Linear life expectancy regression with censored data

BY Y. Q. CHEN
Program in Biostatistics, Division of Public Health Sciences,
Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, U.S.A.
yqchen@scharp.org

AND S. CHENG
Department of Epidemiology & Biostatistics,
University of California, San Francisco, California 94143, U.S.A.
scheng@biostat.ucsf.edu

Summary

In the statistical literature, the life expectancy is usually characterised by the mean residual life function. Regression models are thus needed to study the association between the mean residual life functions and their covariates. In this article, we consider a linear mean residual life model and further developed inference procedures in the presence of potential censoring. The new model and inference procedures are applied to the Stanford heart transplant data. Additional semiparametric efficiency calculations and information bound are also considered.

Some key words: Additive model; Counting process; Estimating equation; Mean residual life; Semiparametric model
Suppose that failure time $T$ is a nonnegative random variable on a probability space $\{\Omega, \mathcal{F}, \mathcal{P}\}$. In the statistical literature, the life expectancy of $T$ at time $t > 0$, $e(t)$ say, is usually characterised and studied by way of its mean residual life function, $m(t) = E(T - t \mid T > t) = e(t) - t$.

When covariates are present, the proportional mean residual life model of Oakes & Dasu (1990) can be used to study the association between $m(t)$ and the covariates:

$$m(t \mid Z) = m_0(t) \exp(\beta^T Z),$$

where $Z$ are the $p$-vector covariates and $\beta$ are the associated parameters. To estimate $\beta$ in (1), Maguluri & Zhang (1994) and Oakes & Dasu (2003) studied some estimation procedures when there is no censoring. Recently, Chen & Cheng (2005) developed quasi partial score estimating equations when censoring is present and $m_0(\cdot)$ is unspecified.

As a result of the nature of $e(t)$, the shape of $m(t)$ has an intrinsic constraint, in that $e(t) = m(t) + t$ should be monotonically nondecreasing in $t \geq 0$. In the proportional mean residual life model, however, $m(t \mid Z) + t = m_0(t) \exp(\beta^T Z) + t$ may not always satisfy this constraint for certain $\beta \in \mathbb{R}^p$, 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 compatible with the actual underlying process, in the case of human ageing, for example. To cope with this constraint, we consider instead a linear mean residual life model,

$$m(t \mid Z) = m_0(t) + \beta^T Z.$$  

It is apparent the additive structure in model (2) complies with the intrinsic constraint. Moreover, model (2) is equivalent to $e(t \mid Z) = e_0(t) + \beta^T Z$. Thus, in practice, the parameter $\beta$ in model (2) can be interpreted as the average difference in life expectancies per unit change in $Z$. For example, when $Z = 0/1$ is the treatment indicator in a randomised clinical trial, $\beta$ can be considered as the treatment effect measured in terms of life expectancy. In general, $m_0(t) + \beta^T Z$ is required to be nonnegative.
To apply the new linear model in real applications, it is desirable that $m_0(\cdot)$ be unspecified without restrictive parametric assumptions, and the fact that the survival outcomes are often censored has to be accommodated. The rest of this article aims to develop and study some inference procedures for the new linear model.

2 Estimation and Inference

2.1 Estimation of baseline function

Suppose that there are $n$ subjects in the dataset, and let $T_i$ and $C_i$ be the failure and censoring times, respectively, for $i = 1, 2, \ldots, n$. The observed dataset thus consists of $n$ independent identically distributed copies of $\{(X_i, \Delta_i, Z_i); i = 1, 2, \ldots, n\}$, where $X_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$ and $Z_i$ are the covariates, respectively. Given $Z_i$, $T_i$ and $C_i$ are assumed independent. Without confusion in notation, subscripts may be occasionally suppressed. Let $N(t) = I(X \leq t, \Delta = 1)$ and $Y(t) = I(X \geq t)$.

Since the baseline residual life function in model (2) is unspecified, we first develop an estimator for $m_0(t)$ as if the true $\beta$ were known. Let the true values of $\beta$ and $m_0(\cdot)$ be $\beta^*$ and $m^*(\cdot)$, respectively; and consider the filtration defined by $\mathcal{F}_t = \sigma\{N_i(u), Y_i(u), Z_i; 0 \leq u \leq t, i = 1, 2, \ldots, n\}$. Then $E\{dN(t) \mid \mathcal{F}_t; \beta^*, m^*\} = Y(t)d\Lambda(t \mid Z; \beta^*, m^*)$, where $\Lambda(\cdot)$ denotes the cumulative hazard function. Let $dM_i(t; \beta^*, m^*) = dN_i(t) - Y_i(t)d\Lambda(t \mid Z_i; \beta^*, m^*)$, $i = 1, 2, \ldots, n$, where $\{M_i(t; \beta^*, m^*)\}$ are martingales with respect to $\mathcal{F}_t$.

Applying the inversion formula in Oakes & Dasu (1990) to the linear model, we know that the survival function of $Z$ is

$$S(t \mid Z; \beta^*, m^*) = \frac{m^*_s(0) + \beta^*_s Z}{m_s(t) + \beta^*_s Z} \exp\left\{-\int_0^t \frac{du}{m_s(u) + \beta^*_s Z}\right\}.$$ 

As a result, $\{m_s(t) + \beta^*_s Z\}d\Lambda(t \mid Z; \beta^*, m^*) = d\{t + m_s(t)\}$. Thus, it is reasonable to estimate $m_0(\cdot)$ by the estimating equation,

$$\sum_{i=1}^n \{\hat{m}_0(t) + \beta^*_s Z_i\}dN_i(t) - \sum_{i=1}^n Y_i(t)d\{t + \hat{m}_0(t)\} = 0, \quad (3)$$

3
which is in fact \(-\hat{m}_0(t)\sum_i dN_i(t)/\sum_i Y_i(t) + d\hat{m}_0(t) = \sum_i \{\beta^T Z_i dN_i(t) - Y_i(t) dt\}/\sum_i Y_i(t)\). Then this first-order ordinary differential equation yields a closed-form solution of

\[
\hat{m}_0(t; \beta_*) = \hat{S}(t)^{-1} \int_t^\tau \hat{S}(u) \frac{\sum_i \{Y_i(u) du - \beta^T Z_i dN_i(u)\}}{\sum_i Y_i(u)},
\]

where \(\hat{S}(t) = \exp\{-\int_0^t \sum_i dN_i(u)/\sum_i Y_i(u)\}\). Here \(\tau = \inf\{t : \text{pr}(X > t) = 0\} < \infty\) to avoid technical discussion about the tail behaviour of censored data. By the convention of Reid (1981) and James (1986), \(\Delta\) would be redefined as 1 for the last observation to ensure a meaningful \(\hat{m}_0(\cdot)\).

Since \(\sum_i \{m_*(t) + \beta^T Z_i \} dN_i(t) - \sum_i Y_i(t) d\{t + m_*(t)\} = \sum_i \{m_*(t) + \beta^T Z_i \} dM_i(t; \beta_*, m_*)\), by subtracting it from (3), we obtain that \(\sum_i \{\hat{m}_0(t; \beta_*) - m_*(t)\} dN_i(t) + \sum_i Y_i(t) d\{\hat{m}_0(t; \beta_*) - m_*(t)\} = -\sum_i \{m_*(t) + \beta^T Z_i \} dM_i(t; \beta_*, m_*)\). Thus, this leads to

\[
\hat{m}_0(t; \beta_*) - m_*(t) = -\hat{S}(t)^{-1} \int_t^\tau \hat{S}(u) \frac{\sum_i \{m_*(u) + \beta^T Z_i \} dM_i(u; \beta_*, m_*)}{\sum_i Y_i(u)},
\]

and hence the consistency and asymptotic normality of \(\hat{m}_0(t; \beta_*)\) hold according to the standard martingale theory for counting processes. In addition, if we let \(\hat{\mu}_*(t; \beta) = -\partial \hat{m}_0(t; \beta)/\partial \beta\), it follows that \(\hat{\mu}_*(t; \beta_*) = \hat{S}(t)^{-1} \int_t^\tau \hat{S}(u) \sum_j Z_j dN_j(u)/\sum_j Y_j(u) = \mu_*(t) + o_p(1)\), where \(\mu_*(t) = E\{Z S_* (t | Z)\} / E\{S_* (t | Z)\}\) and \(S_*(t | Z) = \text{pr}\{X \geq t | Z\}\), respectively.

2.2 Extended Buckley-James estimation

The linear mean residual life model (2) implies that \(E(T | Z) = m_0(0) + \beta^T Z\). It is thus closely related to the standard linear model in Miller (1976) and Buckley & James (1979), which assumes that \(T - \gamma_0 - \gamma_1 Z\) has the identical zero-mean distribution with \(E(T | Z) = \gamma_0 + \gamma_1^T Z\). Since the Buckley-James estimation procedure (Buckley & James, 1979) has been shown to be a reliable estimation procedure in the linear regression models with censored data (Miller & Halpern, 1982; Lin & Wei, 1992), we consider an extension of this procedure to the estimation of the linear mean residual life model.
Let \( \alpha(T, Z; \beta) = Z\{T - m_0(0) - \beta^T Z\} \), and let that \( \xi(X, \Delta, Z; \beta) = \Delta \alpha(X, Z; \beta) + (1 - \Delta)E\{\alpha(T, Z; \beta) \mid T > X, Z\} \). Then \( E(\xi \mid Z) = E(\alpha \mid Z) = 0 \), which leads to the unbiased estimating equations

\[
g_1(\beta) = n^{-1/2} \sum_{i=1}^{n} \xi(X_i, \Delta_i, Z_i; \beta) = 0, \tag{5}
\]
as in the Buckley-James procedure for the standard linear regression model. However, both \( \alpha(T, Z; \beta) \) and \( E\{\alpha(T, Z; \beta) \mid T > X, Z\} \) depend on the unknown \( m_0(\cdot) \), in addition to \( \beta \). In the linear regression model, Buckley & James (1979) required that the residuals of \( T - m_0(0) - \beta^T Z \) share the same distribution, which was then estimated by the Kaplan-Meier product-limit estimator. However, this is usually not satisfied in the linear mean residual life model. It is thus less straightforward to use Efron’s (1967) self-consistency representation of the Kaplan-Meier estimator to simplify \( g_1(\beta) \). Instead, we extend the Buckley-James procedure to the model (2) using the proposed baseline estimator in §2.1.

Consider the natural estimator of \( \alpha(T, Z; \beta) \) defined by \( \hat{\alpha}(T, Z; \beta) = Z\{T - \hat{m}_0(0) - \beta^T Z\} \). In addition, by the inversion formula, \( S(t \mid Z; \beta) \) is estimated by \( \hat{S}(t \mid Z; \beta) = \{\hat{m}_0(0) + \beta^T Z\}^{-1}\{\hat{m}_0(t) + \beta^T Z\}^{-1}\exp[- \int_{t}^{1} \{\hat{m}_0(u) + \beta^T Z\}^{-1} du] \). Therefore, \( E\{\alpha(T, Z; \beta) \mid T > X, Z\} \) can be estimated by \( \hat{E}\{\alpha(T, Z; \beta) \mid T > X, Z\} = -\hat{S}(X \mid Z; \beta)^{-1} \int_{X}^{T} \hat{\alpha}(u, Z; \beta) d\hat{S}(u \mid Z; \beta) \).

By substituting \( \alpha(T, Z; \beta) \) and \( E\{\alpha(T, Z; \beta) \mid T > X, Z\} \) by these estimators in \( g_1(\beta) \), we obtain the final estimating equations

\[
\hat{g}_1(\beta) = n^{-1/2} \sum_{i=1}^{n} \left[ \Delta_i \hat{\alpha}(X_i, Z_i; \beta) + (1 - \Delta_i) \hat{E}\{\alpha(T, Z_i; \beta) \mid T > X_i\} \right] = 0. \tag{6}
\]

Let \( D_1(\beta) = \lim_n -n^{-1/2} \partial \hat{g}_1(\beta)/\partial \beta \). When \( \beta = \beta_* \), \( D_1(\beta) = -E[\Delta \partial \hat{\alpha}(X, Z; \beta_*)/\partial \beta + (1 - \Delta) \partial \hat{E}\{\alpha(T, Z; \beta_*) \mid T > X, Z\}/\partial \beta] \). In addition, by a Multivariate Central Limit Theorem, \( n^{-1/2}\hat{g}_1(\beta_*) \) approaches a zero-mean normal distribution with covariance matrix \( V_1 = E\{\Delta \hat{\alpha}(X, Z; \beta_*)^{\otimes 2}\} + E[\{(1 - \Delta) \hat{E}\{\alpha(T, Z; \beta_*) \mid T > X, Z\}\}^{\otimes 2}] \), asymptotically. Denote by \( \hat{\beta}_{n1} \) the solution to \( \hat{g}_1(\beta) = 0 \). Following an application of Taylor expansion, \( \hat{\beta}_{n1} \) is asymptotically normal:

\[
n^{1/2}(\hat{\beta}_{n1} - \beta_*) \rightarrow N(0, D_1^{-1} V_1 D_1^{-1}),
\]
in distribution, as \( n \to \infty \), provided that \( D_1 \) is nonsingular. Here \( V_1 \) and \( D_1 \) can be estimated by their empirical estimators, \( \hat{V}_1 = n^{-1} \sum_i [\Delta_i \hat{\alpha}(X_i, Z_i; \hat{\beta}_0)]^{\otimes 2} + (1 - \Delta_i)\bar{E}\{\alpha(T, Z_i; \hat{\beta}_0) \mid T > X_i, Z_i\}^{\otimes 2} \) and \( \hat{D}_1 = n^{-1} \sum_i [\Delta_i \partial \hat{\alpha}(X_i, Z_i; \hat{\beta}_0)/\partial \beta + (1 - \Delta_i)\partial \bar{E}\{\alpha(T, Z_i; \hat{\beta}_0) \mid T > X_i, Z_i\}/\partial \beta] \).

Here, \( \hat{\alpha}(\cdot) \) is an ad hoc choice for the estimating functions, and therefore the estimator \( \hat{\beta}_0 \) is not necessarily efficient. In fact, in the original Buckley-James estimation, when \( \epsilon \) is normal with known variance in the standard linear regression model of \( Y = \gamma_0 + \gamma_1^T Z + \epsilon \), \( \alpha(T, Z; \gamma) = Z(T - \gamma_0 - \gamma_1^T Z) \) is the score function of the regression parameter \( \gamma_1 \). Its efficient use for distributions other than normal is however unclear. Later, in James (1986), \( \alpha(T, Z; \gamma) \) was naturally generalised to be \( \partial \log f(T \mid Z; \gamma)/\partial \gamma \), where \( f(\cdot) \) is the known density function of \( T \). Similarly for the current linear mean residual life model (2), one possibly more efficient choice for \( \alpha \) is

\[
\tilde{\alpha}(T, Z; \beta) = \frac{\partial \log f(T \mid Z; \beta)}{\partial \beta} = -Z \left[ \frac{1}{m_0(T) + \beta^T Z} - \int_0^T \frac{\{1 + m_0'(u)\}du}{\{m_0(u) + \beta^T Z\}^2} \right],
\]

instead of \( \alpha(T, Z; \beta) = Z(T - m_0(0) - \beta^T Z) \). In particular, when the baseline mean residual life function is constant of \( m_0(t) \equiv \alpha_0 \), then \( \tilde{\alpha}(T, Z; \beta) = Z(\alpha_0 + \beta Z)^{-2}(T - \alpha_0 - \beta Z) = Zm(t \mid Z)^{-2}(T - \alpha_0 - \beta Z) \). More discussion of efficient estimation can be found in §2.4.

### 2.3 Quasi partial score estimation

An alternative approach is to construct similar quasi partial score equations to those in Chen & Cheng (2005). Note that the estimating functions of \( g_2(\beta^*_s; m_*) = n^{-1/2} \sum_{i=1}^n \int_0^T Z_i\{m_*(t) + \beta^*_s Z_i\}dN_i(t) - Y_i(t)\{1 + m'_*(t)dt\} \) are unbiased. It is therefore natural to replace \( m_*(\cdot) \) with its estimator, \( \hat{m}_0(\cdot; \beta) \), and use \( \hat{g}_2\{\beta; \hat{m}_0(\beta)\} = 0 \) to solve for \( \beta \) to provide an estimator. Since \( d\{t + \hat{m}_0(t; \beta)\} = \sum_{i=1}^n \{\hat{m}_0(t; \beta) + \beta^T Z_i\}dN_i(t)/\sum_i Y_i(t) \), as in (3), \( \hat{g}_2\{\beta; \hat{m}_0(\beta)\} = 0 \) is simplified as

\[
n^{-1/2} \sum_{i=1}^n \int_0^T \{Z_i - \overline{Z}(t)\} \{\hat{m}_0(t; \beta) + \beta^T Z_i\}dN_i(t) = 0. \tag{8}
\]

Here, \( \overline{Z}(t) = \sum_i Y_i(t)Z_i/\sum_i Y_i(t) \) has the same limit as \( \mu_*(t) \). Therefore, \( -n^{-1/2}\partial \hat{g}_2(\beta_*)/\partial \beta \) tends to \( D_2 = E[\int_0^T \{Z - \mu_*(t)\}^{\otimes 2}dN(t)] \) almost surely, as \( n \to \infty \). Consider a decomposition
of \( \hat{g}_2{\beta_s; \hat{m}_0(\beta_s)} \) of the form \( \hat{g}_2{\beta_s; \hat{m}_0(\beta_s)} = [\hat{g}_2{\beta_s; \hat{m}_0(\beta_s)} - \hat{g}_2(\beta; m_*]) + \hat{g}_2(\beta; m_*) \), as in Tsiatis (1991). By the martingale representation of \( \hat{m}(t; \beta) - m_*(t) \) in (4), the first term in this decomposition equals

\[
-n^{-1/2} \sum_{i=1}^{n} \int_0^T \frac{Z_i - \bar{Z}(t)}{S(t)} \left[ \sum_{j=1}^{n} \int_t^T \frac{\hat{S}(u)}{\sum_i Y_i(u)} \{m_*(u) + \beta^T Z_j \} dM_j(u) \right] dN_i(t). \tag{9}
\]

Let \( \bar{Z}(t) = \hat{S}(t) \int_0^T \hat{S}(u)^{-1} \sum_j \{Z_j - \bar{Z}(u)\} dN_j(u)/\sum_j Y_j(t) \). Then (9) can be further written as

\[
-n^{-1/2} \sum_i \int_0^T \bar{Z}(t) \{m_*(t) + \beta^T Z_i \} dM_i(t).
\]

Since the second term of \( \hat{g}_2(\beta; m_*) \) in the decomposition equals \( n^{-1/2} \sum_i \int_0^T \{Z_i - \bar{Z}(t)\} \{m_*(t) + \beta^T Z_i \} dM_i(t) \), we obtain that

\[
\hat{g}_2(\beta_s) = \hat{g}_2{\beta_s; \hat{m}_0(\beta_s)} = n^{-1/2} \sum_i \int_0^T \{Z_i - \bar{Z}(t) - \bar{Z}(t)\} \{m_*(t) + \beta^T Z_i \} dM_i(t).
\]

As a result, \( \hat{g}_2(\beta_s) \) tends in distribution to a zero-mean normal distribution with the asymptotic variance \( V_2(\beta_s) = \text{Var}[\int_0^T Y_1(t) \{Z - \mu_*(t) - \tilde{\mu}_*(t)\}^\otimes \{1 + m'_*(t)\} \{m_*(t) + \beta^T Z \} dt] \), where \( \tilde{\mu}_*(t) = \lim_n \bar{Z}(t) \), as \( n \to \infty \). Let \( \tilde{\beta}_{qP} \) be the solution to \( \hat{g}_2(\beta) = 0 \). By Taylor expansion,

\[
n^{1/2}(\tilde{\beta}_{qP} - \beta_s) = \{-n^{-1/2} \partial \hat{g}_2(\beta_s)/\partial \beta\}^{-1} \cdot \hat{g}_2(\beta_s) + o_p(1).
\]

Therefore,

\[
n^{1/2}(\tilde{\beta}_{qP} - \beta_s) \to N(0, D_2^{-1} V_2 D_2^{-1}),
\]

in distribution, where \( D_2 \) and \( V_2 \) can be consistently estimated by \( \hat{D}_2 = n^{-1} \sum_i \int_0^T \{Z_i - \bar{Z}(t)\}^\otimes dN_i(t) \), and \( \hat{V}_2 = n^{-1} \sum_i \int_0^T Y_i(t) \{Z_i - \bar{Z}(t) - \bar{Z}(t)\}^\otimes \{1 + \tilde{m}_0(t)\} \{\tilde{m}_0(t) + \tilde{\beta}_{qP}^T Z_i\} dt \).

In addition, user-defined weight functions can be also included in the proposed quasi partial estimating functions, as in \( n^{-1/2} \sum_i \int_0^T W(t) \{Z_i - \bar{Z}(t)\}^\otimes \{\tilde{m}_0(t; \beta) + \beta^T Z_i\} dN_i(t) = 0 \), where the \( F_t \)-predictable \( W(\cdot) \) converges to a deterministic function of \( w(\cdot) \). Then the asymptotic properties of its solution, \( \hat{\beta}_w \), say, can be similarly derived.

### 2.4 Efficiency considerations

Given the ad hoc nature of the estimating equations for the proposed \( \hat{\beta}_{u1} \) and \( \hat{\beta}_{qP} \), more efficient estimation procedures may be of interest. To gain some insight into the efficient estimation of the semiparametric linear mean residual life model, we follow the procedure in Lai & Ying (1992) and Lin & Ying (1994) in calculating a semiparametric information
bound for this model. Let $\beta_0 \in \mathbb{R}^p$ be a parameter and let $\eta(\cdot)$ be a fixed function. Consider the parametric submodels of (2), defined by

$$m(t \mid Z) = m_0(t) + \beta_s^T Z_s(t),$$  \tag{10}$$

where $\beta_s = (\beta_0^T, \beta^T)^T$ and $Z_s(\cdot) = (\eta(\cdot)^T, Z^T)^T$, respectively. Then the associated loglikelihood function of $\beta_s$ is $l(\beta_s) = \sum_i \int_0^T \log \left\{ \frac{m_0(t) + \theta^T \eta(t)}{m_0(t) + \beta_s^T Z_s(t)} \right\} dN_i(t) - Y_i(t) \frac{1 + m_0(t) + \theta^T \eta(t)}{m_0(t) + \beta_s^T Z_s(t)} dN_i(t)$. Therefore, the score functions at $\beta_{ss} = (0, \beta^T)^T$ are $l_\beta^0(\beta_{ss}) = \partial l(\beta_s)/\partial \beta|_{\beta_s = \beta_{ss}} = -\sum_i \int_0^T Z_i / \{m_0(t) + \beta_s^T Z_s(t)\} dN_i(t) - Y_i(t) \{1 + m_0(t)\}/\{m_0(t) + \beta_s^T Z_s(t)\} dt$ and $l_\beta^0(\beta_{ss}) = \partial l(\beta_s)/\partial \beta_0|_{\beta_s = \beta_{ss}} = \sum_i \int_0^T \theta(t) / \{1 + m_0(t) - \eta(t)\}/\{m_0(t) + \beta_s^T Z_s(t)\} dN_i(t) - Y_i(t) \{1 + m_0(t)\}/\{m_0(t) + \beta_s^T Z_s(t)\} dt$, respectively. We therefore obtain that, in distribution,

$$n^{-1/2} \begin{bmatrix} l_\beta^0(\beta_{ss}) \\ l_\beta^0(\beta_{ss}) \end{bmatrix} \rightarrow N \{0, I(\eta)\}, \text{ where } I(\eta) = \begin{bmatrix} I_{\beta \beta}(\eta) & I_{\beta \beta_0}(\eta) \\ I_{\beta_0 \beta}(\eta) & I_{\beta_0 \beta_0}(\eta) \end{bmatrix},$$

under proper regularity conditions similar to those in Lai & Ying (1992), where $I_{\beta \beta}(\eta) = E\{l_\beta^0(\beta_{ss})\}$, $I_{\beta \beta_0}(\eta) = I_{\beta_0 \beta}(\eta)^T = E\{l_\beta^0(\beta_{ss})\}$ and $I_{\beta_0 \beta_0}(\eta) = E\{l_\beta^0(\beta_{ss})\}$. Actual calculation leads to that

$$I_{\beta \beta}(\eta) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^n \int_0^T E \left[ \frac{Y_i(t) \{1 + m_0(t)\} Z_i^2}{\left\{m_0(t) + \beta_s^T Z_s(t)\right\}^3} \right] dt,$$

$$I_{\beta \beta_0}(\eta) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^n \int_0^T E \left[ \frac{Y_i(t) \{1 + m_0(t)\}}{\left\{m_0(t) + \beta_s^T Z_s(t)\right\}^2} \left( \frac{\eta(t)}{1 + m_0(t)} - \frac{\eta(t)}{m_0(t) + \beta_s^T Z_s(t)} \right)^T \right] dt,$$

$$I_{\beta_0 \beta_0}(\eta) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^n \int_0^T E \left[ \frac{Y_i(t) \{1 + m_0(t)\}}{\left\{m_0(t) + \beta_s^T Z_s(t)\right\}^2} \left( \frac{\eta(t)}{1 + m_0(t)} - \frac{\eta(t)}{m_0(t) + \beta_s^T Z_s(t)} \right)^2 \right] dt.$$

We shall say that matrix A is ‘larger’ than matrix B if $A - B$ is nonnegative definite. Then, by an application of the Cauchy-Schwarz inequality, $I_{\beta \beta}(\eta) = \{I_{\beta \beta}(\eta) - I_{\beta \beta_0}(\eta)I_{\beta_0 \beta}(\eta)^T\}^{-1}$ reaches its maximum at $\eta = \eta_0$. Here, $\eta_0(\cdot)$ satisfies the ordinary differential equation $\eta_0(t) E[Y(t)/\{1 + m_0(t)\}] - \eta_0(t) E[Y(t)/\{m_0(t) + \beta_s^T Z_s(t)\}] = E[Y(t)Z_s(t)/\{m_0(t) + \beta_s^T Z_s(t)\}]$, with the solution of $\eta_0(t) = \overline{P}(t)^{-1} \int_t^0 \overline{P}(u) \overline{Q}(u) du$. Here, $\overline{P}(t) = \exp(-\int_t^0 E[Y(u)/\{m_0(u) + \beta_s^T Z_s(u)\}] dt)$.
construct a regular semiparametric estimator with its asymptotic variance reaching 

\[ I_{\beta}^{-1}(\eta_0) = \left( \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \int_0^{\tau} E \left[ Y_i(t) \left\{ Z_i \left( \frac{\{m_0(t) + \beta^T Z_i\}^2}{\{m_0(t) + \beta^T Z_i\}^2} - \mu_0(t; \beta) \right) \right\} \right] dt \right)^{-1}, \quad (11) \]

where \( \mu_0(t; \beta) = \lim_n E[\sum_i Y_i(t) Z_i/\{m_0(t) + \beta^T Z_i\}] / E[\sum_i Y_i(t)/\{m_0(t) + \beta^T Z_i\}] \).

The sample-splitting technique in Lai & Ying and Lin & Ying (1994) can be used to construct a regular semiparametric estimator with its asymptotic variance reaching \( I_{\beta}^{-1}(\eta_0) \). The entire data would be randomly divided into two evenly disjoint groups. One has the first \( n_0 = [n/2] \) samples and the other has the remaining \( n_1 = n - n_0 \) samples. Let \( \hat{\beta}_j \) and \( \hat{m}_{0,j}(\cdot) \) be the ad hoc estimators of \( \beta \) and \( m_0(\cdot) \) estimated from the \( (j+1) \)st group for \( j = 0, 1 \). For example, they can be obtained from either the extended Buckley-James estimation procedure or the quasi partial estimation procedure. Consider the following estimating functions for \( \beta \):

\[ g(\beta) = \sum_{j=0}^{1} \sum_{i=1}^{n_j} \int_0^{\tau} Z_i - \hat{Z}(t; \hat{\beta}_{1-j}, \hat{m}_{0,1-j}) \left[ dN_i(t) - \frac{Y_i(t) d\{t + \hat{m}_{0,1-j}(t)\}}{\hat{m}_{0,1-j}(t) + \beta^T Z_i} \right], \]

where \( \hat{Z}(t; \beta, m_0) = \sum_i [Y_i(t) Z_i/\{m_0(t) + \beta^T Z_i\}] / \sum_i [Y_i(t)/\{m_0(t) + \beta^T Z_i\}] \). By standard counting process arguments, the asymptotic behaviour for one group is not affected by its estimates of \( \beta \) and \( m_0 \) from another group, but they asymptotically approach the true values. Thus the asymptotic variance of the semiparametric estimator \( \hat{\beta} \) such that \( g(\hat{\beta}) = 0 \) equals \( I_{\beta}^{-1}(\eta_0) \). According to Bickel et al. (1993), the upper parametric information bound for \( \beta \) in the model (10) at \( \beta_* \) would lead to the semiparametric information bound for \( \beta \) at \( \beta_* \) in the model (2) as well. Therefore the semiparametric information bound of the linear mean residual life model is \( I_{\beta}^{-1}(\eta_0) \); that is, by the Cramér-Rao inequality, the variance of any regular semiparametric estimator \( \tilde{\beta} \) in the linear mean residual life model is larger than \( (I_{\beta_{\beta}} - I_{\beta_0} I_{\beta_0}^{-1} I_{\beta_0})^{-1} \) for any \( \eta_i \) if \( n^{1/2}(\tilde{\beta} - \beta_* \) converges to a zero-mean normal.

9
3 Examples

We consider the well-known Stanford heart transplantation data analysed in Miller & Halpern (1982). The time-to-event outcome of this dataset is the lifetime since first heart transplantation between October 1967 and February 1980. Two covariates were originally used in analysis, namely, age at the time of first transplant and the T5 mismatch score, which measures the degree of tissue incompatibility between the hearts of the initial donor and recipient with respect to HLA antigens. To provide a comparison with their results, we apply the linear mean residual life model to the lifetimes with these two covariates. The results are tabulated in Table 1, along with those from the partial likelihood estimates of the Cox proportional hazards models and from the Buckley-James estimates of the linear regression models for the base-10 log-transformed lifetimes.

[Table 1 about here]

All the estimates in the linear mean residual life models suggest that both age and the T5 mismatch score are negatively associated with the patients’ life expectancy, although none of them is significant for the T5 mismatch scores. The covariate age is not only a significant predictor for the patients’ hazard and their average survival lifetimes, but also for their life expectancy throughout time. Among the 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. In addition, as demonstrated in Miller & Halpern (1982), a quadratic relationship to the covariate of age is plausible for both the Cox proportional hazards model and the linear regression model. We therefore fitted the linear mean residual life model with the T5 mismatch scores omitted because their nonsignificance as shown in Table 1. Table 2 shows that, both age and squared age are significant predictors of the life expectancy. However, their effects are of different signs: age itself is positively correlated with the life expectancy while the quadratic age has a negative influence on life expectancy. This demonstrates similar patterns to those shown by the Cox model and the linear regression
model. Again, the efficient estimation procedure yields smaller variances than do the other ad hoc procedures.

We can assess the fit of the linear mean residual life model using the goodness-of-fit test of Gill & Schumacher (1987). We consider two different weight functions, $W_1(t)$ and $W_2(t)$, say, in the weighted quasi partial estimating equations of $g_{w_1}(\beta)$ and $g_{w_2}(\beta)$, respectively. Their corresponding estimators are denoted by $\hat{\beta}_{w_1}$ and $\hat{\beta}_{w_2}$, respectively. Then a significant difference of $\hat{\beta}_{w_1} - \hat{\beta}_{w_2}$ should suggest lack of fit. To implement this, we adapt the reliable variation suggested in Wei et al. (1990), involving the goodness-of-fit test statistic defined by

$$
\kappa(W_1, W_2) = \min_{\beta \in \mathcal{N}(\hat{\beta}_{w_1})} \{(g_{w_1}(\beta)^T, g_{w_2}(\beta)^T) \Sigma_w^{-1} \hat{\beta}_{w_1} (g_{w_1}(\beta)^T, g_{w_2}(\beta)^T)^T \},
$$

where $\mathcal{N}(\hat{\beta}_{w_1})$ is a neighbourhood of $\hat{\beta}_{w_1}$. Here $\Sigma_w(\hat{\beta}_{w_1})$ is the covariance matrix of $(g_{w_1}(\hat{\beta}_{w_1})^T, g_{w_2}(\hat{\beta}_{w_1})^T)^T$, which can be estimated by computer-intensive methods, such as the bootstrap. Under the null hypothesis that the linear mean residual life model is adequate, $\kappa(W_1, W_2)$ is approximately distributed as $\chi^2_p$. For the Stanford data, $\kappa(W_1, W_2) = 4.81$ ($p = 0.09$) for the linear mean residual life model based on age and T5, which suggests a mild lack of fit. For the linear mean residual life model based on age and squared age, the test statistic is 1.17 ($p > 0.5$) which does not reject the model.

4 Discussion

Some simulations were conducted to assess the finite-sample properties of the proposed methods. We took the sample size $n$ to be 100 or 200 with two covariates $Z = (Z_1, Z_2)^T$ for each of $n$ subjects. The covariate $Z_1$ is a Ber(0.5) random variable and $Z_2 \sim Un[0, 1]$. We chose $m_*(t) = t + 1$, which corresponds to the Pareto distribution with the baseline survival function being $(1 + t)^{-2}$. Failure times were generated according to the linear mean residual life model. The true parameters were $(0, 0)^T$ or $(1, 1)^T$. Independent censoring times were generated from the uniform distribution so as to allow about 30% of the observations to be
The simulation results are summarised in Table 3, with each entry based on 1000 simulated datasets. The results show that the estimators of $\beta_*$ are virtually unbiased and the nominal 95% confidence intervals have acceptable coverage probabilities. More extensive simulation studies are needed to evaluate the efficiency of the proposed methods.

There is, however, the critical challenge of the tail behaviour of underlying distributions of failure times. When the underlying failure times are heavily right-skewed and censored early, as in the case of 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 those in Gill (1983) can be extended. In general, it is difficult to determine a reasonable upper limit $\tau$ without the robustness of the proposed methodologies being compromised. One possible approach is modify the fully unspecified $m_0(\cdot)$ by including a parametric component in the tail. For instance, let $\tilde{\tau}$ be a prespecified truncation time. Then it is assumed that

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

where $m_r$ is some positive constant. This means that the baseline mean residual life function is unspecified up to the truncation time $\tilde{\tau}$, while it becomes exponential after $\tilde{\tau}$. Then it is straightforward to extend the proposed methodologies to the whole positive real line.

**ACKNOWLEDGEMENT**

The authors would like to thank the associate editor and an anonymous reviewer for their constructive comments and suggestions that led to a significant improvement of this paper. They would also thank Professors Zhiliang Ying, Chen-Hsin Chen and Louis Chen for fruitful discussion during the first author’s visit to the National University of Singapore in 2005-6, which was partially funded by the Institute for Mathematical Sciences, National University of Singapore.
References


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

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Age</th>
<th>T5 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}_1$</td>
<td>s.e.</td>
</tr>
<tr>
<td>Cox model</td>
<td>0.030</td>
<td>0.011</td>
</tr>
<tr>
<td>Linear regression model</td>
<td>-0.015</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Linear mean residual life model

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Buckley-James</td>
<td>-20.045</td>
<td>9.315</td>
<td>-60.891</td>
<td>162.503</td>
</tr>
<tr>
<td>Quasi partial score</td>
<td>-23.517</td>
<td>10.227</td>
<td>-113.822</td>
<td>151.109</td>
</tr>
<tr>
<td>Efficient estimation</td>
<td>-21.168</td>
<td>9.031</td>
<td>-100.025</td>
<td>171.035</td>
</tr>
</tbody>
</table>

Cox model, estimates obtained in original time-scale; Linear regression model, Buckley-James estimates obtained in 10-base log-transformed scale; linear mean residual life model, estimates obtained in original time-scale; s.e., estimated standard error
Table 2: Regression estimates with standard errors for 152 Stanford heart transplant patients who survived at least 10 days

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Age</th>
<th>Age²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}_1$</td>
<td>s.e.</td>
</tr>
<tr>
<td>Cox model</td>
<td>-0.146</td>
<td>0.055</td>
</tr>
<tr>
<td>Linear regression model</td>
<td>0.107</td>
<td>0.037</td>
</tr>
<tr>
<td>Linear mean residual life model</td>
<td>107.532</td>
<td>51.066</td>
</tr>
<tr>
<td>Buckley-James</td>
<td>107.951</td>
<td>51.268</td>
</tr>
<tr>
<td>Quasi partial score</td>
<td>108.078</td>
<td>50.381</td>
</tr>
</tbody>
</table>

Cox model, estimates obtained in original time-scale; Linear regression model, Buckley-James estimates obtained in 10-base log-transformed scale; linear mean residual life model, estimates obtained in original time-scale; s.e., estimated standard error
### Table 3: Summary of simulation studies

<table>
<thead>
<tr>
<th>$\beta_s$</th>
<th>$n$</th>
<th>Estimation</th>
<th>Cov. Prob.</th>
<th>Bias</th>
<th>Cov. Prob.</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,0)</td>
<td>100</td>
<td>BJ</td>
<td>0.951</td>
<td>-0.014</td>
<td>0.948</td>
<td>-0.017</td>
</tr>
<tr>
<td>(0,0)</td>
<td>100</td>
<td>QP</td>
<td>0.944</td>
<td>0.011</td>
<td>0.937</td>
<td>-0.012</td>
</tr>
<tr>
<td>(0,0)</td>
<td>200</td>
<td>BJ</td>
<td>0.943</td>
<td>0.004</td>
<td>0.951</td>
<td>-0.006</td>
</tr>
<tr>
<td>(0,0)</td>
<td>200</td>
<td>QP</td>
<td>0.935</td>
<td>-0.003</td>
<td>0.951</td>
<td>-0.005</td>
</tr>
<tr>
<td>(1,1)</td>
<td>100</td>
<td>BJ</td>
<td>0.946</td>
<td>-0.009</td>
<td>0.965</td>
<td>0.018</td>
</tr>
<tr>
<td>(1,1)</td>
<td>100</td>
<td>QP</td>
<td>0.949</td>
<td>-0.007</td>
<td>0.951</td>
<td>0.015</td>
</tr>
<tr>
<td>(1,1)</td>
<td>200</td>
<td>BJ</td>
<td>0.944</td>
<td>0.004</td>
<td>0.953</td>
<td>-0.007</td>
</tr>
<tr>
<td>(1,1)</td>
<td>200</td>
<td>QP</td>
<td>0.955</td>
<td>0.003</td>
<td>0.948</td>
<td>0.005</td>
</tr>
</tbody>
</table>

BJ, the Buckley-James approach; QP, the quasi partial score equation approach; Cov. Prob., coverage probabilities of the 95% nominal confidence intervals; Bias, mean difference between the estimate and the true value.