An additive-multiplicative restricted mean residual life model

Zahra Mansourvar
Torben Martinussen
Thomas Scheike

Research Report 15/02

Department of Biostatistics
University of Copenhagen
An additive-multiplicative restricted mean residual life model

Zahra Mansourvar, Torben Martinussen, Thomas H. Scheike

Department of Biostatistics, University of Copenhagen

Abstract
The mean residual life measures the expected remaining life of a subject who has survived up to a particular time. When survival time distribution is highly skewed or heavy tailed, the restricted mean residual life must be considered. In this paper, we propose an additive-multiplicative restricted mean residual life model to study the association between the restricted mean residual life function and potential regression covariates in the presence of right censoring. This model extends the proportional mean residual life model using an additive model as its covariate dependent baseline. For the suggested model some covariate effects are allowed to be time-varying. To estimate the model parameters, martingale estimating equations are developed, and the large sample properties of the resulting estimators are established. In addition, to assess the adequacy of the model, we investigate a goodness of fit test that is asymptotically justified. The proposed methodology is evaluated via simulation studies and further applied to a kidney cancer data set collected from a clinical trial.

Keywords: Additive model; Counting process; Martingale estimating equation; Model checking; Proportional model; Restricted mean residual life; Right censoring; Time-varying effect.

1 Introduction
The mean residual life function is of interest in biomedical and reliability research, actuarial studies, demography and other disciplines. For a non-negative survival time \( T \) with finite expectation, the mean residual life function at time \( t \geq 0 \) is defined as \( m(t) = E(T - t \mid T > t) \). This function can sometimes serve as more desirable tool than the survival function and the hazard function. Whereas the
hazard function at time $t$ provides information about a small interval after $t$, the mean residual life function at $t$ considers information about the whole remaining interval after $t$. For instance, it may be more informative to tell a breast cancer patient how long she can survive, on average given her survival up to time $t$ as compared to her instantaneous survival chance. Many studies of mean residual life function have been investigated, mostly focusing on its probabilistic properties. A comprehensive review of previous research on this function is given by Sun and Zhang (2009).

To evaluate the effects of covariates on the mean residual life function, Oakes and Dasu (1990) proposed the proportional mean residual life model

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

(1)

where $m(t \mid X) = E(T - t \mid T > t; X)$ is the mean residual life function corresponding to a vector covariate $X$, $m_0(t)$ is some unspecified baseline mean residual life function when $X = 0$, and $\beta$ is an unknown vector of regression parameters. Inference on the parameters in model (1) has been discussed in Zahedi (1992), which is maximum likelihood method for complete survival data when the baseline mean residual life function is parametric. Maguluri and Zhang (1994) used the underlying proportional hazard structure of the model to develop estimation procedures for $\beta$ in model (1), when $m_0(t)$ is unknown, but mainly for uncensored survival data. To accommodate censoring, Chen et al. (2005) extended the estimation procedure of Maguluri and Zhang (1994) to censored survival data using the inverse probability of censoring weighting technique of Robins and Rotnitzky (1992) to estimate the parameters in model (1). Chen and Cheng (2005) employed a quasi partial score approach based on counting process theory to develop new inference procedures for the regression analysis of model (1) when censoring is present and $m_0(t)$ is not known.

As an alternative to the proportional mean residual life model, Chen and Cheng (2006) considered a semi-parametric linear mean residual life model as follows

$$m(t \mid X) = m_0(t) + X^T \beta,$$

(2)

in which all components are defined similarly as in (1). They developed two estimation procedures when $m_0(t)$ is unknown. In the presence of censoring, either a Buckley-James approach (Buckley and James, 1979) or a martingale-based estimating equation similar to those in Chen and Cheng (2005) can be used. Model (2) has also been studied by Chen (2007).
A more general class of regression models, termed as the transformed mean residual life models, were proposed by Sun and Zhang (2009). To estimate the model parameters, they made use of inverse probability of censoring weighting technique and developed a system of estimating equations under right censoring when the baseline mean residual life function is unspecified. More recently, Sun et al. (2012) considered a flexible class of semi-parametric mean residual life models where some effects may be time-varying and some may be constant over time. In the presence of right censoring, they applied inverse probability of censoring weighting approach and developed inference procedures for estimating the model parameters.

The mean residual life function is especially useful when the tail behavior of the survival time distribution is of interest. However in survival analysis, because of the inevitable right censoring of event times, the crucial upper tail of the survival time distribution can not be observed. Thus, the mean residual life function of $T$ may not be estimable at some points. To overcome this issue, some investigators supposed an extra assumption that the support of censoring time is longer than that of the survival time to ensure the mean residual life function is estimable. Such an assumption has been adopted by Chen and Cheng (2005, 2006), Chen et al. (2005), Sun and Zhang (2009) and Sun et al. (2012). Although this assumption appears to hold in some situations, it may fail if the survival time is highly right-skewed or if the follow-up time is too short. As an alternative, instead of the mean residual life function, we consider the restricted mean residual life function which for a pre-specified time $\tau > 0$ is $m_\tau(t) = E(T_\tau - t \mid T_\tau > t)$ $(0 < t < \tau)$, where $T_\tau = \min(T, \tau)$. For example, a patient diagnosed with breast cancer may be interested in knowing how much longer she is expected to survive within the next $\tau$ years given she has survived so far. The choice of $\tau$ to specify the restricted mean residual life function is crucial, which may be determined at the study design stage based on clinical relevance and feasibility of conducting the study. The value of $\tau$ should be chosen by subject matters, and should not be derived by the biostatistician. An important property of the restricted mean residual life function is $m_\tau(\tau) = 0$.

The proportional and linear mean residual life models (1) and (2) postulate different relationships between the mean residual life function and covariates, and sometimes it will not be obvious which of the models are to be desired. These two models may often be utilized to complement each other and to supply different summary measures. In this paper, we propose a flexible model that combines the
proportional and linear restricted mean residual life model which takes the form

\[ m_r(t \mid X, Z) = \{X^T \alpha_r(t)\} \exp(Z^T \beta_r), \]

where \( m_r(t \mid X, Z) = E(T_r - t \mid T_r > t; X, Z) \), \( X \) and \( Z \) are bounded covariate vectors of dimensions \((p + 1)\) and \( q \), respectively, and \( \alpha_r(t) \) and \( \beta_r \) are conformal vectors of unknown time-dependent and time-independent regression parameters with \( \alpha_r(\tau) = 0 \). The first component of \( X \) is set to be 1, which contributes to the baseline restricted mean residual life function. Under model (3), some covariate effects work additively on the restricted mean residual life function and yield non-parametric and time-varying effects. Other covariates have constant multiplicative effects. The model extends the proportional mean residual life model (1) by allowing the baseline mean residual life function to depend on covariates through an additive model. It further extends the linear mean residual life model (2) by considering time-varying effects and therefore provides a very flexible class of models.

We develop inference procedures for estimating the parameters of model (3), using the martingale theory. Our proposed method is much easier than the inverse probability of censoring weighting approach applied by Sun et al. (2012) in which the tail estimates may be unstable when the weights are very small. Furthermore, to use the inverse probability of censoring weighting approach, one needs to fit a model for the censoring times. The censoring times are usually of no scientific interest, and moreover, a misspecified censoring model might invalidate the estimates of the parameters of interest.

The paper is structured as follows. In Section 2, we discuss semi-parametric inference procedures for estimating the non-parametric component \( \alpha_r(t) \) and parametric component \( \beta_r \) in model (3) by applying martingale estimating equations in the presence of right censoring. The asymptotic properties of the proposed estimators are presented in Section 3, and here we also discuss how to construct a uniform confidence band for the restricted mean residual life function. In Section 4, we develop a goodness of fit test for checking the adequacy of the model and provide a test for investigating whether or not an additive component has a time-varying effect. Section 5 reports some results from simulation studies conducted for evaluating the proposed methods. To illustrate the use of the model, in Section 6, we apply the model to a data on patients with advanced kidney cancer. Section 7 contains some discussion, and regularity conditions and theoretical proofs are collected in the Appendix.
2 Inference procedures

Let \( T_\tau, X \) and \( Z \) be defined as before and \( C \) be the potential censoring time and assume that \( T_\tau \) is independent of \( C \) given the covariates \( X \) and \( Z \). Let \((T_{\tau i}, C_i, X_i, Z_i)\), where \( i = 1, \ldots, n \) be independent replicates of \((T_\tau, C, X, Z)\). When right censoring is present, the observed data set thus consists of \( n \) independent copies of \((\tilde{T}_{\tau i}, \Delta_i, X_i, Z_i)\), where \( \tilde{T}_{\tau i} = \min(T_{\tau i}, C_i) \) and \( \Delta_i = I(T_{\tau i} \leq C_i) \). Here, \( I(.) \) is the indicator function. Furthermore, let \( N_i(t) = I(\tilde{T}_{\tau i} \leq t, \Delta_i = 1) \) be a counting process, \( Y_i(t) = I(\tilde{T}_{\tau i} \geq t) \) be an at risk indicator process, and let \( \Lambda_i(t) \) be the cumulative hazard function of \( T_{\tau i} \). In the following, for the sake of notation, \( \alpha_\tau(t) \) and \( \beta_\tau \) are interchanged with \( \alpha(t) \) and \( \beta \).

It is shown in Corollary 1.4.1 of Fleming and Harrington (1991) that

\[
E \{ dN_i(t) \mid \mathcal{F}_t ; \alpha_0(\cdot), \beta_0 \} = Y_i(t) d\Lambda_i(t; \alpha_0, \beta_0),
\]

where \( \alpha_0(t) \) and \( \beta_0 \) are the true values of the parameters \( \alpha(t) \) and \( \beta \) in (3), respectively, and the filtration of \( \mathcal{F}_t \) defined by

\[
\mathcal{F}_t = \sigma \{ N_i(u), Y_i(u), X_i, Z_i : 0 \leq u \leq t; \ i = 1, \ldots, n \}.\]

Let

\[
dM_i(t; \alpha_0, \beta_0) = dN_i(t) - Y_i(t) d\Lambda_i(t \mid X_i, Z_i; \alpha_0, \beta_0) \quad (i = 1, \ldots, n),
\]

where \( \{M_i(t; \alpha_0, \beta_0)\} \) are zero-mean martingale with respect to \( \mathcal{F}_t \). Therefore without assuming any particular form for the time-varying effect \( \alpha(t) \), it is natural to estimate \( \alpha_0(t) \) and \( \beta_0 \) from estimating equations parallel to the partial score equations

\[
\sum_{i=1}^{n} Y_i(t)X_i \{ dN_i(t) - Y_i(t) d\Lambda_i(t \mid X_i, Z_i; \alpha, \beta) \} = 0 \quad (0 \leq t < \tau),
\]

\[
\sum_{i=1}^{n} \int_{0}^{\tau} Y_i(t)Z_i \{ dN_i(t) - Y_i(t) d\Lambda_i(t \mid X_i, Z_i; \alpha, \beta) \} = 0.
\]

It is well known that the survival function of \( T_\tau \) given \( X \) and \( Z \) is

\[
S_\tau(t \mid X, Z) = \frac{m_\tau(0 \mid X, Z)}{m_\tau(t \mid X, Z)} \exp \left\{ - \int_{0}^{t} \frac{du}{m_\tau(u \mid X, Z)} \right\} \quad (t < \tau).
\]
An asymptotically equivalent expression is product-integral, see Andersen et al. (1993, § II.6) and where Volterra equation (Andersen et al., 1993, p. 91; see Yang and Prentice, 1999) can in (9) has a backward recursive structure, but admits an explicit solution. The estimation of make the convention that where 

\[ Z. Mansourvar et al. \]

Then under model (3), one obtains that

\[ X_i^T \alpha(t) dA_i(t \mid X_i, Z_i; \alpha, \beta) = \exp(-Z_i^T \beta)dt + X_i^T d\alpha(t) \quad (i = 1, \ldots, n). \]

Therefore, by analogy with (4), the following estimating equations can be used for estimating \( \alpha_0(t) \) and \( \beta_0 \) in model (3), respectively

\[ \sum_{i=1}^{n} Y_i(t) X_i [X_i^T \alpha(t) dN_i(t) - Y_i(t) \{\exp(-Z_i^T \beta)dt + X_i^T d\alpha(t)\}] = 0, \tag{6} \]

\[ \sum_{i=1}^{n} \int_0^T Y_i(t) Z_i [X_i^T \alpha(t) dN_i(t) - Y_i(t) \{\exp(-Z_i^T \beta)dt + X_i^T d\alpha(t)\}] = 0. \tag{7} \]

In view of (6), it can be obtained that

\[ d\alpha(t) = \left\{X^T(t)X(t)\right\}^{-1} X^T(t) \left\{d\tilde{N}(t)X(t)\alpha(t) - \exp(-Z\beta)dt\right\}, \tag{8} \]

where \( X(t) = \{Y_1(t)X_1, \ldots, Y_n(t)X_n\}^T \) is a design matrix of dimension \( n \times (p + 1) \), \( d\tilde{N}(t) = \text{diag}\{dN_i(t)\} \) and \( \exp(-Z\beta) = \{\exp(-Z_1^T \beta), \ldots, \exp(-Z_n^T \beta)\}^T \). We make the convention that \( \{X^T(t)X(t)\}^{-1} = 0 \) when the inverse does not exist. Since \( \alpha(\tau) = 0 \) in model (3), we suggest the following estimating equation for estimation of \( \alpha_0(t) \) (given \( \beta \)) by integrating (8) backwards in time

\[ \hat{\alpha}(t) = \int_t^T \left\{X^T(s)X(s)\right\}^{-1} X^T(s) \left\{\exp(-Z\beta)ds - d\tilde{N}(s)X(s)\hat{\alpha}(s^+)\right\}, \tag{9} \]

where \( \hat{\alpha}(s^+) \) is the right hand limit of \( \hat{\alpha}(t) \) at \( s \). The estimator of \( \alpha_0(t) \) displayed in (9) has a backward recursive structure, but admits an explicit solution. The Volterra equation (Andersen et al., 1993, p. 91; see Yang and Prentice, 1999) can easily be adapted to yield the solution for equation (9) as

\[ \hat{\alpha}(t; \beta) = P(t) \int_t^T P(s^+)^{-1} \left\{X^T(s)X(s)\right\}^{-1} X^T(s) \exp(-Z\beta)ds, \]

where \( P(t) = \prod_{[t, \tau]} [I - \{X^T(s)X(s)\}^{-1}X^T(s)d\tilde{N}(s)X(s)] \), with \( \prod_{[t, \tau]} \) denoting a product-integral, see Andersen et al. (1993, § II.6) and \( I \) is the identity matrix. An asymptotically equivalent expression is

\[ \hat{\alpha}(t; \beta) = Q(t)^{-1} \int_t^T Q(s^+) \left\{X^T(s)X(s)\right\}^{-1} X^T(s) \exp(-Z\beta)ds, \tag{10} \]

where \( Q(t) = \prod_{[t, \tau]} [I + \{X^T(s)X(s)\}^{-1}X^T(s)d\tilde{N}(s)X(s)] \).
An estimator for $\beta_0$ is given by substituting $\hat{\alpha}(t; \beta)$ for $\alpha(t)$ in (7) and solving that for $\beta$. Hence, using expression (8), the score equation for $\beta_0$ can be simplified as

$$U\{\tau, \beta; \hat{\alpha}(t; \beta)\} = \int_0^\tau \{Z^T(t) - H_1(t)X^T(t)\} \left\{d\hat{N}(t)X(t)\hat{\alpha}(t; \beta) - \exp(-Z\beta)dt\right\},$$

(11)

where $Z(t) = \{Y_1(t)Z_1, \ldots, Y_n(t)Z_n\}^T$ is a design matrix of dimension $n \times q$ and $H_1(t) = Z^T(t)X(t)\{X^T(t)X(t)\}^{-1}$. It can also be written that

$$U\{\tau, \beta; \hat{\alpha}(t; \beta)\} = \int_0^\tau Z^T(t)C(t) \left\{d\hat{N}(t)X(t)\hat{\alpha}(t; \beta) - \exp(-Z\beta)dt\right\},$$

where $C(t) = I - X(t)\{X^T(t)X(t)\}^{-1}X^T(t)$ is the projection matrix that projects onto the orthogonal space spanned by $X(t)$. Let $\tilde{\beta}$ denote the solution to $U\{\tau, \beta; \hat{\alpha}(t; \beta)\} = 0$. The corresponding estimator of $\alpha_0(t)$ given $\tilde{\beta}$ is $\hat{\alpha}(t) = \hat{\alpha}(t; \tilde{\beta})$.

3 Asymptotic properties

In the Appendix, we establish the asymptotic properties of the estimators given in the previous section. We show the existence, uniqueness and strong consistency of $\hat{\beta}$ together with the strong consistency of $\hat{\alpha}(t)$. We also derive asymptotic normality results, based on expansions in terms of sums of independent and identically distributed random variables. The results are summarized in the following theorems.

Theorem 1 Under the regularity conditions stated in the Appendix, $\hat{\beta}$ exists and is unique. Moreover, $\hat{\beta}$ is strongly consistent for $\beta_0$, and $\hat{\alpha}(t) \rightarrow \alpha_0(t)$ almost surely uniformly in $t \in [0, \tau]$ as $n \rightarrow \infty$.

Theorem 2 Under the regularity conditions stated in the Appendix, we have

(i) $n^{-1/2}U\{\tau, \beta_0; \hat{\alpha}(t; \beta_0)\}$ converges to a normal distribution with mean zero and a variance-covariance matrix that can be consistently estimated by $\hat{\Sigma}$, where

$$\hat{\Sigma} = n^{-1} \sum_{i=1}^n \hat{\xi}_i\hat{\xi}_i^T,$$

$$\hat{\xi}_i = \int_0^\tau \{Z_i(t) - H_1(t)X_i(t) - H_2(t)X_i(t)\} X_i^T(t)\hat{\alpha}(t) d\hat{M}_i(t),$$
propose to use a resampling scheme as in Lin et al. (2000) to approximate the dis-
ance function at \( Z \). Mansourvar et al. can be approximated by that of the zero-mean Gaussian process 
Appendix that the distribution of the process \( X \) with covariates 
(ii) \( n^{1/2}(\hat{\beta} - \beta_0) \) is asymptotically normal with mean zero and a variance-covariance 
matrix that can be consistently estimated by \( \hat{D}(\tau, \hat{\beta}) = \hat{D}(\tau, \hat{\beta})^{-1} \hat{\Sigma} \hat{D}(\tau, \hat{\beta})^{-1} \), where 
\[
\hat{D}(\tau, \hat{\beta}) = -n^{-1} \sum_{i=1}^{n} \int_{t}^{\tau} \{ Z_i(t) - H_1(t)X_i(t) - H_2(t)X_i(t) \} \exp(-Z_i^T\hat{\beta}Z_i(t))dt.
\]

**Theorem 3** Under the regularity conditions stated in the Appendix, \( n^{1/2}\{\hat{\phi}(t) - \phi_0(t)\} \ (0 \leq t \leq \tau) \) converges weakly to a zero-mean Gaussian process whose covariance function at \( (s, t) \) can be consistently estimated by \( \hat{\Gamma}(s, t) = n^{-1} \sum_{i=1}^{n} \hat{\phi}_i(s)\hat{\phi}_i(t)^T \) where 
\[
\hat{\phi}_i(t) = \hat{\zeta}_i(t) + \hat{G}(t; \hat{\beta}) \hat{D}(\tau, \hat{\beta})^{-1} \hat{\xi}_i,
\]
and
\[
\hat{\zeta}_i(t) = -nQ^{-1} \int_{t}^{\tau} Q(s^+) \{ X^T(s)X(s) \}^{-1} X_i(s)X_i^T(s)\hat{\phi}_i(s)d\hat{M}_i(s),
\]
\[
\hat{G}(t; \hat{\beta}) = -Q(t^{-1}) \int_{t}^{\tau} Q(s^+) \{ X^T(s)X(s) \}^{-1} X^T(s) \text{diag}\left\{ \exp(-Z_i^T\hat{\beta}) \right\} Z(s)ds.
\]

In many applications, it is of interest to predict the restricted mean residual life function for a given patient with certain set of covariates. Assume that a subject with covariates \( X_0 \) and \( Z_0 \) has the restricted mean residual life function \( m_r(t \mid X_0, Z_0) = X_0^T\alpha(t) \exp(Z_0^T\beta) \), that obviously can be estimated by \( \hat{m}_r(t \mid X_0, Z_0) = X_0^T\hat{\alpha}(t) \exp(Z_0^T\hat{\beta}) \). To construct a uniform confidence band for \( m_r(t \mid X_0, Z_0) \), we propose to use a resampling scheme as in Lin et al. (2000) to approximate the distribution of \( n^{1/2}\{\hat{m}_r(t \mid X_0, Z_0) - m_r(t \mid X_0, Z_0)\} \). It is demonstrated in the Appendix that the distribution of the process \( n^{1/2}\{\hat{m}_r(t \mid X_0, Z_0) - m_r(t \mid X_0, Z_0)\} \) can be approximated by that of the zero-mean Gaussian process \( \Delta m_r(t) = n^{-1/2} \sum_{i=1}^{n} \hat{\epsilon}_i(t)\hat{\Omega}_i \), where
\[
\hat{\epsilon}_i(t) = \exp(Z_0^T\hat{\beta}) \left\{ X_0^T\hat{\phi}_i(t) + X_0^T\hat{\alpha}(t) Z_0^T \hat{D}(\tau, \hat{\beta})^{-1} \hat{\xi}_i \right\}, \quad (12)
\]
and \((\Omega_1, \cdots, \Omega_n)\) are independent standard normal variables which are independent of the observed data. It also follows that \(\tilde{V}(t) = n^{-1}\sum_{i=1}^{n} \xi_i^2(t)\) is a consistent estimator of the variance of \(n^{1/2}\{m_\tau(t \mid X_0, Z_0) - m_\tau(t \mid X_0, Z_0)\}\). Therefore an approximate \((1 - \alpha)\)-confidence band for \(m_\tau(t \mid X_0, Z_0) \pm n^{-1/2} C_\alpha \tilde{V}^{1/2}(t)\), where \(C_\alpha\) is the \((1 - \alpha)\)-quantile of \(\sup_{t \in [0, \tau]} |\Delta_m^k(t)\tilde{V}^{-1/2}(t)|\) where \(\Delta_m^k(t)\) is the \(k\)th resampled process for \(1 \leq k \in \mathbb{N}\).

4 Test procedures

To evaluate the goodness of fit of model (3), and to test if covariates with time-varying effects are indeed time-dependent or significantly different from zero, we develop some asymptotically justified test procedures in the following. The suggested class of test statistics are based on cumulative martingale residuals.

First, to evaluate the goodness of fit of the covariates included in the multiplicative part of the model, we consider the cumulative score processes as in Lin et al. (1993). By the representation \(U(\tau, \beta)\) in (11), the observed score process can be written as

\[
U(t, \hat{\beta}) = \sum_{i=1}^{n} \int_{t}^{\tau} Z_i(s)X_i^T(s)\hat{\alpha}(s)d\hat{M}_i(s),
\]

which involves transformations of the martingale residual processes \(\hat{M}_i(t)\). Using the resampling approach (Lin et al., 2000), it can be established that the asymptotic distribution of \(U(t, \hat{\beta})\) is equivalent to the asymptotic distribution of

\[
\hat{U}(t, \hat{\beta}) = \sum_{i=1}^{n} \left\{ \hat{\xi}_i(t) - \hat{D}(t, \hat{\beta})\hat{D}(\tau, \hat{\beta})^{-1}\hat{\xi}_i \right\} \Omega_i,
\]

where \(\hat{\xi}_i(t) = \int_{t}^{\tau} \{Z_i(s) - H_1(s)X_i(s) - H_2(s)X_i(s)\}X_i^T(s)\hat{\alpha}(s)d\hat{M}_i(s)\), and \(\hat{D}(t, \hat{\beta}) = -n^{-1}\sum_{i=1}^{n} \int_{t}^{\tau} \{Z_i(s) - H_1(s)X_i(s) - H_2(s)X_i(s)\} \exp(-Z_i^T\hat{\beta})Z_i^T(ds)\). Thus to approximate the distribution of \(\hat{U}(t, \hat{\beta})\), we can simulate a number of realizations from \(\hat{U}(t, \hat{\beta})\) by repeatedly generating independent standard normal samples \((\Omega_1, \ldots, \Omega_n)\) while holding the observed data \((\hat{T}_i, \Delta_i, X_i, Z_i)\) fixed, where \(i = 1, \ldots, n\). If the model is proportional for the suggested covariates, then the \(j\)th component of the score process should behave as expected under the null. To assess how unusual the observed process \(U(t, \hat{\beta})\) behaves under model (3), we can plot it along with a few, say 50, realizations from \(\hat{U}(t, \hat{\beta})\). Moreover, this graphical inspection is naturally accompanied by the supremum test statistic

\[
F_1 = \sup_{t \in [0, \tau]} |U_j(t, \hat{\beta})|.
\]
The $p$-value of this test can be obtained by generating a large number of, say 1000, realizations from $\sup_{t \in [0, \tau]} |U_j(t, \hat{\beta})|$ and comparing them with the observed value of $\sup_{t \in [0, \tau]} |U_j(t, \hat{\beta})|$. However, the test statistic $F_1$ checks the proportionality of the effect of the $j$th covariate on the multiplicative part of model (3), it can be affected by any misspecification of the covariates in the additive part of the model.

Second, to do inference about the non-parametric components of model (3), a resampling scheme can approximate the asymptotic distribution of $n^{1/2} \{\hat{\alpha}(t) - \alpha_0(t)\}$ following the proof of Theorem 3 in the Appendix under the corresponding null hypothesis. Therefore, to test whether covariate $j$ that is included in the additive part of the model is significant with the null hypothesis $\alpha_j(t) = 0$ for all $t \in [0, \tau]$, we suggest the test statistic

$$F_2 = \sup_{t \in [0, \tau]} |\hat{\alpha}_j(t)|.$$  

The test statistic $F_2$ considers departures of $\hat{\alpha}_j(t)$ from 0. Furthermore, to test if an additive component has a time-varying effect with the null hypothesis $\alpha_j(t) = \gamma$ for all $t \in [0, \tau]$ for a constant free parameter but not necessarily known $\gamma$, we consider the test statistic

$$F_3 = \sup_{t \in [0, \tau]} \left| \hat{\alpha}_j(t) - \tau^{-1} \int_0^\tau \hat{\alpha}_j(t) dt \right|.$$  

The test statistic $F_3$ measures departures between $\hat{\alpha}_j(t)$ and one estimate of the constant effect $\gamma$ under the null, $\tau^{-1} \int_0^\tau \hat{\alpha}_j(t)$. Similarly, approximate critical value for these two test statistics can be simulated using the resampling approach. Also, to test the significance of constant multiplicative effects ($\beta$), we can use the Wald test.

5 Simulation study

We conducted a simulation study to evaluate the finite sample properties of the estimators of both non-parametric and parametric terms of model (3). In the study, we considered the following additive-multiplicative restricted mean residual life model

$$m_{\tau}(t \mid X_1, Z_1, Z_2) = \{\alpha_0(t) + X_1 \alpha_1(t)\} \exp(Z_1 \beta_1 + Z_2 \beta_2) \quad (0 < t < \tau), \quad (13)$$

where $Z_1$ is a uniform random variable on $(0, 1)$ and independent of $X_1$ and $Z_2$ that are independent Bernoulli random variables with success probability 0.5. We
set \( \alpha_0(t) = \tau - t \), \( \alpha_1(t) = (t - \tau) \exp(-1 - \tau) \) and \( \tau = 5 \). In view of (5), the survival function of \( T_\tau \) corresponding to model (13) is

\[
S_\tau(t \mid X_1, Z_1, Z_2) = \left\{ \frac{(\tau - t)/\tau}{(a-1)} \right\},
\]

where \( a = \exp(-Z_1\beta_1 - Z_2\beta_2)/\{1 - X_1\exp(-1 - \tau)\} \). Applying the inverse transform sampling, the restricted survival times \( T_\tau \) were generated from the foregoing survival function. In order to obtain a proper estimator of model (13), \( m_\tau(t \mid X_1, Z_1, Z_2) + t \) must be non-decreasing in \( t \) (see the Characterization Theorem of Hall and Wellner, 1984). This condition can be satisfied in (13) when \( a > 1 \). Therefore, the true parameter of \( (\beta_1, \beta_2)^T \) was taken to be \( (-0.2, -0.2)^T \) to impose this restriction. Independent censoring times were generated from the uniform distribution on \((0, c)\), where the constant \( c \) was selected to result in, on average, about 20\% or 40\% of observations censored. Each configuration was based on 1000 replications with sample sizes \( n = 100 \) and 300.

Table 1 summarizes the simulation results for the estimates of \( \beta_0 \) and \( \beta_1 \) along with the estimates of \( \alpha_0(t) \) and \( \alpha_1(t) \) at time points \( t = 0.5, 2.5 \) and 4.5. The results indicate that the proposed estimators perform quite well and provide virtually unbiased estimates. Some bias is seen for the baseline at the beginning time points, but it diminishes when the sample size is increased. The estimated standard errors are close to the empirical standard errors, and the nominal 95\% confidence intervals have acceptable coverage probabilities. We also considered other simulation configurations and obtained similar results (not shown here).

To investigate the performance of the model checking method, we also conducted some simulations to assess the size and power of the test based on the supremum statistic \( F_1 \). We assumed the true model is the additive-multiplicative restricted mean residual life model (13) in which \( Z_2 \) is omitted from the multiplicative part of model. For significance level of 0.05, the empirical size of the supremum test \( F_1 \) was estimated at 0.081, 0.064 and 0.063 for \( n = 100, 200 \) and 400, respectively with 20\% uniform censoring. We used 1000 realizations of the corresponding statistic \( F_1 \) with 1000 replications of the data. It shows the empirical size is close to the nominal size with sufficient sample size. To study the power of the supremum test \( F_1 \), we generated survival times \( T \) from the proportional mean residual life model (1). Here we let \( X = (X_1, X_2)^T \) where \( X_1 \) is a uniform random variable on \((0, 1)\) and \( X_2 \) is a Bernoulli random variable with success probability 0.5. We chose \( m_0(t) = 1 + 2t \) if \( X_2 = 0 \) and \( m_0(t) = 1.5 + t \) otherwise. The true parameter of \( \beta = (\beta_1, \beta_2)^T \) was \( (0.1, -0.1)^T \). Then the restricted survival times \( T_\tau \) were computed with \( \tau = 4 \). The estimated power of the supremum test \( F_1 \) at
the significance level of 0.05 was, respectively, 0.365, 0.595 and 0.871, for \( n = 100, 200 \) and 400 with no censoring. It indicates that the test has a reasonable power to detect deviations from the true model. The power increased with increases in sample size from 100 to 400.

Table 1: Summary of the simulation results. The columns are bias \( \times 10 \), sample standard error \( \times 10 \), mean of the standard error of the estimate \( \times 10 \), and empirical coverage probability \( \times 100 \)

<table>
<thead>
<tr>
<th>C%</th>
<th>Bias</th>
<th>SE</th>
<th>SEE</th>
<th>CP</th>
<th>Bias</th>
<th>SE</th>
<th>SEE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_1 )</td>
<td>20</td>
<td>0.016</td>
<td>1.201</td>
<td>1.165</td>
<td>94.7</td>
<td>0.012</td>
<td>0.706</td>
<td>0.680</td>
</tr>
<tr>
<td>40</td>
<td>0.014</td>
<td>1.307</td>
<td>1.295</td>
<td>95.1</td>
<td>0.004</td>
<td>0.736</td>
<td>0.747</td>
<td>96.1</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>20</td>
<td>0.015</td>
<td>0.693</td>
<td>0.674</td>
<td>94.3</td>
<td>0.014</td>
<td>0.380</td>
<td>0.392</td>
</tr>
<tr>
<td>40</td>
<td>-0.016</td>
<td>0.754</td>
<td>0.738</td>
<td>94.8</td>
<td>0.018</td>
<td>0.433</td>
<td>0.430</td>
<td>94.5</td>
</tr>
<tr>
<td>( \alpha_0(0.5) )</td>
<td>20</td>
<td>0.316</td>
<td>3.281</td>
<td>3.142</td>
<td>94.7</td>
<td>0.131</td>
<td>1.867</td>
<td>1.817</td>
</tr>
<tr>
<td>40</td>
<td>0.512</td>
<td>3.618</td>
<td>3.489</td>
<td>96.7</td>
<td>0.091</td>
<td>2.003</td>
<td>1.983</td>
<td>95.4</td>
</tr>
<tr>
<td>( \alpha_0(2.5) )</td>
<td>20</td>
<td>0.238</td>
<td>2.020</td>
<td>1.926</td>
<td>94.6</td>
<td>0.086</td>
<td>1.135</td>
<td>1.117</td>
</tr>
<tr>
<td>40</td>
<td>0.301</td>
<td>2.235</td>
<td>2.163</td>
<td>96.2</td>
<td>0.094</td>
<td>1.249</td>
<td>1.237</td>
<td>94.0</td>
</tr>
<tr>
<td>( \alpha_0(4.5) )</td>
<td>20</td>
<td>0.086</td>
<td>0.419</td>
<td>0.402</td>
<td>93.4</td>
<td>0.026</td>
<td>0.244</td>
<td>0.237</td>
</tr>
<tr>
<td>40</td>
<td>0.130</td>
<td>0.481</td>
<td>0.461</td>
<td>95.2</td>
<td>0.038</td>
<td>0.268</td>
<td>0.269</td>
<td>95.1</td>
</tr>
<tr>
<td>( \alpha_1(0.5) )</td>
<td>20</td>
<td>0.082</td>
<td>3.060</td>
<td>2.975</td>
<td>94.3</td>
<td>0.031</td>
<td>1.753</td>
<td>1.731</td>
</tr>
<tr>
<td>40</td>
<td>0.240</td>
<td>3.125</td>
<td>3.178</td>
<td>95.7</td>
<td>0.072</td>
<td>1.873</td>
<td>1.838</td>
<td>94.6</td>
</tr>
<tr>
<td>( \alpha_1(2.5) )</td>
<td>20</td>
<td>0.030</td>
<td>1.791</td>
<td>1.765</td>
<td>95.2</td>
<td>0.029</td>
<td>1.016</td>
<td>1.035</td>
</tr>
<tr>
<td>40</td>
<td>0.011</td>
<td>1.988</td>
<td>1.965</td>
<td>94.8</td>
<td>0.037</td>
<td>1.179</td>
<td>1.152</td>
<td>95.4</td>
</tr>
<tr>
<td>( \alpha_1(4.5) )</td>
<td>20</td>
<td>0.001</td>
<td>0.403</td>
<td>0.394</td>
<td>94.7</td>
<td>0.008</td>
<td>0.239</td>
<td>0.236</td>
</tr>
<tr>
<td>40</td>
<td>0.008</td>
<td>0.469</td>
<td>0.461</td>
<td>94.2</td>
<td>0.000</td>
<td>0.285</td>
<td>0.280</td>
<td>94.5</td>
</tr>
</tbody>
</table>

C%, the censoring percentage; Bias, the mean of 1000 point estimates of the parameter minus the true value; SE, the sample standard error of the estimates; SEE, the mean of the standard error of the estimates; CP, the empirical normal 95% coverage probability.
6 Application

To illustrate the proposed estimation procedure, we consider data from a randomized trial recruiting patients with advanced kidney cancer between February 1992 and November 1997 (Royston et al., 2004). The aim of this trial was to compare the effects of interferon-α (IFN) with medroxyprogesterone acetate (MPA) on the overall survival of patients that had metastasized (i.e. spread from the original cancer site to other organs). A 28% reduction in the mortality rate in the interferon-α group was reported (MRCRC, 1999). The present analysis is based on data updated to June 2001, including 322 deaths out of 347 patients during follow-up and the rest were alive, i.e., censored at the last follow-up (Royston and Parmar, 2011). In this study, 172 patients were randomly assigned to IFN and 175 to MPA. Royston and Parmar (2011) proposed the use of restricted mean survival time at $\tau = 4$ years.

We now apply the methodologies described in Sections 2 and 4 to this clinical trial in order to identify the potential factors that might affect the restricted mean residual life. In addition to the treatment factor, we also consider the World Health Organization (WHO) performance status. For treatment, we let it takes 1 if the patient was in the IFN group and 0 otherwise, and for WHO performance status, we define it takes 1 if the patient had restriction in activities of daily living and 0 if the patient had the best quality of life. To illustrate our method, we initially fit the proportional restricted mean residual life model with $\tau = 4$ years using the two covariates mentioned earlier in the multiplicative part of model (3). This model indicates that both treatment and WHO performance status are significant covariates ($p$-values are $0.004$ and $<0.001$, respectively). In order to evaluate the adequacy of the model for each of these covariates, we considered the score process test. The test process $F_1$ along with 50 resampled processes under the null are shown in Figure 1 where it is seen that the test process for WHO performance status is clearly outside of the expected range. The $p$-values of the supremum test statistics based on 5000 resamplings are 0.48 and $<0.001$ for treatment and WHO performance status, respectively. This indicates that the model does not fit well when WHO performance status is included in the multiplicative part of the model.

We therefore considered an additive-multiplicative restricted mean residual life model still with $\tau = 4$ years where WHO performance status has additive effect and treatment has multiplicative effect. The supremum test $F_1$ yields a $p$-value of 0.43 for treatment, suggesting no strong evidence for lacking fit of the multiplicative part of the model. The estimate of treatment effect is 0.3 with estimated standard
error of 0.11, suggesting the treatment has a significant positive effect on the restricted mean residual life ($p$-value=0.007). These estimates are quite similar to those based on the proportional restricted mean residual life model, where the effect of treatment is 0.29 with estimated standard error of 0.11 ($p$-value=0.009).

Figure 2 displays the estimated functions for the two additive components of the model (baseline and WHO performance status) with 95% point wise confidence intervals.

It is not clear according to Figure 2 if the effect of WHO performance status is significantly time-varying. We thus performed a test based on the test statistic $F_3$, to evaluate whether the baseline and WHO performance status vary with time. This test gives $p$-values of <0.001 and 0.04 for baseline and WHO performance status, respectively based on 5000 simulations of the test statistic, indicating that both effects are time-varying. The overall test $F_2$ yields $p$-values of <0.001 and 0.01, meaning that the effect of baseline and WHO performance status are significantly different from 0 at the conventional 5% significance level. The non-parametric estimation of the restricted mean residual life function was easily calculated for four different combinations of treatment and WHO performance status. It can be obtained by substituting the estimate of survival function $S(t)$ for each combination based on the additive non-parametric Aalen model (Aalen, 1980).
into $m_r(t) = S(t)^{-1} \int_t^\tau S(u)du \ (t < \tau)$. In Figure 3(a), we have shown the non-parametric estimate of the restricted mean residual life function for a patient in IFN group with the best quality of life. In addition, the estimate of the restricted mean residual life function obtained from the additive-multiplicative restricted mean residual life model and from the proportional restricted mean residual life model are given. It indicates that the suggested additive-multiplicative restricted mean residual life model fits well as there is not considerable difference between the estimate of the restricted mean residual life function based on the mentioned model and the fully non-parametric estimate. As we also expected the proportional restricted mean residual life model does not fit well since the estimate of the restricted mean residual life function based on this model has substantially larger difference than the other estimates. Similarly, the other three combinations of treatment and WHO performance status depict the additive-multiplicative restricted mean residual life model fits better than the proportional restricted mean residual life model (not shown here).

Finally, we have illustrated the behavior of estimation of the restricted mean residual life function based on the final model (additive-multiplicative restricted mean residual life) for a patient in IFN group and a patient in MPA group with restriction in activities of daily living in Figure 3(b). It displays that a patient in
IFN group had consistently longer mean residual life within the next 4 years than that in MPA group with restriction in activities of daily living. It also suggests that the restricted mean residual life function increases for the initial period \([0, 1.5]\) and then decreases for the mentioned groups of patients. Hence, if the patient survives the first year, the prognosis is better than compared to what it was at the beginning of the time period.

7 Discussion

The mean residual life can be influenced by the tail behavior of underlying distributions of survival times. When the underlying survival time has an extremely long tail and censored early, as in the case of long-term survivors on cancer treatment, the mean residual life function is not estimable on the whole positive real line without extra assumptions. The restricted mean residual life might then be a useful alternative. Many authors studied the mean residual life function, but it
seems to be no work on the restricted mean residual life function. An alternative strategy, instead of modeling the mean residual life, is to model the median or any other quantile of the residual life. Some researches have already been done in this field, for example by Jeong et al. (2008), Jung et al. (2009) and Ma and Yin (2010).

Although model (3) is very flexible with time-varying coefficients in additive part of the model, in many practical settings, it is of interest to investigate if some of the regression coefficients do not depend on time. When some of the non-parametric components do not appear to depend on time, a possible extension of model (3) is

$$m_r(t \mid X, Z) = \{X_1^T \alpha_r(t) + X_2^T \gamma_r(t)\} \exp(Z^T \beta_r),$$

where $X = (X_1, X_2)$, $\gamma_r$ is a vector of unknown regression coefficients that does not depend on time and $f_r(t)$ is a known and specific function with $f_r(\tau) = 0$. This model involves a semi-parametric additive baseline, and can be analyzed similarly to what has been proposed in this paper.

In many clinical studies, patients will often be at risk from more than one event, which are either mutually exclusive (competing risks) or occur sequentially (multi-state models). These models are very useful for describing event history data, affording a better understanding of the disease process, and leading to a better knowledge of the evolution of the disease over time. In illness-death multi-state model (with healthy, illness and death states), there is a lot of interest in estimating the expected remaining number of years spent in health. Further research is needed to model expected remaining number of years spent in health and assess how covariates affect this model in the presence of potential censoring. This is a potential topic for our future work.

**Acknowledgement**

This work was supported by the Ministry of Science, Research and Technology of Iran.
Appendix

Conditions

Let $H_1(t)$ and $H_2(t)$ denote the quantities defined in the text. Also let $h_1(t)$ and $h_2(t)$ denote their limits, respectively. In order to study the asymptotic properties of the proposed estimators, we need the following regularity conditions:

Condition A1. $P(C \geq \tau) > 0$, and $N(\tau)$ is bounded almost surely.

Condition A2. The covariates $X$ and $Z$ are bounded.

Condition A3. $\alpha_0(t)$ is right continuous with left-hand limits, and has bounded total variation on $[0, \tau]$.

Condition A4. The matrix

$$D(\tau, \beta) = -E \left[ \int_0^\tau \{Z_i(t) - h_1(t)X_i(t) - h_2(t)X_i(t)\} \exp(-Z_i^T \beta)Z_i^T(t)dt \right],$$

is non-singular.

Proof of Theorem 1

The first step is to show that $\hat{\alpha}(t; \beta)$ for given $\beta$ converges uniformly in probability to a limit, say $\alpha(t)$. It follows from (9) that

$$\hat{\alpha}(t; \beta) - \alpha(t) = \int_t^\tau \left\{ X^T(s)X(s) \right\}^{-1} X^T(s) \exp(-Z\beta) ds$$

$$- \int_t^\tau \left\{ X^T(s)X(s) \right\}^{-1} X^T(s)d\tilde{N}(s)X(s)\hat{\alpha}(s^+; \beta) + \int_t^\tau d\alpha(s)$$

$$= \int_t^\tau \left\{ X^T(s)X(s) \right\}^{-1} X^T(s)d\tilde{N}(s)X(s)\alpha(s)$$

$$- \int_t^\tau \left\{ X^T(s)X(s) \right\}^{-1} X^T(s)d\tilde{M}(s)X(s)\alpha(s)$$

$$- \int_t^\tau \left\{ X^T(s)X(s) \right\}^{-1} X^T(s)d\alpha(s)$$

$$- \int_t^\tau \left\{ X^T(s)X(s) \right\}^{-1} X^T(s)d\tilde{N}(s)X(s)\hat{\alpha}(s^+; \beta) + \int_t^\tau d\alpha(s),$$
A.1. By the uniform strong law of large numbers (Pollard, 1990, p. 41), it can be proved that the process

\[ \alpha(t; \beta) - \alpha(t) = -Q(t)^{-1} \int_0^t Q(s^+) \{ X^T(s)X(s) \}^{-1} X^T(s)d\tilde{M}(s)X(s)\alpha(s). \]

Now, recall that \( Q(t) = \prod_{t, \tau} [I + \{ X^T(s)X(s) \}^{-1} X^T(s)d\tilde{N}(s)X(s)] \). Assuming \( Q(t) \) converges uniformly in probability to a deterministic function \( q(t) \) and following the Lemma of Lin (2000, p. 45) since \( Q(t) \) is monotone along with uniform weak law of large numbers, it can be obtained that

\[
\hat{\alpha}(t; \beta_0) - \alpha_0(t) = -n^{-1}q(t)^{-1} \sum_{i=1}^n \int_0^t q(s) \omega(s)X_i(s)X_i^T(s)\alpha_0(s)dM_i(s) + o_p(1),
\]

where \( \omega(s) \) is the limit in probability of \( n \{ X^T(s)X(s) \}^{-1} \). Here and subsequently, the \( o_p(1) \) is uniform in \( t \). By Lenglart’s inequality (Andersen et al., 1993, p. 86), the above term converges uniformly in probability to zero. Hence (15) implies that \( \hat{\alpha}(t; \beta_0) \) converges almost surely to \( \alpha(t; \beta_0) = \alpha_0(t) \) in \( t \in [0, \tau] \).

The next step is to show that the proposed estimator \( \hat{\beta} \) is unique and consistent. Denote minus the derivative of \( U(\tau; \beta; \hat{\alpha}(t; \beta)) \) with respect to \( \beta \) by \( I(\tau, \beta) \), where

\[ I(\tau, \beta) = -\int_0^\tau \left\{ Z^T(t) - H_1(t)X^T(t) \right\} \left[ d\tilde{N}(t)X(t) \frac{\partial \hat{\alpha}(t; \beta)}{\partial \beta} + \text{diag} \{ \exp(-Z_i^T \beta) \} Z(t)dt \right]. \]

Substituting the derivative of \( \hat{\alpha}(t; \beta) \) in (10) with respect to \( \beta \) into the foregoing equation and then interchanging the order of integration, we get

\[ n^{-1}I(\tau, \beta) = -n^{-1} \int_0^\tau \left\{ Z^T(t) - H_1(t)X^T(t) - H_2(t)X(T(t)) \right\} \text{diag} \{ \exp(-Z_i^T \beta) \} Z(t)dt. \]

It can be proved that \( H_1(t) \) and \( H_2(t) \) are bounded variation and thus can be written as the difference of two increasing functions. Hence the process \( I(\tau, \beta) \) can be written as sums or products of monotone functions in \( t \) and all components of \( \beta \) and therefore it is manageable (Pollard, 1990, p. 38; Bilias et al., 1997, Lemma A.1). By the uniform strong law of large numbers (Pollard, 1990, p. 41), it
follows that $n^{-1}I(\tau, \beta)$ converges almost surely to a non-random function $D(\tau, \beta)$ uniformly in $\beta$ where
\[
D(\tau, \beta) = -E \left[ \int_{0}^{\tau} \{ Z(t) - h_1(t)X_1(t) - h_2(t)X_2(t) \} \exp(-Z(t)\beta)Z(t)dt \right],
\]
with $h_1(t)$ and $h_2(t)$ being the limit in probability of $H_1(t)$ and $H_2(t)$, respectively.

It can also be checked that $n^{-1}U(\tau, \beta_0) \to 0$ almost surely, and $D(\tau, \beta)$ is non-singular by condition (A4). Therefore, the uniform convergence of $n^{-1}I(\tau, \beta)$ and the continuity of $D(\tau, \beta)$ imply that there exists a small neighborhood of $\beta_0$ in which $n^{-1}I(\tau, \beta)$ is non-singular when $n$ is large enough. Hence it follows from the inverse function theorem (Rudin, 1976, p. 221) that within a small neighborhood of $\beta_0$, there exists a unique solution $\hat{\beta}$ to $U(\tau, \beta) = 0$ for all sufficiently large $n$. Since this neighborhood of $\beta_0$ can be arbitrarily small, the preceding proof also implies that $\beta$ is strongly consistent. It then follows from the uniform convergence of $\hat{\alpha}(t; \beta_0)$ to $\alpha(t; \beta_0)$ that $\hat{\alpha}(t) \equiv \hat{\alpha}(t; \hat{\beta}) \to \alpha(t; \beta_0) \equiv \alpha_0(t)$ almost surely uniformly in $t \in [0, \tau]$.

**Proof of Theorem 2**

(i) Consider a decomposition of $n^{-1/2}U \{ \tau, \beta_0; \hat{\alpha}(t; \beta_0) \}$ of the form
\[
n^{-1/2}U \{ \tau, \beta_0; \hat{\alpha}(t; \beta_0) \} = n^{-1/2}U \{ \tau, \beta_0; \alpha_0(t) \} + n^{-1/2}[U \{ \tau, \beta_0; \hat{\alpha}(t; \beta_0) \} - U \{ \tau, \beta_0; \alpha_0(t) \}].
\]

(16)

The first term on the right side of (16) can be written as the following
\[
n^{-1/2}U \{ \tau, \beta_0; \alpha_0(t) \} = n^{-1/2} \int_{0}^{\tau} \left\{ Z(t) - H_1(t)X(t) \right\} \left\{ d\hat{N}(t)X(t)\alpha_0(t) - \exp(-Z(t)\beta)dt \right\}
\]
\[
= n^{-1/2} \int_{0}^{\tau} \left\{ Z(t) - H_1(t)X(t) \right\} d\hat{M}(t)X(t)\alpha_0(t)
\]
\[
= n^{-1/2} \sum_{i=1}^{n} \int_{0}^{\tau} \left\{ Z_i(t) - H_1(t)X_i(t) \right\} X_i(t)\alpha_0(t) dM_i(t).
\]

By the representation of $\hat{\alpha}(t; \beta_0) - \alpha_0(t)$ in (14), the second term on the right side of (16) is equal to
\[
- n^{-1/2} \int_{0}^{\tau} \left\{ Z(t) - H_1(t)X(t) \right\} d\hat{N}(t)X(t)Q(t)^{-1}
\]
\[
\left[ \int_{0}^{\tau} Q(s^+) \left\{ X^T(s)X(s) \right\}^{-1} X^T(s)d\hat{M}(s)X(s)\alpha_0(s) \right].
\]
Interchanging the order of integration, the foregoing term is equal to
\[-n^{-1/2} \sum_{i=1}^{n} \int_{0}^{\tau} H_2(t)X_i(t)X_i^T(t)\alpha_0(t)dM_i(t).\]

Thus it can be obtained that
\[n^{-1/2}U\{\tau, \beta_0; \hat{\alpha}(t; \beta_0)\} = n^{-1/2} \sum_{i=1}^{n} \int_{0}^{\tau} \{Z_i(t) - H_1(t)X_i(t) - H_2(t)X_i(t)\} X_i^T(t)\alpha_0(t)dM_i(t).\]

By the uniform strong law of large numbers and using the Lemma of Lin (2000, p. 45) since \(H_1(t)\) and \(H_2(t)\) are bounded variation, \(n^{-1/2}U\{\tau, \beta_0; \hat{\alpha}(t; \beta_0)\}\) can be decomposed as a sum of independent and identically distributed terms
\[n^{-1/2}U\{\tau, \beta_0; \hat{\alpha}(t; \beta_0)\} = n^{-1/2} \sum_{i=1}^{n} \xi_i + o_p(1), \quad (17)\]

where
\[\xi_i = \int_{0}^{\tau} \{Z_i(t) - h_1(t)X_i(t) - h_2(t)X_i(t)\} X_i^T(t)\alpha_0(t)dM_i(t).\]

Hence by the multivariate central limit theorem, \(n^{-1/2}U\{\tau, \beta_0; \hat{\alpha}(t; \beta_0)\}\) converges in distribution to zero-mean normal distribution whose variance-covariance matrix \(\Sigma = E(\xi_i^{\otimes 2})\) can be consistently estimated by \(\hat{\Sigma} = n^{-1} \sum_{i=1}^{n} \xi_i^{\otimes 2}\), defined in Theorem 2(i).

(ii) A Taylor series expansion of the score function (11) around \(\hat{\beta}\) gives
\[n^{1/2}(\hat{\beta} - \beta_0) = \{n^{-1}I(\tau, \beta_0)\}^{-1} n^{-1/2}U\{\tau, \beta_0; \hat{\alpha}(t; \beta_0)\},\]

where \(\beta_0\) is on the line segment between \(\beta_0\) and \(\hat{\beta}\). To prove asymptotic normality of \(n^{1/2}(\hat{\beta} - \beta_0)\), it suffices to prove weak convergence of \(n^{-1/2}U\{\tau, \beta_0; \hat{\alpha}(t; \beta_0)\}\) to a Gaussian process and to prove convergence in probability of \(n^{-1}I(\tau, \beta_0)\) to a non-singular matrix. From the uniform convergence of \(n^{-1}I(\tau, \beta)\) to \(D(\tau, \beta)\) along with the consistency of \(\hat{\beta}\) and expression (17), asymptotic approximation for \(n^{1/2}(\hat{\beta} - \beta_0)\) can be displayed by
\[n^{1/2}(\hat{\beta} - \beta_0) = n^{-1/2}D(\tau, \beta_0)^{-1} \sum_{i=1}^{n} \xi_i + o_p(1). \quad (18)\]

Thus, it follows that \(n^{1/2}(\hat{\beta} - \beta_0)\) is asymptotically normal with mean zero and covariance matrix \(D(\tau, \beta_0)^{-1}\hat{\Sigma}D(\tau, \beta_0)^{-1}\), which can be consistently estimated by \(D(\tau, \hat{\beta})^{-1}\hat{\Sigma}D(\tau, \hat{\beta})^{-1}\) as defined in Theorem 2(ii).
Proof of Theorem 3

To show the weak convergence of \(n^{1/2}\{\hat{\alpha}(t) - \alpha_0(t)\}\), we first note that

\[
n^{1/2}\{\hat{\alpha}(t) - \alpha_0(t)\} = n^{1/2}\{\hat{\alpha}(t; \hat{\beta}_0) - \alpha_0(t)\} + n^{1/2}\{\hat{\alpha}(t; \hat{\beta}) - \hat{\alpha}(t; \hat{\beta}_0)\}.
\]

It follows from (15) that

\[
n^{1/2}\{\hat{\alpha}_n(t; \hat{\beta}_0) - \alpha_0(t)\} = n^{-1/2} \sum_{i=1}^n \zeta_i(t) + o_p(1),
\]

where

\[
\zeta_i(t) = -q(t)^{-1} \int_t^\tau q(s)\omega(s)X_i(s)X_i^T(s)\alpha_0(s)dM_i(s).
\]

Taking the Taylor expansion of \(\hat{\alpha}(t; \hat{\beta})\), together with the consistency of \(\hat{\beta}\) and the uniform strong law of large numbers, we have

\[
n^{1/2}\{\hat{\alpha}(t; \hat{\beta}) - \hat{\alpha}(t; \beta_0)\} = G(t; \beta_0) n^{1/2}(\hat{\beta} - \beta_0) + o_p(1),
\]

where \(G(t; \beta)\) denotes the limit in probability of \(\partial\hat{\alpha}(t; \beta)/\partial\beta\). Therefore, it follows from (18) that uniformly in \(t \in [0, \tau]\),

\[
n^{1/2}\{\hat{\alpha}(t) - \alpha_0(t)\} = n^{-1/2} \sum_{i=1}^n \phi_i(t) + o_p(1), \tag{19}
\]

where

\[
\phi_i(t) = \zeta_i(t) + G(t; \beta_0)D(\tau, \beta_0)^{-1}\xi_i,
\]

are independent and identically distributed zero-mean random variables for each \(t\). By the multivariate central limit theorem, \(n^{1/2}\{\hat{\alpha}(t) - \alpha_0(t)\}\) converges in finite-dimensional distribution to a zero-mean Gaussian process for \(0 \leq t \leq \tau\). The processes \(\{\zeta_i(t); i = 1, \ldots, n\}\) can be written as sums or products of monotone functions of \(t\), since any function of bounded variation can be expressed as the difference of two increasing functions. Therefore they are manageable and using the functional central limit theorem (Pollard, 1990, p. 53), the first term on the right-hand side of (19) is tight. The second term is also tight because \(n^{-1/2} \sum_{i=1}^n \xi_i\) converges in distribution and \(G(t; \beta_0)\) is a deterministic function. Thus, \(n^{1/2}\{\hat{\alpha}(t) - \alpha_0(t)\}\) is tight and converges weakly to a zero-mean Gaussian process whose covariance function \(\Gamma(s, t) = E\{\phi(t)\phi(t)\}\) can be consistently estimated by \(\hat{\Gamma}(s, t)\) defined in Theorem 3.
 proves that $n^{1/2} \{ \hat{m}_r(t | X_0, Z_0) \}$ can be written in terms of a sum of independent and identically distributed zero-mean decompositions. For independent standard normal samples $(\Omega_1, \ldots, \Omega_n)$, following the resampling approach as in Lin et al. (2000), the asymptotic distributions of $\Delta_1 = n^{-1/2} \hat{D}(\tau, \hat{\beta})^{-1} \sum_{i=1}^n \hat{\xi}_i \Omega_i$ and $\Delta_2(t) = n^{-1/2} \sum_{i=1}^n \hat{\phi}_i(t) \Omega_i$ are equivalent to the asymptotic distributions of $n^{1/2} (\hat{\beta} - \beta_0)$ and $n^{1/2} \{ \hat{\alpha}(t) - \alpha_0(t) \}$, respectively. Therefore applying the functional delta method to these asymptotically equivalent processes leads to a resampling version of the restricted mean residual life function estimator. The term on the right side of equation (20) is asymptotically equivalent to $n^{1/2} \sum_{i=1}^n \epsilon_i(t)$, where $\epsilon_i(t) = \exp(Z_0^T \hat{\beta}) \left\{ X_0^T \phi_i(t) + X_0^T \alpha_0(t) Z_0^T \hat{D}(\tau, \hat{\beta})^{-1} \hat{\xi}_i \right\}$. Define

$$
\hat{\epsilon}_i(t) = \exp(Z_0^T \hat{\beta}) \left\{ X_0^T \hat{\phi}_i(t) + X_0^T \hat{\alpha}(t) Z_0^T \hat{D}(\tau, \hat{\beta})^{-1} \hat{\xi}_i \right\}.
$$

It can be derived that under regularity conditions $n^{1/2} \{ \hat{m}_r(t | X_0, Z_0) - m_r(t | X_0, Z_0) \}$ has the same asymptotic distribution as $\Delta_{m_r}(t) = n^{-1/2} \sum_{i=1}^n \hat{\epsilon}_i(t) \Omega_i$.

References


13/01 Forman, J.L. & Sørensen M. A new approach to multi-modal diffusions with applications to protein folding.


13/03 Wolbers, M., Koller, M.T., Witteman, J.C.M. & Gerds, T.A. Concordance for prognostic models with competing risks.

13/04 Andersen, P.K., Canudas-Romo, V. & Keiding, N. Cause-specific measures of life lost.

13/05 Keiding, N. Event history analysis.


13/07 Hansen, S., Andersen, P.K. & Parner, E. Events per variable for risk differences and relative risks using pseudo-observations.


13/11 Olsbjerg, M. & Christensen K.B. SAS macro for marginal maximum likelihood estimation in longitudinal polytomous Rasch models.

14/01 Ambrogi, F. & Andersen, P.K. Predicting Smooth Survival Curves through Pseudo-Values.

14/02 Olsbjerg, M. & Christensen, K.B. Modeling local dependence in longitudinal IRT models.

14/03 Olsbjerg, M. & Christensen, K.B. LIRT: SAS macros for longitudinal IRT models.

14/04 Jacobsen, M. & Martinussen, T. A note on the large sample properties of estimators based on generalized linear models for correlated pseudo-observations.

14/05 De Neve, J. & Gerds, T. A note on the interpretation of the Cox regression model.

14/06 Gerds, TA. The Kaplan-Meier theater.

14/07 Martinussen, T., Holst, K.K. & Scheike, T. Cox regression with missing covariate data using a modified partial likelihood method.

15/01 Mansourvar, Z., Martinussen, T., Scheike, T. Semiparametric regression for restricted mean residual life under right censoring.

15/02 Mansourvar, Z., Martinussen, T., Scheike, T. An additive-multiplicatve restricted mean residual life model.