$. Then it is straightforward to extend the proposed methodologies to the whole positive real line. The second approach is by way of the cure mixture model. That is, assume that the failure times are mixture of two subpopulations: relatively short-term survivors and relatively long-term survivors, denoted by $\rho = 1$ and 0, respectively. As in Lu & Ying (2004), a failure time $\widetilde{T}$ is assumed to be

$$\widetilde{T} = \rho T + (1-\rho)\infty,$$

in notation. Here the $\text{supp}\{F_{\text{T}}(t)\}$ is finite and T follows the linear model (2). In fact, as in Farewell (1982), the probability of $\rho = 1$ can be further modelled by the generalized linear models such as the logistic model.

As in the additive hazards model by Lin & Ying (1994), the linear combination form of $\beta^{\text{T}} Z$ is chosen mainly for easy interpretation and simple inference procedures. Compared with the Oakes-Dasu proportional model, it is not restricted to maintain the derived monotonicity in baseline functions. However, it does have constraint such that the modelled mean residual life function to be nonnegative. One solution is to replace the linear form with its exponentiated term, as suggested in Lin & Ying (1994). In fact, there are quite a few modelling routines for the hazard functions can be similarly adapted to the proposed linear model in this article. For instances, we can consider the following linear model,

$$m(t \mid Z) = \sum_{j=1}^{p} Z_j(t) m_{0j}(t),$$

as in Aalen (1980); or the additive-multiplicative model,

$$m(t \mid Z) = m_0(t) h_1(\beta_1^{\text{T}} Z_1) + h_2(\beta_2^{\text{T}} Z_2),$$

as in Lin & Ying (1995).

References

Buckley, J. & James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429-36.

CHEN, Y. Q. & CHENG, S. (2004). Semiparametric regression analysis of mean residual life with censored survival data. *Biometrika*, in press.

EFRON, B. (1967). The two sample problem with censored data. *Proc. 5th Berkeley Symp.* **4**, 831-53.

FAREWELL, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38**, 1041-6.

GILL, R. D. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11**, 49-58.

JAMES, I. R. (1986). On estimating equations with censored data, *Biometrika* **73**, 35-42.

KOUL, H., SUSARLA, V. & VAN RYZIN, J. (1981). Regression with randomly right-censored data, *Ann. Statist.* **9**, 1276-1288.

LAI, T. L. & YING, Z. (1992). Asymptotic efficiency estimation in censored and truncated regression models. *Statist. Sinica* **2**, 17-46.

LIN, D. Y. & YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.

LIN, D. Y. & YING, Z. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *Ann. Statist.* **5**, 1712-34.

LIN, J. S. & WEI, L.-J. (1992). Linear regression analysis based on Buckley-James estimating equations. *Biometrics* **48**, 679-681.

LU, W. & YING, Z. (2004). On semiparametric transformation cure models. *Biometrika* **91**, 331-43.

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

MILLER, R. G. (1976). Least squares regression with censored data. *Biometrika* **63**, 449-64.

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*, Ed. J. E. Kolassa, and D. Oakes, Institute of Mathematical Statistics Lecture Notes Monograph Series, Vol. **43**, 105-16.

REID, N. (1981). Influence functions for censored data. *Ann. Statist.* **9**, 78-92.

SUSARLA, V. & VAN RYZIN, J. (1980). Large sample theory for an estimator of the mean survival time from censored samples. *Ann. Statist.* **8**, 1002-16.

YING, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76-99.

Last revised: August 1, 2004

Table 1: *Regression estimates with standard errors for 157 Stanford heart transplant patients*

Coefficients		Age		T5 Score	
		$\widehat{\beta}_1$	s.d.	$\widehat{\beta}_2$	s.d.
Cox model		0.0294	0.0115	0.0481	0.0330
Linear regression model		-0.0161	0.0083	-0.0294	0.0343
Linear mean residual life model					
	Buckley-James	-0.0273	0.0138	-0.0231	0.0427
	Quasi partial score	-0.0282	0.0137	-0.0252	0.0401
	Efficient estimation	-0.0252	0.0118	-0.0256	0.0354

Cox model, estimates based on partial likelihood; Linear regression model, estimates based on Buckley-James procedure

Table 2: *Regression estimates with standard errors for 152 Stanford heart transplant patients who survived at least 10 days*

Coefficients		Age		Age^2	
		$\widehat{\beta}_1$	s.d.	$\widehat{\beta}_2$	s.d.
Cox model		-0.1457	0.0554	0.0023	0.0007
Linear regression model		0.1083	0.0417	-0.0017	0.0005
Linear mean residual life model					
	Buckley-James	0.1728	0.0662	-0.0027	0.0012
	Quasi partial score	0.1725	0.0686	-0.0028	0.0012
	Efficient estimation	0.1771	0.0547	-0.0023	0.0010

Cox model, estimates based on partial likelihood; Linear regression model, estimates based on Buckley-James procedure

Figure 1: Mean residual life functions, their corresponding hazard functions and hazard ratios, when $m_0(t) = \phi_0 + \phi_1 t$. The graphs in each row are for decreasing, constant and increasing mean residual life functions, respectively. Solid lines are for $Z = 0$ and dashed lines for $Z = 1$, respectively.

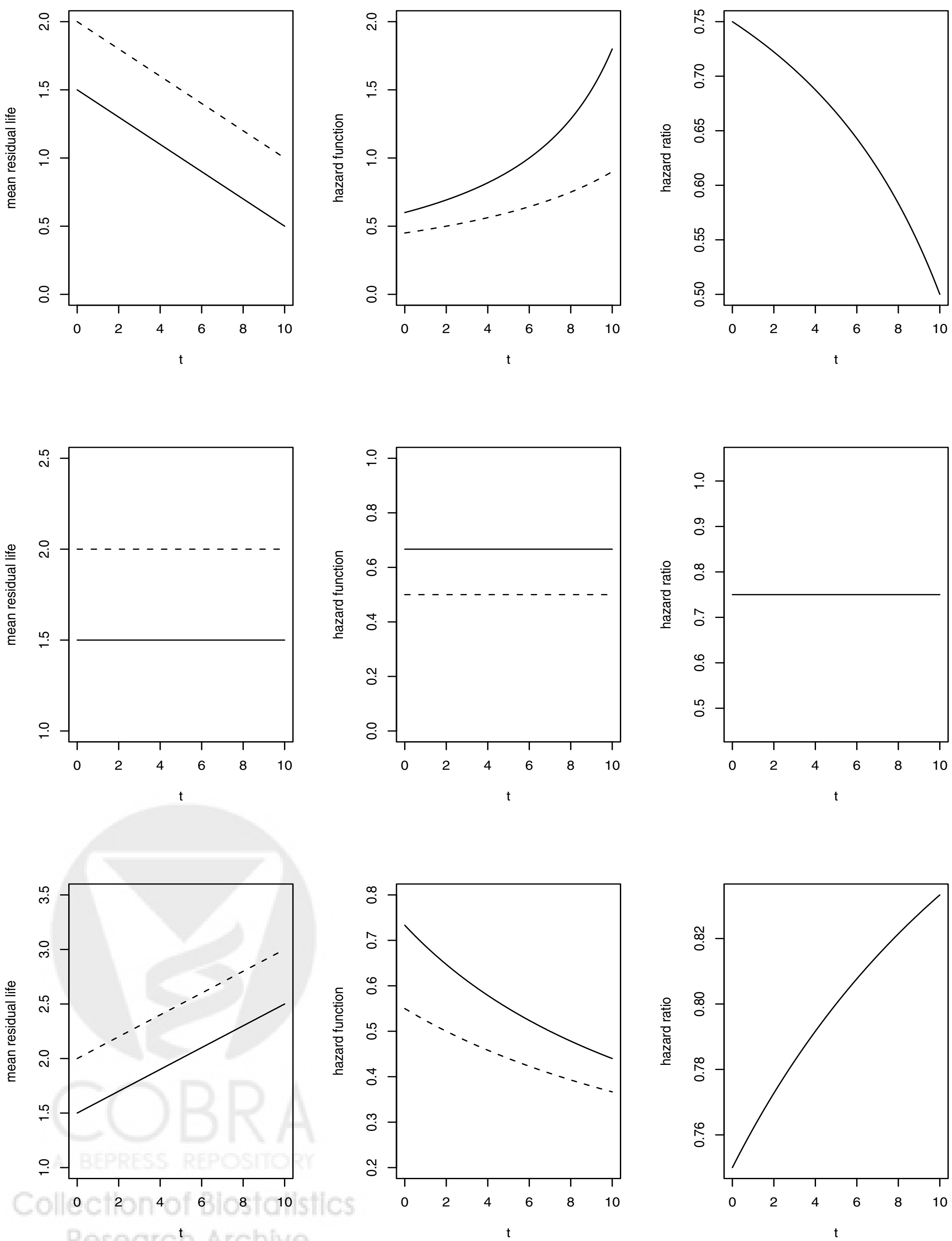

Figure 2: Mean residual life functions and their corresponding hazard functions, when $m_0(t) = 1/(1+t)$. Solid lines are for $Z = 0$ and dashed lines for $Z = 1$, respectively.