Lifetime Data Anal (2010) 16:385–408
DOI 10.1007/s10985-010-9152-2

# Goodness-of-fit tests for additive mean residual life model under right censoring

**Zhigang Zhang · Xingqiu Zhao · Liuquan Sun**

Received: 8 April 2009 / Accepted: 8 January 2010 / Published online: 26 January 2010


**Abstract** The mean residual life (MRL) measures the remaining life expectancy and is useful in actuarial studies, biological experiments and clinical trials. To assess the covariate effect, an additive MRL regression model has been proposed in the literature. In this paper, we focus on the topic of model checking. Specifically, we develop two goodness-of-fit tests to test the additive MRL model assumption. We explore the large sample properties of the test statistics and show that both of them are based on asymptotic Gaussian processes so that resampling approaches can be applied to find the rejection regions. Simulation studies indicate that our methods work reasonably well for sample sizes ranging from 50 to 200. Two empirical data sets are analyzed to illustrate the approaches.

**Keywords** Additive mean residual life model · Right-censored survival data · Gaussian process · Goodness-of-fit test

Z. Zhang (✉)
Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center,
307 East 63rd Street, New York, NY 10065, USA
e-mail: zhangz@mskcc.org

X. Zhao
Department of Applied Mathematics, The Hong Kong Polytechnic University, Room HJ625,
Core J, Stanley Ho Building, Kowloon, Hong Kong
e-mail: xingqiu.zhao@polyu.edu.hk

L. Sun
Institute of Applied Mathematics, Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, Beijing 100190, People's Republic of China
e-mail: slq@amt.ac.cn

## 1 Introduction

The mean residual life function (MRLF), $m(t)$, defined at time $t$ for a nonnegative survival time $T$ with finite expectation, is

$$m(t) = E(T - t|T > t).$$

This function essentially measures the remaining life expectancy of a subject when the subject has survived to time $t$. MRLF has a one-to-one correspondence with the survival function and the hazard function of $T$. For example, $m(t) = \frac{\int_t^\infty S(u)du}{S(t)}$ where $S(t)$ is the survival function of $T$, i.e., $S(t) = P(T > t)$. Therefore, MRLF characterizes the stochastic behavior of a survival variable $T$ theoretically and provides practical usage in research fields such as clinical trials and reliability or actuarial studies, etc (Cox 1961).

Many probabilistic properties of MRLF have been investigated. Cox (1961) gave a general introduction of MRLF. For properties including characterization and representation theorems, limiting behaviors and multivariate versions, see Swartz (1973), Balkema and de Haan (1974), Hollander and Proschan (1975) and Kotz and Shanbhag (1980). More recent works towards statistical applications can be found in Hall and Wellner (1984), Arnold and Zahedi (1988) and Oakes and Dasu (1990).

The work of Oakes and Dasu (1990) also contained a proportional or multiplicative MRL model, which was used by the authors to describe the effect of a dichotomous covariate. Their model takes the form

$$m_1(t) = \psi m_0(t), \tag{1}$$

where $m_r(t)$ is the MRLF associated with group $r, r = 0, 1$. This model was later generalized by Maguluri and Zhang (1994) to accommodate continuous covariates as follows

$$m(t|Z) = m_0(t) \exp(\beta' Z). \tag{2}$$

In the above, $m(t|Z)$ denotes the MRLF corresponding to the $p \times 1$ vector covariate $Z$, $m_0(t)$ is some unknown baseline MRLF, and $\beta$ is an unknown $p \times 1$ vector of regression parameters. Chen et al. (2005) extended the estimation procedure of Maguluri and Zhang (1994) to accommodate right-censored survival data. Another inference procedure for model (2) was also proposed by Chen and Cheng (2005).

Just as the additive hazards regression model is an attractive alternative to the multiplicative hazards regression model, the additive MRL model has also been proposed as an alternative to (2). This model takes the form

$$m(t|Z) = m_0(t) + \beta' Z, \tag{3}$$

in which all components are defined similarly as in (2). Model (3) was first considered in Chen and Cheng (2006). In that paper and a later paper by Chen (2007) the authors

proposed several estimation procedures. When right-censored data are present, either a Buckley–James approach or a martingale-based estimating equation can be used for estimating $\beta_0$, the true value of $\beta$.

A more general class of regression models, termed as the transformed MRL models, were studied by Sun and Zhang (2009). This general class of models contain (2) and (3) as special cases. In that paper the authors applied the inverse probability censoring weighting technique and proposed a class of generalized estimating equations for the regression parameter and the baseline MRLF. Via numerical studies, they also showed that their estimators are more efficient than those of Chen and Cheng (2005, 2006) under independent censoring and other mild conditions.

Sometimes one is also interested in checking model adequacy. For regression models utilizing the hazard function, there exist many approaches in the literature. See Fleming and Harrington (1991) and Kalbfleisch and Prentice (2002) for extensive discussion of this topic. For regression models utilizing the MRLF, the research seems quite limited. Yuen et al. (2003) proposed a goodness-of-fit test for model (2) when there is no censoring involved. In this paper, we propose two goodness-of-fit tests for model (3) based on appropriately constructed stochastic processes which are asymptotically Gaussian when right censoring is present.

The rest of the paper is organized as follows. Section 2 discusses the construction of these two goodness-of-fit tests. Section 3 reports some results from simulation studies conducted for evaluating the proposed methods. Two real-life data sets are analyzed using our methods in Section 4. Section 5 studies the local alternatives and power of the tests. Section 6 contains some concluding remarks. All technical proofs are summarized in the Appendix.

## 2 Construction of test statistics

Let $T$ and $Z$ be defined as before, and $C$ be the censoring time. Suppose that $\{T_i, C_i, Z_i; i = 1, \ldots, n\}$ are independent replicates of $\{T, C, Z\}$. When right censoring is present, we observe $\{X_i, \Delta_i, Z_i; i = 1, \ldots, n\}$, where $X_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$ and $I(\cdot)$ is the indicator function. In this section, we focus on goodness-of-fit tests for the additive MRL model (3) and make no attempt to develop new estimators for $\beta_0$. Methods for estimating $\beta_0$ are available in Chen and Cheng (2006), Chen (2007) and Sun and Zhang (2009).

We now establish some additional notation and assumptions. Let $H_1(t, z) = P\{X_i \leq t, Z_i \leq z, \Delta_i = 1\}$ and $H(t, z) = P\{X_i > t, Z_i \leq z\}$, where the notation "$Z \leq z$" means that each component of $Z$ is less than or equal to the corresponding component of $z$ when both sides are vectors. Here $Z$ is assumed to be bounded (see C2 in the Appendix A.1) and thus we may take $Z$ to be nonnegative with a finite upper bound without loss of generality. Also define $G(t)$ to be the survival function of the censoring time $C$, i.e., $G(t) = P(C > t)$. To avoid lengthy technical discussion of the tail behavior of the limiting distributions, we assume that $P\{C > \tau\} > 0$, where $0 < \tau = \inf\{t : P(T \geq t) = 0\} < \infty$. This assumption essentially says that the support of the censoring time is greater than that of the survival time, a necessary condition for MRLF to be identifiable (Sun and Zhang 2009). We further assume that

the censoring time $C$ is independent of $T$ and $Z$. Remarks on this assumption will be given in the discussion.

To construct the test statistics, we shall use the following equality

$$m_0(t) = \frac{1}{H(t,z)} \left\{ \int_t^\tau \frac{(s-t)G(t)}{G(s)} H_1(ds, z) - \int_0^z \beta_0' w H(t, dw) \right\}, \tag{4}$$

provided that $H(t, z)$ is not zero (note that because of the assumption $P(C > \tau) > 0$, $G(s)$ is always positive when $s \le \tau$). To see (4), we start with the right-hand-side

$$\begin{aligned}
&\frac{1}{H(t,z)} \left\{ \int_t^\tau \frac{(s-t)G(t)}{G(s)} H_1(ds, z) - \int_0^z \beta_0' w H(t, dw) \right\} \\
&= \frac{1}{H(t,z)} \left\{ \int_t^\tau (s-t)G(t) \int_0^z P(s < T \le s + ds, w < Z \le w + dw) \right. \\
&\quad \left. - \int_0^z \beta_0' w H(t, dw) \right\} \\
&= \frac{1}{H(t,z)} \left\{ P(C > t) \int_0^z m(t|Z = w) P(T > t|Z = w) P(w < Z \le w + dw) \right. \\
&\quad \left. - \int_0^z \beta_0' w H(t, dw) \right\} \\
&= \frac{1}{H(t,z)} \left\{ \int_0^z m(t|Z = w) P(X > t, w < Z \le w + dw) - \int_0^z \beta_0' w H(t, dw) \right\} \\
&= \frac{1}{H(t,z)} \left\{ \int_0^z m_0(t) H(t, dw) \right\} = m_0(t).
\end{aligned}$$

In the above, the second-to-last equality is due to the model assumption of (3) and $\int_0^z$ stands for $\int_0^{z^{(1)}} \dots \int_0^{z^{(p)}}$ where $z^{(j)}$, $j = 1, \dots, p$ is the $j$th component of the $p \times 1$ vector $z$. Denoting the right-hand side of (4) by $V(t, z)$, we shall test the null hypothesis

$$H_0 : m_0(t) = V(t, z) \tag{5}$$

for all $z$ and all $t$. In other words, when model (3) is indeed correct, $V(t, z)$ should be independent of $z$ for all $t$.

Motivated by this fact, as a measure of fit for model (3), we first estimate $V(t,z)$ by

$$V_n(t,z) = \frac{1}{H_n(t,z)} \left\{ \int_t^\tau \frac{(s-t)\hat{G}(t)}{\hat{G}(s)} H_{1n}(ds,z) - \int_0^z \hat{\beta}' w H_n(t,dw) \right\}, \quad (6)$$

where $\hat{G}(t)$ is the Kaplan–Meier estimator of $G(t)$ based on $\{X_i, 1-\Delta_i; i=1,\ldots,n\}$, $\hat{\beta}$ is a $n^{1/2}$-consistent and regular estimator of $\beta_0$ (some remarks are given at the end of this section) and $H_n$ and $H_{1n}$ are the empirical counterparts of $H$ and $H_1$, respectively. That is, $H_n(t,z) = n^{-1}\sum_{i=1}^n I(X_i > t, Z_i \le z)$ and $H_{1n}(t,z) = n^{-1}\sum_{i=1}^n I(X_i \le t, Z_i \le z, \Delta_i = 1)$. To ensure that the denominator $H_n(t,z)$ is not zero, at least with a probability converging to 1 asymptotically, we will confine $t \in \mathcal{T}$ where $\mathcal{T}$ is a compact subset of $[0,\tau]$ and $z \in \mathcal{Z}$ where $\mathcal{Z}$ is a compact subset of the range of $Z$ such that $H(t,z)$ is bounded away from 0 on $\mathcal{T}\times\mathcal{Z}$.

Define the process

$$\theta_n(t,z) = n^{1/2}\{V_n(t,z) - V_n(t,z_u)\}, \quad (7)$$

where $z_u$ is a vector of upper bounds for $\mathcal{Z}$. Here the role of the upper bound $z_u$ can be roughly regarded as a "reference level". Theoretically speaking, it can be replaced by any other value that $Z$ may take because under $H_0$, $V(t,z)$ does not depend on $z$. The upper bound $z_u$ is used only for computational convenience and in reality, may simply be chosen to be $(z_u^{(1)},\ldots,z_u^{(p)})'$ where $z_u^{(j)}$ is the largest observed $j$th entry of $Z$ in the sample.

Based on (7), two types of test statistics may be used to check the goodness-of-fit of model (3): the Kolmogorov-Smirnov (KS) type

$$\mathcal{F}_n^{(1)} = \sup_{t\in\mathcal{T};z\in\mathcal{Z}} |\theta_n(t,z)|,$$

and the Cramér–von Mises (CvM) type

$$\mathcal{F}_n^{(2)} = \int_{\mathcal{Z}} \int_{\mathcal{T}} \theta_n(t,z)^2 H_n^0(dt,dz),$$

where $H_n^0$ is the empirical distribution function of $(X_i, Z_i)$. The null hypothesis will be rejected for large values of $\mathcal{F}_n^{(1)}$ or $\mathcal{F}_n^{(2)}$, which indicate serious discrepancy between $V_n(t,z)$ and $V_n(t,z_u)$ for some $z$ at certain $t$. The rationale behind these two test statistics is that, under $H_0$, (7) can be rewritten as

$$\begin{aligned}\theta_n(t,z) &= n^{1/2}\{V_n(t,z) - V(t,z)\} - n^{1/2}\{V_n(t,z_u) - V(t,z_u)\} \\ &= \phi_n(t,z) - \phi_n(t,z_u) \end{aligned} \quad (8)$$

where $\phi_n(t,z) := n^{1/2}\{V_n(t,z) - V(t,z)\}$ can be regarded as the standardized mean residual life process. This standardization is important for studying the asymptotic

properties of $\theta_n(t,z)$. As shown in the Appendix A.2, under model (3) (and hence $H_0$ is true), $\theta_n(t,z)$ converges to a zero-mean Gaussian process $W(t,z)$ whose covariance function at $(t,z)$ and $(t^*,z^*)$ can be estimated consistently by $\hat{\sigma}(t,z;t^*,z^*) = n^{-1}\sum_{i=1}^n \hat{\eta}_i(t,z)\hat{\eta}_i(t^*,z^*)$, where $\hat{\eta}_i(t,z) = \hat{\xi}_i(t,z) - \hat{\xi}_i(t,z_u)$, and

$$\begin{aligned}\hat{\xi}_i(t,z) = {} & \frac{\Delta_i(X_i - t)\hat{G}(t)}{H_n(t,z)\hat{G}(X_i)} I(X_i > t, Z_i \le z) \\ & + \frac{\hat{G}(t)}{H_n(t,z)} \int_t^\tau \int_u^\tau \frac{s-t}{\hat{G}(s)\hat{\pi}(u)} H_{1n}(ds,z) d\hat{M}_i^c(u) \\ & - \frac{V_n(t,z)}{H_n(t,z)} I(X_i > t, Z_i \le z) - \frac{1}{H_n(t,z)} \int_0^z w' H_n(t,dw)\hat{A}^{-1}\hat{\psi}_i \\ & - \frac{\hat{\beta}' Z_i}{H_n(t,z)} I(X_i > t, Z_i \le z). \end{aligned} \tag{9}$$

The definition of $\hat{M}_i^c(t)$, $\hat{\pi}(t)$, $\hat{A}$ and $\hat{\psi}_i$ can be found in the Appendix A.1. Consequently, $\mathcal{F}_n^{(1)}$ and $\mathcal{F}_n^{(2)}$ converges in distribution to $F^{(1)}$ and $F^{(2)}$, respectively, where

$$\begin{aligned} F^{(1)} &= \sup_{t\in\mathcal{T};z\in\mathcal{Z}} |W(t,z)|, \\ F^{(2)} &= \int_{\mathcal{Z}} \int_{\mathcal{T}} W(t,z)^2 H^0(dt,dz), \end{aligned}$$

and $H^0$ is the distribution function of $(X,Z)$.

Obviously, the complicated structure of the expression (9) does not allow for an analytic treatment of the involved distributions. We thus seek for a more computationally feasible alternative that depends on resampling. According to the arguments of Lin et al. (2000), the distribution of the process $W(t,z)$ can be approximated by that of the zero-mean Gaussian process $\widetilde{W}(t,z)$, where

$$\widetilde{W}(t,z) = n^{-1/2}\sum_{i=1}^n \hat{\eta}_i(t,z)\Omega_i,$$

and $(\Omega_1,\ldots,\Omega_n)$ are independent standard normal variables which are independent of the data $\{X_i, \Delta_i, Z_i; i = 1,\ldots,n\}$. Thus, the distributions of $F^{(1)}$ and $F^{(2)}$ can be approximated by $\tilde{F}^{(1)}$ and $\tilde{F}^{(2)}$, respectively, where

$$\tilde{F}^{(1)} = \sup_{t\in\mathcal{T};z\in\mathcal{Z}} |\widetilde{W}(t,z)|,$$

and

$$\tilde{F}^{(2)} = \int_{\mathcal{Z}} \int_{\mathcal{T}} \widetilde{W}(t,z)^2 H_n^0(dt, dz).$$

To approximate the distributions of $F^{(k)}(k = 1, 2)$, we can obtain a large number of realizations from $\tilde{F}^{(k)}$ by repeatedly generating the normal random sample $(\Omega_1, \ldots, \Omega_n)$ while fixing the data $\{X_i, \Delta_i, Z_i; i = 1, \ldots, n\}$ at their observed values, and determine approximate critical values of the two tests using the simulated observations.

We now briefly discuss the estimator $\hat{\beta}$. Theoretically, any $n^{1/2}$-consistent, regular estimator for $\beta_0$ can be used. Obviously, however, it is beneficial to choose an efficient one. So far in the literature three estimation approaches for the additive MRL model have been proposed: Chen and Cheng (2006), Chen (2007) and Sun and Zhang (2009). The method in Chen (2007) does not take into account right censoring and thus cannot be used here. The method in Sun and Zhang (2009) is developed for a general class of transformation models which include model (3) as a special case. In particular, under model (3) their estimating equations simplify to a closed form and do not require iteration or estimation of the baseline MRLF $m_0(t)$. Moreover, simulation results suggest that their estimators are more efficient than that of Chen and Cheng (2006) when the censoring time is independent of the survival time and the covariates (which is assumed in this paper). Therefore, we choose to adopt the method in Sun and Zhang (2009) for deriving $\hat{\beta}$. The computational details are given in the Appendix A.1.

## 3 Simulation studies

We conducted several simulation studies to assess the performance of the two goodness-of-fit tests discussed in Sect. 2. We first consider the scenario when a single covariate is involved. To generate simulated data, we let $Z \sim$ Bernoulli$\{0, 1\}$ with probability 0.5 and $C$ follow an exponential distribution with mean $V$, with $V$ varying to yield different censoring percentages. To evaluate the empirical size of our tests, we generate $T$ according to model (3), that is, the null hypothesis (5) is true. To evaluate the empirical power of our tests, we generate $T$ according to model (2), i.e., the proportional MRL model. It can be easily shown that, under model (2), the null hypothesis (5) is no longer true when $\beta_0 \neq 0$. Specifically, the true $\beta_0$ is chosen to be 0.5 and the baseline function $m_0(t)$ is taken from the Hall–Wellner family, i.e., $m_0(t) = (D_1 t + D_2)^+$, where $D_1 > -1$, $D_2 > 0$ and $Y^+$ denotes $YI(Y \geq 0)$ for any quantity $Y$. In our simulation studies we consider two members from the Hall–Wellner family. One is that $D_1 = -0.5$ and $D_2 = 0.5$, which results in a uniform distribution $[0, 1]$ when $Z = 0$. The other is that $D_1 = -1/3$ and $D_2 = 1/3$, which corresponds to the survival function $F(t) = (1 - t)^2$, $0 \leq t \leq 1$ when $Z = 0$. This distribution is chosen for its right skewness. When $Z$ is nonzero, the formula $F(t|Z) = \frac{m(0|Z)}{m(t|Z)} \exp\left\{-\int_0^t \frac{du}{m(u|Z)}\right\}$ can be used to compute the conditional survival function of $T$, where $m(t|Z)$ follows

**Table 1** Simulation results: empirical sizes

| $n$ | C(%) | KS test | | | CvM test | | |
|---|---|---|---|---|---|---|---|
| | | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
| *Model* (3) *with* $m_0(t) = (-t/2 + 1/2)^+$ *and* $\beta_0 = 0.5$ | | | | | | | |
| 50 | 5 | 0.014 | 0.056 | 0.104 | 0.009 | 0.041 | 0.073 |
| | 15 | 0.012 | 0.063 | 0.127 | 0.004 | 0.041 | 0.089 |
| | 30 | 0.014 | 0.078 | 0.137 | 0.006 | 0.044 | 0.087 |
| 100 | 5 | 0.006 | 0.052 | 0.102 | 0.009 | 0.048 | 0.099 |
| | 15 | 0.015 | 0.059 | 0.116 | 0.010 | 0.042 | 0.095 |
| | 30 | 0.019 | 0.066 | 0.135 | 0.011 | 0.040 | 0.086 |
| 200 | 5 | 0.009 | 0.046 | 0.097 | 0.009 | 0.053 | 0.094 |
| | 15 | 0.009 | 0.053 | 0.117 | 0.012 | 0.059 | 0.098 |
| | 30 | 0.013 | 0.049 | 0.108 | 0.010 | 0.039 | 0.087 |
| *Model* (3) *with* $m_0(t) = (-t/3 + 1/3)^+$ *and* $\beta_0 = 0.5$ | | | | | | | |
| 50 | 5 | 0.016 | 0.074 | 0.144 | 0.011 | 0.055 | 0.114 |
| | 15 | 0.009 | 0.085 | 0.154 | 0.005 | 0.034 | 0.085 |
| | 30 | 0.014 | 0.081 | 0.161 | 0.007 | 0.041 | 0.080 |
| 100 | 5 | 0.020 | 0.079 | 0.139 | 0.015 | 0.053 | 0.102 |
| | 15 | 0.009 | 0.072 | 0.135 | 0.006 | 0.042 | 0.087 |
| | 30 | 0.019 | 0.085 | 0.153 | 0.006 | 0.046 | 0.093 |
| 200 | 5 | 0.014 | 0.060 | 0.124 | 0.010 | 0.048 | 0.098 |
| | 15 | 0.011 | 0.067 | 0.128 | 0.010 | 0.051 | 0.106 |
| | 30 | 0.016 | 0.079 | 0.136 | 0.008 | 0.045 | 0.092 |

model (2) or (3). Note that these distributions have bounded support (which is always true when $D_1 < 0$ in the Hall–Wellner family). For constructing the test statistics, we take the upper bound $z_u = 1$, $\tau$ to be the largest uncensored survival time and generate 3000 realizations from $\tilde{F}^{(k)}(k = 1, 2)$ for finding the critical values. All simulation results are based on 1000 replications.

Table 1 shows the results for evaluating the empirical sizes of the two tests when the type I error rate $\alpha$ takes different values of 0.01, 0.05 and 0.10. Here the simulated data were generated from model (3) with the two different baseline MRLFs $m_0(t)$ mentioned above and $\beta_0 = 0.5$. We also consider various censoring percentages: 5%, 15% and 30% (denoted by C% in the table). Sample sizes are taken to be 50, 100 or 200, which are usually regarded as small to moderate samples in clinical or medical studies. In Table 1, each entry represents the proportion of cases when the null hypothesis was falsely rejected. It can be seen that, when the baseline MRLF is uniform, i.e., $m_0(t) = (-t/2 + 1/2)^+$, the empirical sizes are quite close to their nominal levels for moderate sample sizes. Besides, the actual levels converge to the true nominal levels when the sample sizes increase and are better when there are less censored observations, as we might expect. Similar patterns are observed when $m_0(t) = (-t/3 + 1/3)^+$, that is, the baseline MRLF is right-skewed. However, we notice that

**Table 2** Simulation results: empirical powers

| $n$ | C(%) | KS test | | | CvM test | | |
|---|---|---|---|---|---|---|---|
| | | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
| *Model* (2) *with* $m_0(t) = (-t/2 + 1/2)^+$ *and* $\beta_0 = 0.5$ | | | | | | | |
| 50 | 5 | 0.531 | 0.833 | 0.923 | 0.460 | 0.805 | 0.905 |
| | 15 | 0.360 | 0.685 | 0.819 | 0.273 | 0.588 | 0.749 |
| | 30 | 0.186 | 0.449 | 0.604 | 0.064 | 0.276 | 0.435 |
| 100 | 5 | 0.992 | 0.999 | 0.999 | 0.961 | 0.995 | 0.999 |
| | 15 | 0.956 | 0.995 | 0.998 | 0.836 | 0.975 | 0.991 |
| | 30 | 0.709 | 0.883 | 0.940 | 0.324 | 0.674 | 0.824 |
| 200 | 5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 30 | 0.993 | 0.999 | 1.000 | 0.859 | 0.978 | 0.997 |
| *Model* (2) *with* $m_0(t) = (-t/3 + 1/3)^+$ *and* $\beta_0 = 0.5$ | | | | | | | |
| 50 | 5 | 0.049 | 0.229 | 0.362 | 0.051 | 0.198 | 0.338 |
| | 15 | 0.042 | 0.170 | 0.299 | 0.032 | 0.157 | 0.270 |
| | 30 | 0.015 | 0.109 | 0.218 | 0.006 | 0.064 | 0.155 |
| 100 | 5 | 0.295 | 0.593 | 0.732 | 0.269 | 0.554 | 0.707 |
| | 15 | 0.189 | 0.474 | 0.602 | 0.131 | 0.387 | 0.564 |
| | 30 | 0.091 | 0.318 | 0.453 | 0.030 | 0.169 | 0.299 |
| 200 | 5 | 0.774 | 0.923 | 0.963 | 0.724 | 0.909 | 0.956 |
| | 15 | 0.646 | 0.875 | 0.937 | 0.527 | 0.786 | 0.875 |
| | 30 | 0.347 | 0.673 | 0.784 | 0.153 | 0.400 | 0.597 |

bigger sample sizes are required for the empirical sizes to converge reasonably to their nominal levels, especially for the KS-type test. As a matter of fact, our experience has been that the more right-skewed the baseline MRLF is, the larger sample sizes one needs to achieve reasonable empirical levels (simulation results not shown here). This may be attributed to the tail instability of the estimators of $H(t, z)$, $G(t)$ and $\pi(t)$ (see (9)), as is commonly encountered in survival analysis. We also notice that the CvM-type test appears to be more conservative than the KS-type test.

Table 2 reports the empirical powers for detecting non-additive MRL models, for which we generated the data from model (2), the proportional MRL model with the same choices of $m_0(t)$ and $\beta_0$. Each entry in the table now represents the proportion of cases when the null hypothesis is correctly rejected. As expected, the powers converge to one when the sample sizes become larger. Again, right-skewed distributions require larger sample sizes for convergence. As in Table 1, the CvM-type test appears to be less powerful than the KS-type test, especially when the censoring percentage is high. Some other set-ups (e.g., larger samples with higher censoring percentages) are also considered but the results are not shown here. See Sect. 4.2 for a discussion.

We also considered situations where two covariates are present. As before $Z_1 \sim$ Bernoulli$\{0, 1\}$ with probability 0.5. The second covariate, $Z_2$, follows a normal

**Table 3** Simulation results: two covariates

| $n$ | C(%) | KS test | | | CvM test | | |
|---|---|---|---|---|---|---|---|
| | | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
| *Empirical sizes : model* (2) *with* $m_0(t) = (-t/2 + 1/2)^+$ *and* $\beta_{10} = \beta_{20} = 0.5$. | | | | | | | |
| 50 | 5 | 0.023 | 0.091 | 0.151 | 0.015 | 0.085 | 0.161 |
| | 15 | 0.014 | 0.080 | 0.160 | 0.018 | 0.090 | 0.169 |
| | 30 | 0.023 | 0.090 | 0.175 | 0.015 | 0.081 | 0.161 |
| 100 | 5 | 0.016 | 0.068 | 0.140 | 0.011 | 0.063 | 0.117 |
| | 15 | 0.007 | 0.064 | 0.134 | 0.018 | 0.077 | 0.133 |
| | 30 | 0.020 | 0.069 | 0.149 | 0.015 | 0.072 | 0.146 |
| 200 | 5 | 0.014 | 0.072 | 0.130 | 0.011 | 0.068 | 0.118 |
| | 15 | 0.011 | 0.053 | 0.119 | 0.011 | 0.055 | 0.113 |
| | 30 | 0.016 | 0.064 | 0.131 | 0.015 | 0.055 | 0.127 |
| *Empirical Power : model* (3) *with* $m_0(t) = (-t/2 + 1/2)^+$ *and* $\beta_{10} = \beta_{20} = 0.5$. | | | | | | | |
| 50 | 5 | 0.221 | 0.499 | 0.649 | 0.157 | 0.491 | 0.686 |
| | 15 | 0.117 | 0.314 | 0.482 | 0.072 | 0.284 | 0.464 |
| | 30 | 0.054 | 0.186 | 0.332 | 0.037 | 0.141 | 0.241 |
| 100 | 5 | 0.764 | 0.933 | 0.974 | 0.782 | 0.957 | 0.982 |
| | 15 | 0.532 | 0.791 | 0.890 | 0.380 | 0.759 | 0.886 |
| | 30 | 0.218 | 0.459 | 0.645 | 0.062 | 0.262 | 0.469 |
| 200 | 5 | 0.996 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 15 | 0.971 | 0.998 | 1.000 | 0.943 | 0.997 | 1.000 |
| | 30 | 0.725 | 0.936 | 0.979 | 0.327 | 0.740 | 0.898 |

distribution with mean 0 and standard deviation 0.1. To ensure $Z_2$ is bounded, we truncate $Z_2$ between $-0.2$ and 0.2. Otherwise the set-up is similar to the single covariate cases. Because the double numerical integration is quite time consuming, we consider only the scenario $m_0(t) = (-t/2 + 1/2)^+$ and re-sample 1000 times here. Results are shown in Table 3. It can be seen that the empirical sizes are still reasonably close to their nominal levels and the empirical powers are acceptable as well, though it seems that larger sample sizes are needed for asymptotic approximation when compared to the single covariate cases.

## 4 Applications

As an illustration, we perform our tests on the following two data sets obtained from clinical trials. In order to visualize the additive assumption as made in (3), we focus on binary covariates so that nonparametric estimates of the MRLF can be easily computed and compared graphically. We will also check the independent censoring assumption by using popular semiparametric regression survival models and log-rank tests.

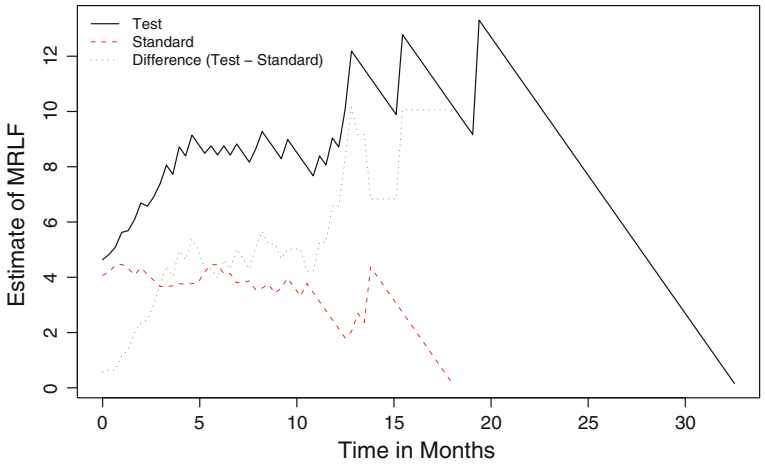


**Fig. 1** Graphical model checking for the Veterans administration lung cancer data: Estimates of the two MRLFs and their difference

### 4.1 Veterans administration lung cancer trial

Our first example is a classical data set that has been analyzed by several authors. Totally 137 advanced lung cancer patients were collected in this trial and were randomized according to one of the two chemotherapeutic agents (standard vs. test). The survival endpoint of this trial is the patient's survival time. Besides the treatment modality, other covariates considered in this trial include the Karnofsky performance score, disease duration, age of the patient, whether or not the patient had prior therapy and tumor cell type. The censoring percentage is 6.6% for this data set. More detailed information and data analysis results can be found in Prentice (1973) and Kalbfleisch and Prentice (2002).

Our objective here is to test whether we can apply model (3) to model the relationship between the survival time ($T$) and the treatment modality ($Z$). We first apply a Cox model to check if the censoring time depends on the treatment. This yields a $p$-value of 0.98 of testing whether the regression coefficient is zero. The log-rank test also gives a $p$-value of 0.99 for testing if the survival distributions of the censoring time for the two treatment modalities are the same. That is, we do not see serious violation of the independent censoring assumption and therefore, we may proceed with the two tests. The KS-type test statistic $\mathcal{F}_n^{(1)}$ is 71.78 and yields a $p$-value of 0.055. The CvM-type test statistic $\mathcal{F}_n^{(2)}$ is 115.06 and yields a $p$-value of 0.067. Here in order to derive the $p$-values, 50,000 re-samplings are used. Both tests suggest that the additive model (3) does not fit the data well.

To further understand why the additive MRL model (3) is not a good choice, we compute and plot the nonparametric estimate of the two MRLFs for the standard group and the test group, respectively. The estimating equation A6 is used for computing these two estimates. Figure 1 shows clearly that the effect of the chemotherapy under test is not additive, for otherwise the difference between the MRLF estimates of the test group and the standard group would be roughly constant.

Interestingly, Fig. 1 implies that there is an upward trend of difference between the test and the standard groups as time increases. Therefore, as suggested by an anonymous reviewer, an extended additive MRL model with time-varying covariate can be applied:

$$m(t|Z) = m_0(t) + \beta' Z + \gamma'(Zt),$$

where $\gamma$ is an unknown parameter to be estimated. Note that the incorporation of this covariate-time interaction can also be used to detect the adequacy of model (3) via the test $H_0 : \gamma_0 = 0$ vs $H_1 : \gamma_0 > 0$. where $\gamma_0$ denotes the true value of $\gamma$. Using a similar estimating approach as in the Appendix A.1, we obtained $\hat{\gamma} = 0.740$ with an estimated standard deviation of 0.301. This yields a $p$-value of 0.007, suggesting a lack of fit if we only consider a time-independent treatment effect and agreeing with our KS- and CvM-type test results mentioned above.

### 4.2 Colorectal adenocarcinoma trial

Recently a clinical trial was conducted at the Memorial Sloan-Kettering Cancer Center to evaluate outcomes of intraoperative high-dose radiotherapy in the management of locally recurrent or primary unresectable colorectal adenocarcinoma (both are considered as clinically advanced disease stages) between November 1992 and December 2007. In this study, the patient's survival time is of interest and is measured from the date of surgery to the date of death or last follow-up. Totally 300 patients are legible for analysis, where 88 are primary unresectable and 212 are locally recurrent patients. The censoring percentage is 48%.

Among all clinical variables collected, the one of most interest to the physicians is the presentation of the disease, i.e., whether the patient is locally recurrent or primary unresectable. Though not conclusive, some clinical trials have shown that primary unresectable patients have better survival outcomes than locally recurrent ones. For our cohort of 300 patients, a routine Cox regression analysis and a log-rank test both confirmed that, without adjusting for other clinical factors, primary unresectable patients tend to have longer survival times than locally recurrent patients. In our analysis, we wish to apply model (3) to draw conclusions on the difference of the MRLFs of the two groups of patients. To see if it is reasonable to use the test statistics developed in Sect. 2 for model checking, we first examine the independent censoring assumption. Both the Cox regression model ($p$-value = 0.32) and the log-rank test ($p$-value = 0.31), switching the role of the censoring and the survival time and using the presentation of the disease as the covariate, suggest that such an assumption is acceptable. Thus we proceed with the computation of the test statistics with the $p$-values evaluated by resampling 50000 times as in Sect. 4.1. It turns out that the KS-type test statistic $\mathcal{F}_n^{(1)}$ is 169.52, which yields a $p$-value of 0.55, and the CvM-type test statistic $\mathcal{F}_n^{(2)}$ is 902.06 which yields a $p$-value of 0.72. Hence we conclude that the additive MRL model (3) can be reasonably fit to this data set by letting $T$ be the survival time and $Z$ be the dichotomous variable indicating whether or not the patient is locally recurrent or primary resectable. Figure 2 depicts the two nonparametric

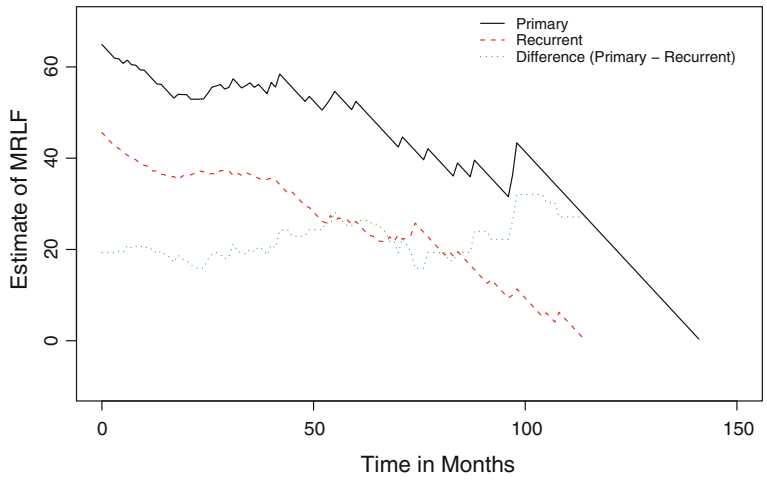


**Fig. 2** Graphical model checking for the colorectal adenocarcinoma trial: Estimates of the two MRLFs and their difference

MRLF estimates. It can be seen that they are approximately parallel (or equivalently, the difference is roughly constant), confirming the additive effect of $Z$ postulated by model (3). Finally, applying model (3) (let $Z = 1$ denote the locally recurrent group and 0 otherwise) and using the estimating Eq. A3 in the Appendix A.1, we obtain $\hat{\beta} = -19.57$ with a standard error as 15.04. This yields a $p$-value of 0.097 for testing $H_0 : \beta_0 = 0$ versus the one-sided alternative $H_1 : \beta_0 < 0$ and suggests for a marginally significant effect. The interpretation of the coefficient estimate $\hat{\beta}$ is that the MRLF of the locally recurrent patients is 19.57 months shorter than that of the primary unresectable patients. Note that this number is quite close to the difference graphically shown in Fig. 2, which further indicates the validity of the parameter estimate.

Because of the high censoring percentage in this data set, one may wonder whether our tests have enough power to reject the null hypothesis even if it is not true. We therefore ran a simulation with sample size 300 and censoring percentage 50% while data are generated from (2), a situation similar to those described in Sect. 3. To mimic the colorectal adenocarcinoma trial, we let the single covariate $Z \sim$ Bernoulli$\{0, 1\}$ with probability 0.3. We considered various situations by setting the type I error rate as 0.05. For instance, we obtained the power as 0.767 for the KS-type test and 0.366 for the CvM-type test when $\beta = 0.4$ (which corresponds to a ratio of $e^{0.4} \simeq 1.5$ for the MRLF of the two arms) and $m_0(t) = (-t/2 + 1/2)^+$. When $\beta = 0.7$ (corresponding to a ratio of $e^{0.7} \simeq 2$)and $m_0(t) = (-t/3 + 1/3)^+$, the power is 0.959 for the KS-type test and 0.565 for the CvM-type test. This suggests that the KS-type test has reasonable powers for rejecting the null at least for proportional MRL alternatives, while the CvM-type test may not be sensitive to such alternatives.

## 5 Power study for local alternatives

We now study the power of our tests for local alternatives converging to the null. Because both the KS-type and the CvM-type test statistics are based on the process $\theta_n(t, z)$, it suffices to derive the asymptotic distribution of $\theta_n(t, z)$ under local alternatives.

Consider a sequence (indexed by $n$) of local alternatives defined as

$$m_n(t|Z) = m_0(t) + \beta' Z + n^{-1/2} R(Z), \tag{10}$$

where $R(Z)$ is an unknown function and does not depend on $\beta$. Under the alternative models (10), (4) becomes

$$m_0(t) = \frac{1}{H(t,z)} \left\{ \int_t^\tau \frac{(s-t)G(t)}{G(s)} H_1(ds, z) - \int_0^z \beta_0' w H(t, dw) - n^{-1/2} \int_0^z R(w) H(t, dw) \right\}. \tag{11}$$

This implies that the empirical process $V_n(t, z)$ defined in (6) may have a similar shift under the local alternatives. To rigorously justify this we will need to investigate the estimator $\hat{\beta}$ under the alternative hypothesis.

Under mild regularity conditions (see Banerjee 2005, Theorem 2.2), it can be easily shown that the sequence of probability measures according to $m_n(t|Z)$ in model (10) is contiguous to the probability measure according to the $m(t|Z)$ in model (3). This means that $\hat{\beta}^{(n)}$, the estimator of $\beta_0$ (see the Appendix A.1 for details) obtained under the alternative hypothesis, is still a consistent estimator for $\beta_0$. However, below we will show that, the asymptotic distribution of $n^{1/2}(\hat{\beta}^{(n)} - \beta_0)$ is no longer centered at zero. Instead, it has a non-random shift.

We first examine $n^{-1/2} U^{(n)}(\beta_0)$ defined in (A2) with the superscript stressing that it is now evaluated under the alternative hypothesis. By the same arguments as those in the appendix and technical report of Sun and Zhang (2009), It can be easily shown that

$$n^{-1/2} U^{(n)}(\beta_0) = n^{-1/2} \sum_{i=1}^n \psi_i^{(n)} + \tilde{R} + o_p(1),$$

where $\tilde{R} = \int_0^\tau E\left[I(T > t) R(Z)(Z - \bar{z}(t))\right] d\tilde{Q}(t)$, $\tilde{Q}(t)$ is the limit of $Q(t)$ and $\psi_i^{(n)}$ is defined similarly as $\psi_i$ in the Appendix A.1. but with

$$M_i^{(n)}(t) = \frac{\Delta_i I(X_i > t)}{G(X_i)} \left[ (X_i - t) - \{m_0(t) + \beta_0' Z_i + n^{-1/2} R(Z_i)\} \right]$$

under the local alternatives. Thus

$$n^{1/2}(\hat{\beta}^{(n)} - \beta_0) = A^{-1}n^{-1/2}\sum_{i=1}^{n}\psi_i^{(n)} + A^{-1}\tilde{R} + o_p(1).$$

Because the difference between $\psi_i^{(n)}$ and $\psi_i$ is at the order of $n^{-1/2}$ and both $\psi_i^{(n)}$ and $\psi_i$ have mean zero under the alternative and the null, respectively, the former has the same limiting distribution as the latter by contiguity of the local alternatives. That is, $n^{1/2}(\hat{\beta}^{(n)} - \beta_0)$ asymptotically follows a normal distribution with mean $A^{-1}\tilde{R}$ and the null variance.

Therefore, the empirical estimator (6), now denoted by $V_n^{(n)}(t, z)$, is actually

$$\begin{aligned} V_n^{(n)}(t, z) &= V_n(t, z) + n^{-1/2}\frac{\int_0^z \tilde{R}'A^{-1}wH_n(t, dw)}{H_n(t, z)} + o_p(n^{-1/2}) \\ &= V_n(t, z) + n^{-1/2}\frac{\int_0^z \tilde{R}'A^{-1}wH(t, dw)}{H(t, z)} + o_p(n^{-1/2}). \end{aligned}$$

Denote $\frac{\int_0^z \tilde{R}'A^{-1}wH(t,dw)}{H(t,z)}$ by $\Gamma(t, z)$, we have

$$\theta_n^{(n)}(t, z) = \theta_n(t, z) + [\Gamma(t, z) - \Gamma(t, z_u)] + o_p(1).$$

In other words, $\theta_n^{(n)}(t, z)$ has a non-random shift of $[\Gamma(t, z) - \Gamma(t, z_u)]$ from the null hypothesis under (10). Thus the proposed KS- and CvM-type test statistics can detect the alternatives converging to the null at the rate of $n^{-1/2}$.

Finally we revisit the re-sampling approach used in Sect. 2 under the local alternatives. It is obvious that

$$\hat{\eta}_i^{(n)}(t, z) = \hat{\eta}_i(t, z) + O_p(n^{-1/2}).$$

Observe that $E(\Omega_i) = 0$ and the difference between $\hat{\eta}_i^{(n)}(t, z)$ and $\hat{\eta}_i(t, z)$ is at the order of $n^{-1/2}$. Therefore, the re-sampled process $\tilde{W}^{(n)}(t, z)$ has the same limiting distribution as $\tilde{W}(t, z)$ under the null hypothesis, meaning that the critical values of the tests remain the same asymptotically.

Note that the above results are essentially the conclusions of Le Cam's third lemma (Bickel et al. 1993, p. 500). However it is quite tedious to verify the condition of joint normality as it would involve the evaluation of the log-likelihood ratio. We thus chose the above ad-hoc approach due to its simplicity and transparency.

## 6 Concluding remarks

In this paper we discuss model checking tools for the additive MRL model, a topic that seems not to have been explored in the literature. Two goodness-of-fit tests are proposed based on an appropriately constructed stochastic process. We establish relevant asymptotic properties and conduct numerical studies. Both simulation and real data analysis show that the proposed tests work reasonably well.

We briefly discuss some differences between the tests proposed here and existing procedures for survival models in the literature. Some goodness-of-fit tests, e.g., the ones developed by Schoenfeld (1980), Andersen (1982), McKeague and Utikal (1991), may require arbitrary discretization of the time axis and covariates. The involved limiting distributions are usually chi-square. Several test statistics based on martingale or score residuals have also been proposed. See Fleming and Harrington (1991, Sect. 4.5) for an excellent summary and illustration. In addition, there exist many graphical model checking tools such as the Cox-Snell residual plot (see, for example, Kalbfleisch and Prentice 2002, Sect. 4.5). Our tests are similar to the ones proposed by Yuen and Burke (1997) and Yuen et al. (2003). The former was for Aalen's additive hazards model and the latter was developed for the proportional MRL model when no censoring is present. However, our tests for the additive MRL model take into account censoring and do not require arbitrary partitioning of the time axis or covariate space. Of course, one limitation of our approach is that it requires the independent censoring assumption. Although it is often true in randomized trials or well-administered studies such as the two examples we see in Sect. 4, it would be attractive to relax this assumption to accommodate more complicated censoring mechanisms.

Here we briefly discuss one possible such extension. Instead of assuming that the censoring time $C$ is independent of $Z$, we shall only assume that $T$ and $C$ are conditionally independent given $Z$. As a matter of fact, this situation is equivalent to the *missing at random* scenario in the literature of missing data, while the independent censoring assumption used before is equivalent to *missing completely at random*. Define $G(t|z) = P(C > t|Z = z)$, that is, the survival function of $C$ given $Z = z$. Then, under the additive MRL model, similar to (4), we have

$$m_0(t) = \frac{1}{H(t,z)} \left\{ \int_0^z \int_t^\tau \frac{(s-t)G(t|w)}{G(s|w)} H_1(ds, dw) - \int_0^z \beta_0' w H(t, dw) \right\}. \quad (12)$$

The derivation of (12) is similar to that of (4) and is thus omitted. Therefore, similar test statistics can be constructed if we are willing to assume a model for estimating $G(t|z)$ when $Z$ is continuous (as suggested by Robins and Ritov (1997), it is wise to postulate such a model due to the so-called "curse of dimensionality"). For instance, we can specify a Cox model to detail the dependence structure between $C$ and $Z$ and replace the homogeneous survival estimator $\hat{G}$ by the one resulting from the Breslow estimator. Due to the martingale structure of the Breslow estimator, similar arguments can be applied to show the limiting Gaussian process. On the other hand, if the covariate $Z$ is categorical, one might just estimate $G(t|z)$ by the Kaplan–Meier method for each value of $z$, and apply our tests with some computational modifications.

Another possible extension is to use different weight functions for the proposed test statistics. For instance, when computing the KS- and CvM-type statistics, instead of adopting equal weights for all $t$, $z$ or all observed sample points, we may assign more weights to early events or the values of covariates that appear to have high density. Such techniques are commonly used in nonparametric tests and are useful in improving powers or detecting different types of alternatives. See, for example, Fleming and Harrington (1991, Chapter 7) and Gill and Schumacher (1987).

It is often desirable to identify the departure from the model assumption when lack-of-fit is detected. For the two tests proposed here, this task seems challenging analytically. The foundation of the two test statistics, the stochastic process $\theta_n(t, z)$, essentially represents the discrepancy of estimates of $m_0(t)$ for different values of the covariate $Z$. However, after taking the supremum (KS-type test) or the integral (CvM-type test), it is no longer possible to tell exactly where the discrepancies are. Therefore, one way to detect such departure when the null hypothesis is rejected is to examine the value of $\hat{\theta}_n(t, z)$ for all $t$ and $z$ in the hope of identifying the source that results in large values of the two test statistics. Another way is to plot the non-parametric estimates of the MRLF for various values of the covariate (for continuous covariates one may need to partition the space of $Z$). As we see in the example, this graphical tool can successfully reveal the departure of model assumption and may even suggest a remedy.

One should be careful about the difference between the MRL models and the hazard models. Models utilizing hazard functions focus on the instantaneous (conditional) probability that a subject will "fail" in the next moment, given that the subject has "survived" up until now, while the MRL models concentrate on the average remaining survival time. Therefore, MRL models are not rank invariant. In other words, the magnitude of the survival observation is important. Consequently, analysis results from these two types of models may not always be the same if influential outliers exist. Moreover, it has been shown by Shaked and Shantikhumar (1993) that the ordering of the MRLF is weaker than the ordering of the hazard function. Therefore, modelling the MRLF can be an attractive alternative when the ordering assumption (e.g., the proportional hazards assumption) on the hazard function is violated.

**Acknowledgements** The first author's research was partly supported by the National Cancer Institute Core Grant (CA008748). The second author's research was partly supported by a grant of the Research Grants Council of Hong Kong. The third author's research was partly supported by the National Natural Science Foundation of China Grants (No. 10731010, 10971015 and 10721101) and the National Basic Research Program of China (973 Program) (No. 2007CB814902). We would like to thank the Editor, Professor Mei-Ling Ting Lee, and two anonymous reviewers for their constructive suggestions that greatly improved the paper.

## Appendix

### A.1 Derivation of $\hat{\beta}$

Since the additive MRL model is a special case of the class of transformed MRL models, the inverse probability censoring weighting (IPCW) estimation procedure proposed in Sun and Zhang (2009) can be applied. Here we present details for calculating $\hat{\beta}$ and provide some remarks for this special case.

Define

$$M_i(t) = \frac{\Delta_i I(X_i > t)}{G(X_i)} \left[ (X_i - t) - \{m_0(t) + \beta_0' Z_i\} \right], \quad i = 1, \ldots, n.$$

Note that under model (3), $M_i(t)$ are zero-mean stochastic processes. Thus, for a given $\beta$, we can estimate $m_0(t)$ by solving the IPCW (Robins and Rotnitzky 1992) estimating equation

$$\sum_{i=1}^{n} \frac{\Delta_i I(X_i > t)}{\hat{G}(X_i)} \left[ (X_i - t) - \{m_0(t) + \beta' Z_i\} \right] = 0, \quad 0 \le t \le \tau.$$

Denote this estimator by $\hat{m}_0(t; \beta)$, and note that we can explicitly express it by

$$\hat{m}_0(t; \beta) = \frac{\sum_{i=1}^{n} \Delta_i \hat{G}(X_i)^{-1} I(X_i > t)\{X_i - t - \beta' Z_i\}}{\sum_{i=1}^{n} \Delta_i \hat{G}(X_i)^{-1} I(X_i > t)}. \tag{A1}$$

To estimate the regression parameter $\beta_0$, the following IPCW generalized estimating equation (Liang and Zeger 1986) can be used

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \frac{\Delta_i I(X_i > t) Z_i}{\hat{G}(X_i)} \left[ (X_i - t) - \{\hat{m}_0(t; \beta) + \beta' Z_i\} \right] dQ(t) = 0, \tag{A2}$$

where $Q(t)$ is an increasing weight function on $[0, \tau]$. Note that the role of the integrator $Q(t)$ here is to take into account contributions of the observations at different time points. In this study we take $Q(t) = n^{-1} \sum_{i=1}^{n} \Delta_i I(X_i \le t)$, that is, $Q(t)$ jumps at every uncensored observations. In view of (A1), the above estimating equation is equivalent to

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \frac{\Delta_i I(X_i > t)}{\hat{G}(X_i)} \{Z_i - \bar{Z}(t; \hat{G})\}\{X_i - t - \beta' Z_i\} dQ(t) = 0, \tag{A3}$$

where $S^{(0)}(t; G) = n^{-1} \sum_{i=1}^{n} \Delta_i G(X_i)^{-1} I(X_i > t)$, $S^{(1)}(t; G) = n^{-1} \sum_{i=1}^{n} \Delta_i G(X_i)^{-1} I(X_i > t) Z_i$, and $\bar{Z}(t; G) = S^{(1)}(t; G)/S^{(0)}(t; G)$.

Let $\hat{\beta}$ denote the solution to $U(\beta) = 0$, we see that it has a closed form

$$\hat{\beta} = \frac{\sum_{i=1}^{n} \int_0^{\tau} \Delta_i I(X_i > t) \hat{G}^{-1}(X_i)\{Z_i - \bar{Z}(t; \hat{G})\}\{X_i - t\} dQ(t)}{\sum_{i=1}^{n} \int_0^{\tau} \Delta_i I(X_i > t) \hat{G}^{-1}(X_i)\{Z_i - \bar{Z}(t; \hat{G})\}^{\otimes 2} dQ(t)},$$

where $v^{\otimes 2} = vv'$ for a vector $v$. Also define $\hat{m}_0(t) := \hat{m}_0(t; \hat{\beta})$, that is, replace $\beta$ by $\hat{\beta}$ in (A1), the corresponding estimator of the unknown baseline MRLF $m_0(t)$. It is easy to see that this estimator has a least square type of interpretation. In particular, $\hat{\beta}$ mimics the slope estimator and $\hat{m}_0(t)$ (which is not of our primary interest) the intercept estimator in normal equations.

Define $\hat{\Sigma} = n^{-1}\sum_{i=1}^{n}\hat{\psi}_i^{\otimes 2}$,

$$\hat{\psi}_i = \int_0^\tau \hat{M}_i(t)\left\{Z_i - \bar{Z}(t;\hat{G})\right\}dQ(t) + \int_0^\tau \frac{\hat{q}(t)}{\hat{\pi}(t)}d\hat{M}_i^c(t),$$

$$\hat{q}(t) = n^{-1}\sum_{i=1}^{n} I(X_i \geq t)\int_0^\tau \hat{M}_i(u)\left\{Z_i - \bar{Z}(u;\hat{G})\right\}dQ(u),$$

$$\hat{M}_i(t) = \frac{\Delta_i I(X_i > t)}{\hat{G}(X_i)}\Big[(X_i - t) - \{\hat{m}_0(t) + \hat{\beta}' Z_i\}\Big],$$

$$\hat{M}_i^c(t) = N_i^c(t) - \int_0^t I(X_i \geq u)d\hat{\Lambda}^c(u),$$

$$\hat{\Lambda}^c(t) = \int_0^t \frac{\sum_{i=1}^n dN_i^c(u)}{n\hat{\pi}(u)}, \quad \hat{\pi}(t) = n^{-1}\sum_{i=1}^{n} I(X_i \geq t),$$

$$N_i^c(t) = I(X_i \leq t, \Delta_i = 0)$$

and $\hat{A} = n^{-1}\sum_{i=1}^{n}\int_0^\tau \frac{\Delta_i I(X_i>t)}{\hat{G}(X_i)}\{Z_i - \bar{Z}(t;\hat{G})\}^{\otimes 2}dQ(t)$. It has been shown in Sun and Zhang (2009) that

$$n^{1/2}(\hat{\beta} - \beta_0) = A^{-1}n^{-1/2}\sum_{i=1}^{n}\psi_i + o_p(1) \tag{A4}$$

($A$ is defined in (C4) below and $\psi_i$ is similarly defined as $\hat{\psi}_i$ with the sample estimates being replaced by their expectations or limits. See Sun and Zhang (2009) for details.) and thus has an asymptotically normal distribution with mean zero and variance that can be consistently estimated by the sandwich estimator $\hat{A}^{-1}\hat{\Sigma}\hat{A}^{-1}$.

In order to derive (A4), Sun and Zhang (2009) assumed certain regularity conditions, which are also adopted here.

(C1) $Q(t)$ converges almost surely to a nonrandom and bounded function $\widetilde{Q}(t)$ uniformly in $t \in [0, \tau]$.
(C2) $Z$ is bounded and $G$ is continuous.
(C3) $m_0(t)$ is continuously differentiable on $[0, \tau]$.
(C4) $A := E\left[\int_0^\tau I(T_i > t)\{Z_i - \bar{z}(t)\}^{\otimes 2}d\widetilde{Q}(t)\right]$ is nonsingular where $\bar{z}(t)$ is the limit of $\bar{Z}(t;G)$.

Interestingly, the estimators derived above can be regarded as an extension of the efficient estimator for one-sample cases without censoring. To see this, suppose we observe an i.i.d. sample $\{T_i; i = 1, \ldots, n\}$. Note that now there is no covariate or censoring involved. In this simple case, Yang (1978) and Bicket et al. (1993, p. 197) have shown that the empirical MRLF estimator based on the relationship between the survival function and the MRLF explored in Sect. 1,

$$\tilde{m}(t) := \frac{\int_t^\infty \tilde{S}(u)du}{\tilde{S}(t)}, \quad t < \max\{T_1, \ldots, T_n\}$$

$\tilde{S}(t)$ being the empirical survival function based on the i.i.d. sample $\{T_i\}_{i=1}^n$, is asymptotically unbiased and Gaussian, uniformly strong consistent and efficient. Note that $\tilde{m}(t)$ can be rewritten as

$$\tilde{m}(t) = \frac{\sum_{i=1}^n I(T_i > t)(T_i - t)}{\sum_{i=1}^n I(T_i > t)} \tag{A5}$$

and is thus essentially the solution of the following estimating equation

$$\sum_{i=1}^n I(T_i > t)(T_i - t - m(t)) = 0.$$

Apparently, the estimator derived in (A1) is an extension of (A5) here.

If right censoring is present in this homogeneous sample, that is, we observe an i.i.d. sample $\{X_i, \Delta_i; i = 1, \ldots, n\}$, then it is not difficult to show that we can use the IPCW estimating equation

$$\sum_{i=1}^n \frac{\Delta_i I(X_i > t)}{\hat{G}(X_i)}(X_i - t - m(t)) = 0 \tag{A6}$$

to derive an asymptotically unbiased estimate for $m(t)$. Compared with (A1), we see that (A6) may also be regarded as a special version of (A1) when $\beta$ is known or set to be zero (because the sample is homogeneous). Such an estimator is used in Sect. 4 for graphical checking of the model's goodness-of-fit.

## A.2 Proof of weak convergence of $\theta_n(t, z)$

We first show that

$$\begin{aligned}\phi_n(t, z) = &\frac{n^{1/2}\{B_n(t, z) - B(t, z)\}}{H(t, z)} - \frac{1}{H(t, z)} \int_0^z w' H(t, dw) n^{1/2}\{\hat{\beta} - \beta_0\} \\ &- \frac{n^{1/2}\{H_n(t, z) - H(t, z)\}}{H(t, z)^2} \left[ B(t, z) - \int_0^z \beta_0' w H(t, dw) \right] \\ &- \frac{n^{1/2}}{H(t, z)} \int_0^z \beta_0' w \left[ H_n(t, dw) - H(t, dw) \right] + o_p(1),\end{aligned} \tag{A7}$$

where

$$B_n(t,z)=\int_t^\tau \frac{(s-t)\hat{G}(t)}{\hat{G}(s)}H_{1n}(ds,z) \quad \text{and} \quad B(t,z)=\int_t^\tau \frac{(s-t)G(t)}{G(s)}H_1(ds,z).$$

To see (A7), we start from the left-hand-side

$$\begin{aligned}
\phi_n(t,z) &= n^{1/2}\{V_n(t,z)-V(t,z)\}\\
&= n^{1/2}\left[\frac{B_n(t,z)}{H_n(t,z)}-\frac{\int_0^z \hat{\beta}' w H_n(t,dw)}{H_n(t,z)}\right]\\
&\quad - n^{1/2}\left[\frac{B(t,z)}{H(t,z)}-\frac{\int_0^z \beta_0' w H(t,dw)}{H(t,z)}\right]\\
&= n^{1/2}\frac{B_n(t,z)-B(t,z)}{H(t,z)} - n^{1/2}B_n(t,z)\left[\frac{1}{H(t,z)}-\frac{1}{H_n(t,z)}\right]\\
&\quad - n^{1/2}\left[\frac{\int_0^z \hat{\beta}' w H_n(t,dw)}{H_n(t,z)}-\frac{\int_0^z \beta_0' w H_n(t,dw)}{H_n(t,z)}\right]\\
&\quad - n^{1/2}\left[\frac{\int_0^z \beta_0' w H_n(t,dw)}{H_n(t,z)}-\frac{\int_0^z \beta_0' w H_n(t,dw)}{H(t,z)}\right]\\
&\quad - n^{1/2}\left[\frac{\int_0^z \beta_0' w H_n(t,dw)}{H(t,z)}-\frac{\int_0^z \beta_0' w H(t,dw)}{H(t,z)}\right].
\end{aligned}$$

Using the strong law of large numbers (Pollard 1990, p. 39), it can be shown that

$$B_n(t,z)\left[\frac{H_n(t,z)-H(t,z)}{H(t,z)H_n(t,z)}\right] = B(t,z)\left[\frac{H_n(t,z)-H(t,z)}{H(t,z)^2}\right]+o_p(n^{-1/2}), \quad \text{(A8)}$$

$$\frac{\int_0^z w H_n(t,dw)}{H_n(t,z)} = \frac{\int_0^z w H(t,dw)}{H(t,z)}+o_p(1) \quad \text{(A9)}$$

$$\text{and} \quad \int_0^z w H_n(t,dw)\left[\frac{1}{H_n(t,z)}-\frac{1}{H(t,z)}\right] = \int_0^z w H(t,dw)\left[\frac{H(t,z)-H_n(t,z)}{H(t,z)^2}\right] + o_p(n^{-1/2}). \quad \text{(A10)}$$

Therefore (A7) holds.

Secondly, using the well known martingale expression of the Kaplan–Meier estimator (Fleming and Harrington 1991, p. 97)

$$\frac{\hat{G}(t) - G(t)}{G(t)} = -\int_0^t \frac{\hat{G}(u-)}{G(u)} \frac{\sum_{i=1}^n dM_i^c(u)}{n\hat{\pi}(u)} \tag{A11}$$

where $M_i^c(t) = N_i^c(t) - \int_0^t I(X_i \geq u) d\Lambda^c(u)$ and $\Lambda^c(t)$ is the cumulative hazard function of $C$, we have

$$\begin{aligned}
\frac{\hat{G}(t)}{\hat{G}(s)} - \frac{G(t)}{G(s)} &= \frac{G(t)\left\{1 - \int_0^t \frac{\hat{G}(u-)}{G(u)} \frac{\sum_{i=1}^n dM_i^c(u)}{n\hat{\pi}(u)}\right\}}{G(s)\left\{1 - \int_0^s \frac{\hat{G}(u-)}{G(u)} \frac{\sum_{i=1}^n dM_i^c(u)}{n\hat{\pi}(u)}\right\}} - \frac{G(t)}{G(s)} \\
&= \frac{G(t)}{G(s)} \left\{ \frac{\int_t^s \frac{\hat{G}(u-)}{G(u)} \frac{\sum_{i=1}^n dM_i^c(u)}{n\hat{\pi}(u)}}{1 - \int_0^s \frac{\hat{G}(u-)}{G(u)} \frac{\sum_{i=1}^n dM_i^c(u)}{n\hat{\pi}(u)}} \right\} \\
&= \frac{G(t)}{G(s)} \int_t^s \frac{\hat{G}(u-)}{G(u)} \frac{\sum_{i=1}^n dM_i^c(u)}{n\hat{\pi}(u)} + o_p(n^{-1/2}) \\
&= \frac{G(t)}{G(s)} \int_t^s \frac{\sum_{i=1}^n dM_i^c(u)}{n\hat{\pi}(u)} + o_p(n^{-1/2}) \,.
\end{aligned}$$

Therefore,

$$\begin{aligned}
B_n(t,z) - B(t,z) &= \int_t^\tau \frac{(s-t)\hat{G}(t)}{\hat{G}(s)} H_{1n}(ds,z) - \int_t^\tau \frac{(s-t)G(t)}{G(s)} H_1(ds,z) \\
&= \int_t^\tau \frac{(s-t)\hat{G}(t)}{\hat{G}(s)} H_{1n}(ds,z) - \int_t^\tau \frac{(s-t)G(t)}{G(s)} H_{1n}(ds,z) \\
&\quad + \int_t^\tau \frac{(s-t)G(t)}{G(s)} H_{1n}(ds,z) - \int_t^\tau \frac{(s-t)G(t)}{G(s)} H_1(ds,z) \\
&= \int_t^\tau \frac{(s-t)G(t)}{G(s)} \{H_{1n}(ds,z) - H_1(ds,z)\} \\
&\quad + \int_t^\tau \frac{(s-t)G(t)}{G(s)} \int_t^s \frac{\sum_{i=1}^n dM_i^c(u)}{n\hat{\pi}(u)} H_{1n}(ds,z) + o_p(n^{-1/2}) \\
&= \int_t^\tau \frac{(s-t)G(t)}{G(s)} \{H_{1n}(ds,z) - H_1(ds,z)\}
\end{aligned}$$

$$+n^{-1}\sum_{i=1}^{n}\int_{t}^{\tau}\int_{u}^{\tau}\frac{(s-t)G(t)}{G(s)}H_1(ds,z)\frac{dM_i^c(u)}{\pi(u)}+o_p(n^{-1/2}). \tag{A12}$$

Thus, combining (A4), (A7) and (A12) and after some algebraic manipulation, we have

$$\theta_n(t,z)=\phi_n(t,z)-\phi_n(t,z_u)=n^{-1/2}\sum_{i=1}^{n}\eta_i(t,z)+o_p(1),$$

where $\eta_i(t,z)=\xi_i(t,z)-\xi_i(t,z_u)$, and

$$\begin{aligned}\xi_i(t,z)=&\frac{\Delta_i(X_i-t)G(t)}{H(t,z)G(X_i)}I(X_i>t,Z_i\le z)\\&+\frac{G(t)}{H(t,z)}\int_{t}^{\tau}\int_{u}^{\tau}\frac{s-t}{G(s)\pi(u)}H_1(ds,z)dM_i^c(u)\\&-\frac{V(t,z)}{H(t,z)}I(X_i>t,Z_i\le z)-\frac{1}{H(t,z)}\int_{0}^{z}w'H(t,dw)A^{-1}\psi_i\\&-\frac{\beta_0'Z_i}{H(t,z)}I(X_i>t,Z_i\le z).\end{aligned}$$

Therefore, by the same arguments as those of Appendix A.5 in Lin et al. (2000), $\theta_n(t,z)$ converges weakly to a zero-mean Gaussian process with covariance function $\sigma(t,z;t^*,z^*)=E\{\eta_i(t,z)\eta_i(t^*,z^*)\}$ at $(t,z)$ and $(t^*,z^*)$, which can be consistently estimated by $\hat{\sigma}(t,z;t^*,z^*)$ given in Sect. 2.

## References


Andersen PK (1982) Testing goodness of fit of Cox's regression and life model. Biometrics 38:67–77

Arnold BC, Zahedi H (1988) On multivariate mean remaining life functions. J Multivar Anal 25:1–9

Balkema AA, de Haan L (1974) Residual life time at great age. Ann Probab 2:792–804

Banerjee M (2005) Likelihood ratio tests under local alternatives in regular semiparametric problems. Stat Sin 15:635–644

Bickel PJ, Klaassen CA, Ritov Y, Wellner JA (1993) Efficient and adaptive inference in semiparametric models. Johns Hopkins University Press, Baltimore, MD

Chen YQ (2007) Additive regression of expectancy. J Am Stat Assoc 102:153–166

Chen YQ, Cheng SC (2005) Semiparametric regression analysis of mean residual life with censored survival data. Biometrika 92:19–29

Chen YQ, Cheng SC (2006) Linear life expectancy regression with censored data. Biometrika 93:303–313

Chen YQ, Jewell NP, Lei X, Cheng SC (2005) Semiparametric estimation of proportional mean residual life model in presence of censoring. Biometrics 61:170–178

Cox DR (1961) Renewal theory. Methuen, London

Fleming TR, Harrington DP (1991) Counting processes and survival analysis. Wiley, New York

Gill RD, Schumacher M (1987) A simple test of the proportional hazards assumption. Biometrika 74: 289–300

Hall WJ, Wellner JA (1984) Mean residual life. In: Csörgő M, Dawson DA, Rao JNK, Md. E. Saleh AK (eds) In: Proceedings of the international symposium on statistics and related topics. North Holland, Amsterdam, pp 169–184

Hollander M, Proschan F (1975) Tests for the mean residual life. Biometrika 62:585–593

Kalbfleisch JD, Prentice RL (2002) The statistical analysis of failure time data. 2. Wiley, New York

Kotz S, Shanbhag DN (1980) Some new approaches to probability distribution. Adv Appl Probab 12:903–921

Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73: 13–22

Lin DY, Wei LJ, Yang I, Ying Z (2000) Semiparametric regression for the mean and rate functions of recurrent events. J R Stat Soc Ser B 62:711–730

Maguluri G, Zhang C-H (1994) Estimation in the mean residual life regression model. J R Stat Soc Ser B 56:477–489

McKeague IW, Utikal KJ (1991) Goodness-of-fit tests for additive hazards and proportional hazards models. Scand J Stat 18:177–195

Oakes D, Dasu T (1990) A note on residual life. Biometrika 77:409–410

Pollard D (1990) Empirical processes: theory and applications. Institute of Mathematical Statistics, Hayward, CA

Prentice RL (1973) Exponential survivals with censoring and explanatory variables. Biometrika 60: 279–288

Robins JM, Ritov Y (1997) Towards a curse of dimensionality appropriate (CODA) asymptotic theory for semiparametric models. Stat Med 16:285–319

Robins JM, Rotnitzky A (1992) Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell N, Dietz K, Farewell V (eds) AIDS epidemiology: methodologic issues. Birkhauser, Boston pp 297–331

Schoenfeld D (1980) Chi-squared goodness-of-fit tests for the proportional hazards regression model. Biometrika 67:145–153

Shaked M, Shantikhumar J (1993) Stochastic orders and their applications. Academic Press, Boston

Sun L, Zhang Z (2009) A class of transformed mean residual life models with censored survival data. J Am Stat Assoc 104:803–815

Swartz GB (1973) The mean residual life function. IEEE Trans Reliab R-22:108–109

Yang GL (1978) Estimation of a biometric function. Ann Stat 6:112–116

Yuen KC, Burke MD (1997) A test of fit for a semiparametric additive risk model. Biometrika 84:631–639

Yuen KC, Zhu LX, Tang NY (2003) On the mean residual life regression model. J Stat Plan Infer 113: 685–698