Lifetime Data Anal (2011) 17:514–532
DOI 10.1007/s10985-011-9197-x

Nonparametric estimation of the conditional mean residual life function with censored data

Alexander C. McLain · Sujit K. Ghosh

Received: 18 June 2010 / Accepted: 28 April 2011 / Published online: 21 May 2011

Abstract The conditional mean residual life (MRL) function is the expected remaining lifetime of a system given survival past a particular time point and the values of a set of predictor variables. This function is a valuable tool in reliability and actuarial studies when the right tail of the distribution is of interest, and can be more informative than the survivor function. In this paper, we identify theoretical limitations of some semi-parametric conditional MRL models, and propose two nonparametric methods of estimating the conditional MRL function. Asymptotic properties such as consistency and normality of our proposed estimators are established. We investigate via simulation study the empirical properties of the proposed estimators, including bootstrap pointwise confidence intervals. Using Monte Carlo simulations we compare the proposed nonparametric estimators to two popular semi-parametric methods of analysis, for varying types of data. The proposed estimators are demonstrated on the Veteran's Administration lung cancer trial.

Keywords Covariates · Local averaging · Mean residual life · Right censoring · Smoothing

1 Introduction

In survival and reliability analysis, the hazard function has long been studied and used to make inferences based on time-to-event data. One shortcoming of the hazard

A. C. McLain (✉)
Division of Epidemiology, Statistics and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD 20892, USA
e-mail: mclaina@mail.nih.gov

S. K. Ghosh
Department of Statistics, North Carolina State University, SAS Hall, Raleigh, NC 27695, USA

function is that its interpretation as the "instantaneous rate of failure" is conceptually difficult to understand and may not be a relevant metric to measure the long-term reliability. The mean residual life (MRL) function of a survival time T (such that $T > 0$), denoted by $m(t)$, is the expected remaining lifetime given survival up to time t. That is,

$$m(t) \equiv E[T - t|T > t] = \int_t^{\tau_T} \frac{S(u)}{S(t)} du, \tag{1}$$

where $S(t) = \Pr(T > t)$ is the survival function, $\tau_T = \inf\{t : S(t) = 0\}$ and we use the convention that $0/0 \equiv 0$. The relation of $S(\cdot)$ to $m(\cdot)$ is given by the well-known inversion formula (Hall and Wellner 1984),

$$S(t) = 1 - F(t) = \frac{m(0)}{m(t)} \exp\left\{-\int_0^t m(u)^{-1} du\right\}. \tag{2}$$

From (1) and (2) it follows that $m(\cdot)$ uniquely determines $S(\cdot)$ and vice versa. The hazard function, denoted by $\lambda(\cdot)$, can be written as $\lambda(t) = \{m'(t)+1\}/m(t)$ where hereafter $'$ is used to denote the derivative of a function with respect to t, $m'(t) \equiv dm(t)/dt$, provided that $m(\cdot)$ is a differentiable function.

The MRL function is especially useful when the tail behavior of the distribution is of interest. In the analysis of reliability and actuarial data there are circumstances where it is appropriate to account for the tail of the distribution. For example, a life insurance company may be interested in the life expectancy of a person, or an engineering firm may wish to estimate the expected remaining lifetime of a system, given survival past time t. In this case, a summary of the conditional distribution rather than the current rate of failure is appropriate. The conditionality of the MRL function is also useful. For example, it is commonplace for medical practitioners to give patients the survival probabilities for death (e.g., for one, two and five years) when a patient initially receives a critical diagnosis. While these survival estimates are informative, they are only useful at the time of diagnosis and no additional information can be given at a followup visit when the patient has survived some length of time. Using the MRL function, information can be provided to medical practitioners that they can use to predict the expected remaining life.

In this paper we propose two methods for nonparametrically estimating the conditional MRL function, defined as $m(t|z) = E[T - t|T > t, \mathbf{Z} = z]$, when baseline covariates (z) are available and survival times are subject to censoring. Considerable research has been done on nonparametric estimation of the conditional survivor and cumulative hazard functions (Dabrowska 1987, 1989, 1992; Gonzalez-Manteiga and Cadarso-Suarez 1994), but little has been done for the MRL function especially in the presence of covariates. There has been a recent surge in the development of semi-parametric estimation methods for the conditional MRL function (see Chen and Cheng 2005; Chen et al 2005; Chen 2007; Sun and Zhang 2009) however, as we discuss in Sect. 2, there are drawbacks to this framework. Previous methods of nonparametric estimation of the MRL function in the absence of covariates have been proposed by

Yang (1977), Chaubey and Sen (1999, 2008), and Abdous and Berred (2005), among others. Sen (1999) proposed a non-parametric mean residual life procedure with a concomitant. This work differs from our proposed estimation procedure in that it does not consider censored observations and uses a Poisson instead of a Bernstein smoothing function. Moreover, Sen (1999) does not provide numerical studies to demonstrate the empirical performance of the proposed models.

The Characterization theorem of Hall and Wellner (1984) provides necessary and sufficient conditions, such that $m(\cdot)$ is a proper MRL function. That is, $F(\cdot)$ is a proper continuous distribution function if and only if $m(\cdot)$ satisfies:

(a) $m(t) \geq 0$ for all $t \geq 0$;
(b) $m(t) + t$ is nondecreasing in t;
(c) if there exists a τ such that $m(\tau) = 0$ then $m(t) = 0$ for all $t \geq \tau$, otherwise, $\int_0^\infty m(t)^{-1} dt = \infty$;
(d) $m(\cdot)$ is a cadlag function with positive increments at discontinuities.

Hereafter these will be referred to as conditions (a–d). Nonparametric procedures for the estimation of the MRL function are appealing since resulting estimates will satisfy these requirements. An MRL function not satisfying (a) for some $t = t^*$ will result in a negative survival function at t^*. An MRL function that satisfies (a), but does not satisfy (b) at t^* will have a negative hazard function at t^*. Condition (c) is needed to ensure that $F(\cdot)$ is a proper distribution function. For convenience, in the remainder of the article we will assume that $m(\cdot)$ is a differentiable function (so condition (d) holds), though this property can be relaxed for our estimation method.

The paper proceeds as follows. In Sect. 2, we will discuss some of the current semi-parametric methods of analysis. This section focuses on the limitations of some of the popular semi-parametric models. Section 3 introduces two methods for non-parametrically estimating the conditional MRL function, when there may be random right censoring. This section contains asymptotic properties of the proposed estimators. We investigate the properties of the proposed estimators in practical settings via simulation study in Sect. 4. This section includes an efficiency comparison between estimators obtained using our proposed estimation method and those obtained using semi-parametric estimation techniques. In Sect. 5, we demonstrate our method on the Veteran's Administration lung cancer trial from Kalbfleisch and Prentice (1980). Proofs and regularity conditions are deferred to the appendix.

2 Semi-parametric estimation of the conditional MRL function

Consider the case where it is of interest to estimate $m(\cdot|z)$ where $z = (z_1, \ldots, z_q)^T$ is a q-dimensional vector of explanatory variables and ^{T} is used to denote the transpose of a vector. In this setting the proportional MRL model (Oakes and Dasu 1990) has been considered,

$$m^{pr}(t|z) = m_0^{pr}(t) \exp(\boldsymbol{\beta}^T z), \tag{3}$$

where $m_0^{pr}(\cdot)$ is a baseline MRL function, and $\boldsymbol{\beta}$ is a q-dimensional vector of regression coefficients. Note that by differentiating (3) with respect to t we obtain $m^{pr\prime}(t|z) =$

$m_0^{pr\prime}(t)\exp(\boldsymbol{\beta}^T \boldsymbol{z})$. If $m_0^{pr\prime}(t) < 0$ for some t, and $\boldsymbol{\beta}^T \boldsymbol{z} > 0$, the situation can arise where $m^{pr\prime}(t|\boldsymbol{z}) < -1$ for that t, violating condition (b). A parametric example where this restriction causes problems is when the baseline population follows a standard lognormal distribution, with MRL function denoted by $m^{LN}(\cdot)$. In this situation $m^{LN\prime}(0) = -1$, hence (3) must be fit under the restriction that $\boldsymbol{\beta}^T \boldsymbol{z} < 0$, for all $\boldsymbol{z}$. Intuitively, it appears that $m_0^{pr\prime}(t) < 0$ would be likely since MRL functions usually decrease to zero when $t \to \infty$.

Chen and Cheng (2006), and Chen (2007) frame the estimation of $m(\cdot|\boldsymbol{z})$ as an additive expectancy regression model. That is, they model the conditional MRL function as

$$m^a(t|\boldsymbol{z}) = m_0^a(t) + \boldsymbol{\gamma}^T \boldsymbol{z}. \tag{4}$$

where $m_0^a(t)$ is a baseline MRL function, and $\boldsymbol{\gamma}$ is a q-dimensional vector of regression coefficients. Estimation of this model must be carried out subject to the constraint that $m^a(t|\boldsymbol{z}) \geq 0$ for all $\boldsymbol{z}$ and $t \geq 0$. Imposing this constraint is especially difficult when $m_0^a(t) \to 0$ as $t \to \infty$, which is the case for many parametric families including the Weibull distribution when the shape parameter is greater than 1. The constraint on the positivity of (4) is similar to the constraint on the additive hazard model (Lin and Ying 1994), however the further constraint imposed by condition (b) makes this situation unique. For example, one could extend (4) with a time-dependent parameter $\boldsymbol{\gamma}(t)$, say $\boldsymbol{\gamma}(t)^T \boldsymbol{z} = \boldsymbol{\gamma}^T \boldsymbol{z} + \tilde{\boldsymbol{\gamma}}^T \boldsymbol{z}t$, where the positivity of (4) would be likely. However, we then have $m^{a\prime}(t|\boldsymbol{z}) = m_0^{a\prime}(t) + \tilde{\boldsymbol{\gamma}}^T \boldsymbol{z}$, hence there is still a constraint on $\tilde{\boldsymbol{\gamma}}$, namely $\tilde{\boldsymbol{\gamma}}^T \boldsymbol{z} \geq -1 - \min_{t\in[0,\tau)}\{m_0^{a\prime}(t)\}$.

Chen (2007) argues that condition (a) can be satisfied in (4) by a linear transformation on the covariates as follows. Without loss of generality we suppose that $z_i \geq 0$ for all $i = 1, \ldots, q$ so that forcing $\boldsymbol{\gamma} \geq 0$, with $m_0^a(\cdot) > 0$, will guarantee that condition (a) holds. The author recommends that if $\gamma_i < 0$ for some i, then we define $\tilde{z}_i = M_i - z_i$ where M_i is large. Then (4) is reparameterized by replacing z_i with $\tilde{z}_i$ for all i such that $\gamma_i < 0$, forcing the parameter for $\tilde{z}_i$ to be positive. The problem with this approach is that estimation procedures for (4) do not have the constraint that $m_0^a(t) > 0$. In Sect. 5 we incorporate this method on a negative estimate of the MRL function and find that reparameterization changes the baseline MRL so that it takes on negative values, and hence was not a solution.

The difficulty in fitting the proportional and additive models is that the parameter space of the regression parameter depends on the characteristics of the unknown baseline MRL function. Construction of a semi-parametric estimator that satisfies conditions (a–d) is still an open ended problem. This issue has been noted recently by Sun and Zhang (2009) who propose the general family of semi-parametric transformation models

$$m^g(t|z) = g\{m_0(t) + \beta^T z\}, \tag{5}$$

that includes (3) and (4) as special cases. On page 814, the authors note “For the class of transformed mean residual life models (5)... may not always satisfy (condition (b)) for certain β unless $m_0(t)$ itself is nondecreasing. The necessary condition for this constraint is that $g\{m_0(t)\} + t$ is nondecreasing. Further research is needed to provide a necessary and sufficient condition for this constraint under model (5) with a general transformation function g."

3 Nonparametric estimation of the conditional MRL function

For the remainder, let $S_T(\cdot|z)$ and $m(\cdot|z)$ denote, respectively, the conditional survival and MRL functions of a continuous nonnegative random variable T, given $\boldsymbol{Z} = z$. Let $S_C(\cdot|z)$ denote the conditional survival function of the censoring variable C, given $\boldsymbol{Z} = z$. We assume the censoring to be conditionally independent of the survival time, given $\boldsymbol{Z} = z$, with S_C and S_T containing no common jumps across all values of z. The observed data set consists of n independent and identically distributed replicates of $(X_i, \delta_i, z_i), i = 1, \ldots, n$ where $X_i \equiv \min(T_i, C_i)$, and $\delta_i = I_{(T_i \leq C_i)}$ is the censoring indicator. The conditional survival function of the observed data is denoted by $S_X(\cdot|z) = S_C(\cdot|z)S_T(\cdot|z)$. To avoid a technical discussion of tail probabilities let τ_X be such that $\tau_X = \inf\{t : S_X(t|z) = 0 \text{ for all } z\} < \infty$.

We use a local averaging method to construct a nonparametric regression procedure for estimating $m(\cdot|z)$. In this article we only consider the *Nadaraya–Watson kernel* method, but the main results still hold for *partitioning*, and *k–nearest neighbor* estimates with slightly different regularity conditions (see chapter 26 of for some related results Györfi et al 2002). Let $K : \mathbb{R}^q \to \mathbb{R}$ be a q-dimensional kernel function and

$$W_{ni}(z|h_n) = \frac{K\left(\frac{z-z_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{z-z_j}{h_n}\right)},$$

where h_n denotes the bandwidth. It is widely known that the choice of the kernel function for nonparametric estimates is not crucial to the performance of the nonparametric estimator and thus we use the standard q-dimensional Epanechnikov kernel for all of our numerical illustrations.

To ensure that our nonparametric method produces an estimate $\hat{m}(\cdot|z)$ that satisfies the conditions of the characterization theorem we start with an estimate of the conditional survivor function and then use (1). Gonzalez-Manteiga and Cadarso-Suarez (1994), following the work of Dabrowska (1987), give the following generalized product-limit estimator,

$$S_n^P(t|z, h_n) = I_{(t \leq X_{(n)})} \prod_{\{i : X_{(i)} \leq t\}} \left\{ \frac{\sum_{j=i+1}^n W_{n(j)}(z|h_n)}{\sum_{j=i}^n W_{n(j)}(z|h_n)} \right\}^{\delta_{(i)}} \tag{6}$$

where $X_{(i)}$ denotes the ith order statistic, and $\{\delta_{(i)}, W_{n(i)}(z|h_n)\}$ denote the corresponding censoring indicator and weight of $X_{(i)}$. As done by previous authors, we set $S_n^P(t|z) = 0$ for all $t > X_{(n)}$ to ensure (1) is finite, this constraint can be relaxed with the estimator discussed below.

Due to the highly volatile nature of MRL estimates a smoothed estimate may be preferred. Consider a smoothed estimate of the conditional survivor function based on Bernstein polynomials,

$$S^B_{n,N}(t|z,h_n,\tau) \equiv \sum_{k=0}^{N} S^P_n\left(\frac{\tau k}{N}\bigg| z,h_n\right)\binom{N}{k}\left(\frac{t}{\tau}\right)^k\left(1-\frac{t}{\tau}\right)^{N-k}$$
$$= \sum_{k=0}^{N} S^P_n\left(\frac{\tau k}{N}\bigg| z,h_n\right)\psi_{k,N}\left(\frac{t}{\tau}\right), \tag{7}$$

where $\tau \geq X_{(n)}$ and $\psi_{k,N}(\cdot)$ are the so-called Bernstein basis functions. Note that even though $S^P_n(t|z,h_n) = 0$ for all $t > X_{(n)}$, $S^B(t|z,h_n) > 0$ for all $t \leq \tau$ and nonzero estimates of the survival, and MRL, can be made up to τ.

The Bernstein polynomial is a natural choice in this circumstance because it does not suffer from boundary bias like some kernel methods. Furthermore, the Bernstein basis has optimal shape preserving properties of all polynomials of the same degree (see Carnicer and Peña 1993). In particular, it inherits the properties of its coefficients, making the estimator in (7) a proper survivor function. A similar estimate of the survival function was proposed by Babu et al (2002) for use in the absence of censoring and covariates. Instead of using a bound, Babu et al (2002) transform survival time to $[0, 1)$ with, for example, $T' = T/(T+1)$. Bernstein polynomials are most efficient when T is uniform over the range, since the “knots” are equidistant. While transformations such as T' are theoretically appealing, we found that in practice they usually take the data further from being uniform over $[0, 1)$. As a result, a larger N (which causes higher variance) is required to gain sufficient detail to the portions of t where the majority of data lie.

From (6) and (7) we introduce two corresponding estimates of the MRL function,

$$m^P_n(t|z,h_n) = \int_t^{\tau_X} \frac{S^P_n(u|z,h_n)}{S^P_n(t|z,h_n)}du, \tag{8}$$

and

$$m^B_{n,N}(t|z,h_n,\tau) = \int_t^{\tau_X} \frac{S^B_{n,N}(u|z,h_n)}{S^B_{n,N}(t|z,h_n)}du. \tag{9}$$

These shall be referred to as the ‘GPLE’ and ‘Bernstein’ MRL estimates, respectively. It is clear from the form of (6) and (7) that conditions (a–d) will hold for m^P_n and $m^B_{n,N}$.

An attractive feature of (8) and (9) is that both have convenient closed form analytical expressions that can be computed using standard functions available in most software. To see this let $T_{(1)}, \ldots, T_{(K)}$ denote the K ordered and distinctly observed survival times. The GPLE estimate, evaluated at t when $T_{(k-1)} < t < T_{(k)}$, has the following closed form

$$m_n^P(t|z,h_n) = (T_{(k)} - t) + m_n^P(T_{(k)}|z,h_n)\frac{S_n^P(T_{(k)}|z,h_n)}{S_n^P(T_{(k-1)}|z,h_n)},$$

where

$$m_n^P(T_{(k)}|z,h_n) = S_n^P(T_{(k)}|z,h_n)^{-1}\sum_{l=k+1}^{K}(T_{(l)} - T_{(l-1)})S_n^P(T_{(l-1)}|z,h_n),$$

for $k = 1,\dots,K$. It is also straightforward to show that the Bernstein MRL function can be expressed as

$$m_{n,N}^B(t|z,h_n,\tau) = \frac{\sum_{k=0}^{N} S_n^P\left(\frac{\tau k}{N}|z,h_n\right)\Psi_{k,N}\left(\frac{t}{\tau}\right)}{\sum_{k=0}^{N} S_n^P\left(\frac{\tau k}{N}|z,h_n\right)\psi_{k,N}\left(\frac{t}{\tau}\right)},$$

with $\Psi_{k,N}(p) = \sum_{l=0}^{k}\psi_{l,N}(p) = B_{1-p}(n-k,k+1)/B_1(n-k,k+1)$, where $B_q(a,b) = \int_0^q t^{a-1}(1-t)^{b-1}dt$ is the incomplete beta function. We now state our theoretical results of the proposed estimators under some standard regularity conditions.

Lemma 1 *(Consistency) Under the assumptions* $(A0)$, *and* $(A2)ii$ *given in the appendix, if* $h_n \to 0$ *and* $nh_n^q \to \infty$,

$$\begin{aligned} &(I)\quad ||m_n^P - m||_{\tau_X} \to 0 \textit{ as } n \to \infty \\ &(II)\quad ||m_{n,N}^B - m||_{\tau_X} \to 0,\ \textit{as } n, N \to \infty \end{aligned}$$

for all $z \in \mathcal{Z}$, *where* $||f||_{\tau_X} \equiv \sup_{t\in[0,\tau_X]}|f(t)|$ *is the supremum norm and* $\mathcal{Z}$ *is as defined in (A2).*

The proof of Lemma 1 can be found in the appendix. Before we state the asymptotic normality of m_n^P, we refer to a result shown by Gonzalez-Manteiga and Cadarso-Suarez (1994) (or for left truncated data Iglesias-Pérez and González-Manteiga 1999) when the covariate is real valued (i.e., $q = 1$). Under $(A1)$–$(A4)$, given in the appendix, $\sqrt{nh_n}\{\hat{S}_n^P(\cdot|z,h_n) - S_T(\cdot|z)\} \to^d G(\cdot|z)$ as $n \to \infty$ for all $z \in \mathcal{Z}$, where $G(\cdot|z)$ is a mean-zero Gaussian process with covariance function

$$\begin{aligned}\Gamma(y,t|z) &= S_T(y|z)S_T(t|z)\frac{(\int K^2)}{f_Z(z)}\left\{\int_0^{y\wedge t}\frac{dH(u|z)}{S_X^2(u|z)}\right\} \\ &\equiv S_T(y|z)S_T(t|z)\frac{(\int K^2)}{f_Z(z)}D(y\wedge t|z) \qquad (10)\end{aligned}$$

where $f_Z(\cdot)$ denotes the density of Z, $\int K^2 = \int_{-\infty}^{\infty}K^2(u)du$, and $H(t|z) = \Pr(X \le t, \delta = 1|Z = z)$.

Theorem 1 *(Asymptotic Normality) Under regularity conditions* $(A0)--(A4)$ *given in the appendix, for* $\tau = \tau_X$, $z \in \mathbb{R}$ *and* $t \in [0, \tau]$ *if* $nh_n^5 \to 0$ *and* $(\log n)^3/(nh_n) \to 0$ *then*

$$\sqrt{nh_n}\{\hat{m}_n^P(t|z, h_n) - m(t|z)\} \to^d \ N(0, \Phi(t|z))$$

for all $z \in \mathcal{Z}$, *with*

$$\Phi(t|z) = \frac{(\int K^2)}{S_T(t|z)^2 f_Z(z)} \left[\int_t^\infty S_T(u|z)B_t(u|z)du + \int_t^\infty S_T(u|z)^2 m(u|z)D(u|z)du \right.$$
$$\left. -m(t|z)^2 S_T(t|z)^2 D(t|z) \right]. \tag{11}$$

where $B_t(u|z) = \int_t^u S_T(w|z)D(w|z)dw$.

The proof of this theorem is a non-trivial application of the functional delta method from the survival function to the MRL function. The proof of Theorem 1 and regularity conditions $(A0)$–$(A4)$ with discussion on their implications are contained in the appendix.

4 Simulation studies

In this section we present results from extensive simulation studies that were run to evaluate the performance of our method for practical sample sizes. In the first subsection, we compare the performance of Berstein and GPLE MRL models to the semi-parametric MRL models discussed in Sect. 2, with varying data types. In the second subsection, we test the performance of bootstrap standard error estimates and Wald type pointwise confidence intervals.

An initial simulation study was run to evaluate various bandwidth selection methods. To quantify the performance of a particular bandwidth selection method we used

$$\sum_{j=1}^{M} \int_0^{X_{(n)}^j} [\hat{m}_j\{t|E(Z), h_n\} - m\{t|E(Z)\}]^2 dt \tag{12}$$

where $\hat{m}_j$ and $X_{(n)}^j$ are the estimated MRL function and the maximum observed survival time in the jth iteration, respectively, $j = 1, \ldots, M$ for $M = 1000$. We simulated data according to a setting identical to that in Sect. 4.2. For both of the nonparametric estimation methods the *biased cross-validation* method performed better compared to other off the shelf bandwidth selection methods (e.g., *unbiased cross-validation, "rules-of-thumb,"*etc. Sheather and Jones 1991), and hence it was used in the simulations. For multivariate nonparametric regression, where the curse of dimensionality comes into play, programs such as `npregbw` in the `R` package `np` (Hayfield and Racine 2008) can be used for bandwidth selection or a semi-parametric dimension reduction approach (cf. Hu et al 2010) can be used to revert the problem to a single dimension.

4.1 Comparison of semi-parametric and nonparametric methods

In this section, using the notation of Sects. 2 and 3, we compare $m^a(\cdot)$, $m^B(\cdot)$ with $N = n/4$, $m^{B2}(\cdot)$ with $N = n/2$, $m^P(\cdot)$, and $m^{pr}(\cdot)$. The estimators $m^{pr}(\cdot)$, and $m^a(\cdot)$ were fit using the methods proposed in Chen and Cheng (2005, 2006). These methods also produce an estimate of the regression parameter, which is not reported since $m(\cdot)$ is the parameter of interest.

There were two settings used to generate the data. The first corresponds to an *additive model* with $m(t|z) = t + 1 + \beta z$, which is the MRL function for a Pareto distribution. The second is a *proportional model* with a nonlinearly decreasing MRL function $m(t|z) = \exp\{-t + \beta z\}$. We generated the censoring variable C from an exponential distribution with rate parameter that resulted in 30% of the observations being censored. Two simulations were run for each setting, one with a Bernoulli$(1/2)$ covariate, the other with a Uniform$(0, 2)$ covariate. The β parameter was set to1 and -1 in the additive and proportional settings, respectively. Let Q_q denote the qth quantile of the distribution of X. The MRL function was estimated for $Q_{0.1}$, $Q_{0.25}$, $Q_{0.5}$, $Q_{0.75}$ and $Q_{0.9}$. The bias of the estimate (BIAS), Monte Carlo standard error (σ_{MC}) and mean squared error (MSE) were calculated from 1000 estimates from a sample size of 200. Two Bernstein methods were included to explore the effect that N has on the results. The τ parameter was set to the maximum observation $X_{(n)}$ in each iteration for the Bernstein methods.

The results for the additive simulation are summarized in Table 1. Overall, the semi-parametric additive MRL model has the lowest MSE for both covariate types. For the categorical covariate the GPLE model, for $Q_{0.1}$ and $Q_{0.25}$, and the Bernstein $(n/4)$ model, for $Q_{0.5}$, $Q_{0.75}$ and $Q_{0.9}$, are competitive in MSE with the semi-parametric additive MRL model. For the continuous covariate the additive model has the lowest MSE for all points. The proportional MRL model has the highest variance and MSE for the continuous covariate. It stays competitive in terms of MSE with the categorical covariate due to the low variance of the estimates, but has bias larger than twice the Monte Carlo standard error for $Q_{0.1}$, $Q_{0.5}$, $Q_{0.75}$ and $Q_{0.9}$. For the continuous covariate the GPLE model performs the best among the nonparametric estimators for $Q_{0.1}$ and $Q_{0.25}$, all three perform similarly for $Q_{0.5}$, and for the latter two points the Bernstein $(n/4)$ has the lowest variance and MSE. The Bernstein $(n/4)$ model appears to perform the best overall, and is robust to the type of covariate.

Table 2 contains results for data simulated from the proportional model. The semi-parametric proportional MRL model performs the best overall for both covariate types. This is the case much more so than in the additive simulation (however, since the simulations were run with such different data types this should not be taken as a comparison of semi-parametric methods). The additive model is greatly outperformed by the non-parametric methods for the categorical covariate. The average estimate of m^a for $Q_{0.75}$

Table 1 Summary of results for the additive simulation where $m(t|z) = t + 1 + \beta z$, with $\beta = 1$

	m^a	m^B	m^{B2}	m^P	m^{pr}
	Categorical covariate				
$Q_{0.1}$					
BIAS	0.26	0.29	0.24	**0.20**	0.22
σ_{MC}	**0.39**	0.41	0.4	0.39	0.51
MSE	0.22	0.25	0.21	**0.19**	0.31
$Q_{0.25}$					
BIAS	0.24	0.2	0.19	**0.15**	0.21
σ_{MC}	**0.41**	0.44	0.47	0.43	0.56
MSE	0.23	0.25	0.22	**0.20**	0.35
$Q_{0.5}$					
BIAS	0.15	0.1	0.08	**0.05**	0.14
σ_{MC}	**0.50**	0.52	0.52	0.53	0.68
MSE	**0.27**	0.28	0.27	0.29	0.48
$Q_{0.75}$					
BIAS	**−0.06**	−0.21	−0.22	−0.22	−0.15
σ_{MC}	0.73	**0.72**	0.75	0.8	0.94
MSE	**0.54**	0.57	0.61	0.68	0.91
$Q_{0.9}$					
BIAS	**−0.58**	−0.82	−0.82	−0.81	−0.95
σ_{MC}	1.19	**1.14**	1.2	1.28	1.4
MSE	**1.76**	1.98	2.11	2.29	2.85
	Continuous covariate				
$Q_{0.1}$					
BIAS	−0.25	**−0.15**	−0.21	−0.26	−0.49
σ_{MC}	0.34	0.52	0.51	0.51	**0.24**
MSE	**0.18**	0.29	0.31	0.33	0.29
$Q_{0.25}$					
BIAS	−0.29	**−0.22**	−0.26	−0.30	−0.55
σ_{MC}	0.39	0.56	0.57	0.58	**0.28**
MSE	**0.24**	0.37	0.39	0.43	0.38
$Q_{0.5}$					
BIAS	−0.41	**−0.38**	−0.40	−0.42	−0.75
σ_{MC}	0.51	0.70	0.71	0.74	**0.36**
MSE	**0.43**	0.63	0.67	0.72	0.69
$Q_{0.75}$					
BIAS	−0.71	−0.73	−0.73	**−0.71**	−1.27
σ_{MC}	0.83	1.04	1.10	1.18	**0.55**
MSE	**1.19**	1.62	1.73	1.89	1.90

Table 1 continued

	m^a	m^B	m^{B2}	m^P	m^{pr}
$Q_{0.9}$					
BIAS	−1.41	−1.47	−1.45	**−1.38**	−2.42
σ_{MC}	1.52	1.73	1.86	2.15	**0.30**
MSE	**4.29**	5.16	5.57	6.52	6.68

The MRL function was estimated at $t = Q_{0.1}, Q_{0.25}, Q_{0.5}, Q_{0.75}$ and $Q_{0.9}$ of the distribution of X and were predicted for $z = 1$ for both the Bernoulli$(1/2)$ and Uniform$(0, 2)$ covariate distributions
The bold values represent the smallest value in a given row indicating the best performing model in terms of bias, s.e., or MSE

and $Q_{0.9}$ are less than zero. In fact, of the 1000 estimates 41.1%, 99.9%, and 100% of the estimates of $m^a(t|1)$ for $t = Q_{0.5}, Q_{0.75}$ and $Q_{0.9}$ were less than zero, respectively. As a result the simulation was altered to use $\tilde{Z} = 1 - Z$, and hence $\tilde{\beta} = -\beta = 1$, for the additive model as recommended in Chen (2007), and ran with the same seed. The MSE*'s for the altered simulation were (2.63, 10.09, 46.77, 139.13, 250.43), respectively, and are very similar to those from the initial run. Furthermore, of the 1000 iterations 39.3%, 99.9%, and 100% of the estimates of $m^a(t|1)$ for $t = Q_{0.5}, Q_{0.75}$ and $Q_{0.9}$ were less than zero, respectively. Clearly, the method for ensuring the positivity of m^a was not successful.

For the categorical covariate, the GPLE model has a lower bias and MSE than both of the Bernstein models for $Q_{0.1}, Q_{0.25}$ and $Q_{0.5}$, but for the latter time points the Bernstein models perform noticeably better than the GPLE method. With the continuous covariate the results are similar. Overall, the Bernstein $(n/2)$ method appears to perform the best for the proportional setting. A reason for the good performance of the GPLE model in the proportional setting is the low bias of its estimates. The data in the proportional setting have less variability than the additive setting, hence the mean squared error reflects more of the estimators bias in the proportional setting versus the additive setting. In summary, we find that estimates from the GPLE model have lower bias and higher variance, the Bernstein $(n/4)$ has higher bias and lower variance, while the Bernstein $(n/2)$ is in the middle of the nonparametric estimators. The numerical results suggest that the Bernstein method is robust to the selection of N and rudimentary strategies for selecting it suffice. The Bernstein method with $N = n/2$ appears to be a good choice for data with not much skewness while $n/4$ (and possibly $n/6$) are appropriate when there is more skewness.

4.2 Bootstrap standard errors and confidence intervals

Due to the complexity of (11) we tested pointwise bootstrap variance estimates of GPLE and Bernstein $(n/2)$ MRL estimators. This simulation used a constant MRL function $m(t|z) = \exp(\beta Z)$ (i.e. from an exponential distribution), with $\beta = 2$ and $Z \sim N(0, 1)$. The censoring variable was independently generated from an exponential distribution, where the rate parameter resulted in approximately 20% of the observations being censored. Using a sample size of $n = 500$, with $B = 500$ bootstrap samples to calculate the standard error for the case $Z = 0$, $m(t|0)$ was estimated for

Table 2 Summary of results for the proportional simulation where $m(t|z) = \exp\{-t + \beta z\}$ with $\beta = -1$

	m^a	m^B	m^{B2}	m^P	m^{pr}
	Categorical covariate				
$Q_{0.1}$					
BIAS	−0.042	0.024	0.013	**0.001**	−0.005
σ_{MC}	0.028	0.027	0.028	0.029	**0.026**
MSE *	2.5	1.3	0.9	0.8	**0.7**
$Q_{0.25}$					
BIAS	−0.094	0.028	0.014	**0.001**	−0.005
σ_{MC}	0.03	0.028	0.029	0.031	**0.023**
MSE*	9.7	1.5	1.0	1.0	**0.6**
$Q_{0.5}$					
BIAS	−0.21	0.033	0.016	**0.0008**	−0.007
σ_{MC}	0.042	0.032	0.034	0.039	**0.020**
MSE *	45.8	2.0	1.4	1.5	**0.4**
$Q_{0.75}$					
BIAS	−0.36	0.035	0.016	**0.007**	−0.008
σ_{MC}	0.057	0.047	0.055	0.083	**0.018**
MSE *	133.0	3.4	3.3	6.9	**0.4**
$Q_{0.9}$					
BIAS	−0.47	**0.010**	−0.022	0.136	−0.014
σ_{MC}	0.071	0.06	0.067	0.098	**0.019**
MSE *	226.2	3.7	4.9	28.0	**0.6**
	Continuous covariate				
$Q_{0.1}$					
BIAS	0.066	0.03	0.02	0.012	**−0.006**
σ_{MC}	0.024	0.033	0.033	0.035	**0.020**
MSE *	4.9	2.0	1.5	1.3	**0.4**
$Q_{0.25}$					
BIAS	0.055	0.034	0.024	0.014	**−0.007**
σ_{MC}	0.024	0.033	0.033	0.035	**0.020**
MSE *	3.7	2.2	1.7	1.4	**0.5**
$Q_{0.5}$					
BIAS	0.025	0.042	0.03	0.019	**−0.009**
σ_{MC}	0.026	0.035	0.036	0.039	**0.020**
MSE *	1.3	3.0	2.2	1.9	**0.5**
$Q_{0.75}$					
BIAS	−0.036	0.051	0.038	0.027	**−0.012**
σ_{MC}	0.036	0.04	0.042	0.05	**0.022**
MSE *	2.5	4.2	3.3	3.2	**0.6**
$Q_{0.9}$					
BIAS	−0.105	0.055	0.042	0.038	**−0.017**

Table 2 continued

σ_{MC}	0.057	0.052	0.059	0.081	**0.026**
MSE *	14.2	5.8	5.2	7.9	**1.0**

The MRL function was estimated at $t = Q_{0.1}, Q_{0.25}, Q_{0.5}, Q_{0.75}$ and $Q_{0.9}$ of the distribution of X and were predicted for $z = 1$ for both the Bernoulli(1/2) and Uniform(0, 2) covariate distributions. MSE* corresponds to the MSE$\times$1000

The bold values represent the smallest value in a given row indicating the best performing model in terms of bias, s.e., or MSE

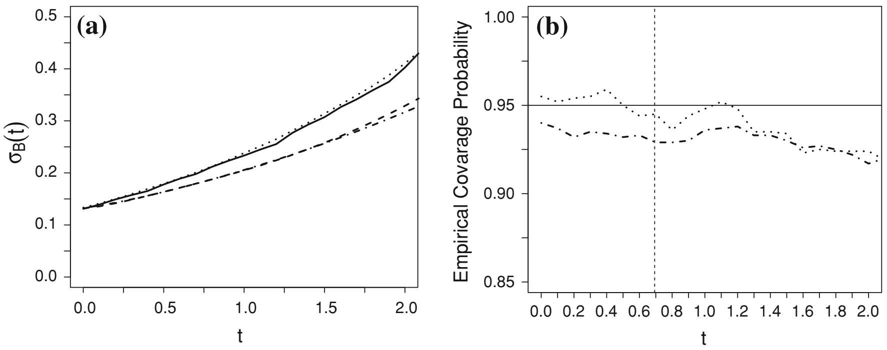

Fig. 1 **a** Comparison of Bootstrap and Monte Carlo standard deviation for the GPLE (*solid line* Monte Carlo, *dotted line* bootstrap) or the Bernstein (*dashed line* Monte Carlo, *dashed dotted line* bootstrap) with $N = n/2$. **b** Pointwise 95% empirical coverage probabilities for the GPLE (*dotted line*) and the Bernstein (*dashed dotted line*) with $N = n/2$, the vertical line represents the median of T given $Z = 0$

values of t up to the 85th percentile of the distribution of T given $Z = 0$. Figure 1a allows for a comparison of the bootstrap estimate of the pointwise standard error and the pointwise Monte Carlo standard error. Figure 1b shows the empirical coverage probabilities of pointwise 95% Wald type confidence intervals. The empirical coverage probabilities for t up to the 1.2 (which is approximately the 70th percentile of T given $Z = 0$) are close to the nominal value for the GPLE model; while the Bernstein ($n/2$) has an average of 0.935 over this range and is slightly permissive. For t greater than the 70th percentile of T given $Z = 0$ the empirical coverage probabilities start to fall off, and confidence intervals should be treated with caution.

5 Data analysis

The Veteran's Administration (VA) data set (Kalbfleisch and Prentice 1980) contains the survival times of 137 (7% censored) advanced lung cancer patients. Information such as age, treatment type, tumor classification, prior therapy, and Karnofsky score were collected on each patient at the time of entry into the study. Karnofsky score is a measure (10–90) of general health, where 10 indicates the patient is completely hospitalized and 90 signifies they are able to care for themselves. The VA data set has been analyzed by numerous authors most of which only consider the data from patients with no prior therapy. The effect of Karnofsky score may differ for the prior therapy

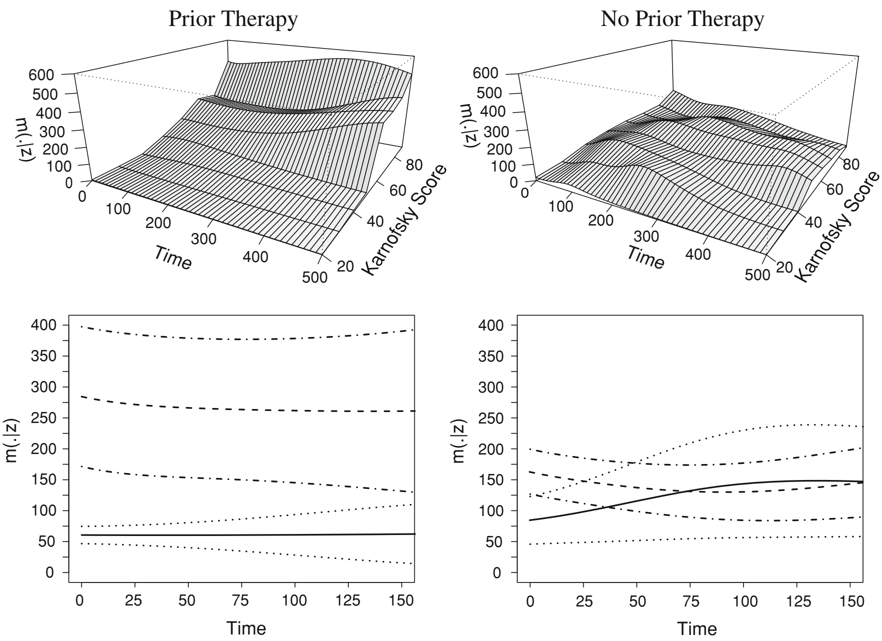

Fig. 2 (*Top*) MRL estimation of survival time of the Veteran's Administration lung cancer trial by Karnofsky Score for the group with and without prior therapy. (*Bottom*) Estimated MRL for the prior therapy and no prior therapy groups with 95% Bootstrap confidence intervals for Karnofsky scores 35 (*solid line*), and 80 (*dashed line*)

and no prior therapy groups. In this data analysis we look at the effect Karnofsky score has on MRL, separately for patients with and without prior therapy.

In Fig. 2, we display the predicted MRL functions (top row) by Karnofsky score for the prior therapy and no prior therapy groups, and 95% pointwise bootstrap confidence intervals (bottom row) for a low Karnofsky score (35) and a high Karnofsky score (80). All figures use the Bernstein method with $N = \lceil n/2 \rceil$, and $\tau = X_{(n)}$, where h_n was obtained via *biased cross-validation*. The figures in the top row show the ability of our method to summarize the MRL function so that an estimate of the remaining life can be given, taking into account the current survival time and covariate information. Furthermore, the figures can be compared to give a visual assessment of how the effect of Karnofsky score differs for patients with prior therapy with those without.

In the bottom row of Fig. 2, we compare the MRL function for patients with high and low Karnofsky scores. From these figures it appears that Karnofsky score has a larger effect for patients with prior therapy versus those with no prior therapy. Among the patients with prior therapy, those with a high Karnofsky score have a significantly longer MRL than those with a low Karnofsky score. Among the patients with no prior therapy, those with a high Karnofsky score have a significantly longer MRL than those with a low Karnofsky score only at baseline. Furthermore, conditional on surviving the first 75 days there is little difference between the low and high Karnofsky score groups. The MRL estimate in the top row of Fig. 2 does not refute the finding that the effect of Karnofsky score differs with time for the no prior therapy group. As a result, we feel that Karnofsky score may only be a significant exposure at the onset for the no prior therapy group.

6 Discussion

In this article we have proposed two nonparametric methods for estimating the conditional MRL function. The GPLE estimator is a natural extension of the conditional survival model from Gonzalez-Manteiga and Cadarso-Suarez (1994) and has been shown to have favorable asymptotic properties. The Bernstein estimators performed slightly better than the GPLE estimators in terms of mean squared error and can be used to create smooth graphs that display the effect covariates have on MRL (see Fig. 2). The benefit of using an MRL type model over a hazard based model is the estimate of $m(t|z)$. This function gives the most commonly used statistic, the average, based on the current cumulative observation length, t, and the values of the individuals explanatory variables, z.

The Bernstein and GPLE methods do not, however, yield an estimate of interpretable regression parameters as do semi-parametric MRL models. This limits the ability of nonparametric approaches to indicate the significance of an exposure's association with MRL. The estimate of $m(\cdot|z)$ presented herein provides a dynamic (time-varying) estimate of the regression coefficient. For instance $m'_j(t|z) \equiv dm(t|z)/dz_j$ can estimate the effect the jth predictor has on the MRL function. If we find that $m'_j(t|z)$ does not vary with t then it suggests that an additive MRL model might be suitable for the given data. We can also estimate $d\log\{m(t|z)\}/dz_j$, if we find that it does not vary with t then a proportional MRL model might be suitable. Still more research is needed in realistic and theoretically valid semi-parametric analysis techniques with MRL functions.

Acknowledgements We would like to acknowledge the help of an associate editor and three referees whose insightful comments improved the overall quality of this manuscript. Furthermore, we would like to acknowledge Professor Yogendra Chaubey and Professor Tom Gerig for help on earlier versions of this manuscript. The research of the first author was supported in part by the Intramural Research Program of the National Institute of Health, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development.

Appendix

In order to prove the asymptotic properties given in Sect. 3, the following regularity conditions will be needed.

(A0) $\inf\{t : S_X(t|\mathbf{Z} = z) = 0\} = \inf\{t : S_T(t|\mathbf{Z} = z) = 0\} < \infty$ for all $z \in \mathcal{Z}$ ($\mathcal{Z}$ is defined in $(A2)$).

(A1) T, C, and $\mathbf{Z}$ are absolutely continuous random variables.

(A2) There exists a Z^* such that $\Pr(|Z_j| > Z^*) = 0$ for all $j = 1, 2, \ldots, q$. Let $\mathcal{Z} = [-z_1, z_2]$ be a region contained in the support of $f_Z(\cdot)$ such that $0 < \alpha \leq$

$\inf\{f_Z(z) : z \in \mathcal{Z}_\delta\} < \sup\{f_Z(z) : z \in \mathcal{Z}_\delta\} < \infty$ for some $\mathcal{Z}_\delta = [-z_1 - \delta, z_2 + \delta]$ with $\delta > 0$. Furthermore, for all $z \in \mathcal{Z}$:

(i.) T and C are conditionally independent given $\mathbf{Z} = z$.

(ii.) There exists $a, b \in [0, \infty)$ and $\theta > 0$ such that $S_X(t|z) \geq \theta$ for all $t \in [a, b]$.

(A3) The first and all partial second derivatives with respect to t and z of $f_Z(z)$, $H(t|z)$ and $S_X(t|z)$ exist and are continuous for $(t, z) \in [0, \tau_X) \times \mathcal{Z}_\delta \subseteq \mathbb{R}$.

(A4) The single dimensional kernel function K, is a symmetric density such that $K(x) = 0$ for all $|x| > 1$, and $\int K^2(x)dx < \infty$.

The convergence in Lemma 1 is stated for all $z \in \mathcal{Z}$ with the supremum metric, hence assumption $(A0)$ is needed to ensure that the support of T is observable for all $z \in \mathcal{Z}$. This assumption is, however, restrictive when the censoring is heavy. In such circumstances the convergence for outlying values of Z and t is questionable. The assumption of the continuity of Z, is needed for Theorem 1 where the derivative of $f_Z()$ is used. A discrete Z would be asymptotically equivalent, at $Z = z$, to a un-conditional nonparametric estimator where the sample is restricted to using z only, since $h_n \to 0$ (see for related resultsYang 1977). The assumptions $(A1)$–$(A4)$ are inherited from Iglesias-Pérez and González-Manteiga (1999), whose results are used, and are similar to those given by Dabrowska (1989), and Gonzalez-Manteiga and Cadarso-Suarez (1994).

Proof of Lemma 1 Proof of (I) in lemma1 follows from the work of Stone (1977) and Pintér (2001) on consistency of nonparametric regression estimates. Pintér (2001) showed that under the conditions of Stone (1977), which are satisfied by $(A0)$ and $(A2)ii$, that $||S_n^P - S||_{\tau_X} \to 0$ as $n \to \infty$ almost surely. Since $\tau_X < \infty$ and S_T is continuous, $m(\cdot|z)$ is a bounded continuous function by it's definition. As a result, $m(\cdot|z)$ is a Hadamard differentiable functional of $S(t|z)$ and the continuous mapping theorem can be used to claim the same convergence properties (see van der Vaart and Wellner 1996, Theorem 1.9.5).

The proof of (II) follows a similar argument. That is, it is sufficient to show that $S_{n,N}^B$ is uniformly convergent. Define $S_N^{B_0}$ as

$$S_N^{B_0}(t|z, \tau) \equiv \sum_{k=0}^{N} S_T(\tau k/N|z)\psi_{k,N}(t/\tau),$$

where $||S_{n,N}^B - S_T||_{\tau_X} \leq ||S_N^{B_0} - S_T||_{\tau_X} + ||S_{n,N}^B - S_N^{B0}||_{\tau_X}$. Feller (1965) shows that if S_T is a bounded continuous function (true in this case) $||S_N^{B_0} - S_T||_{\tau_X} \to 0$ as $N \to \infty$. Furthermore,

$$\begin{aligned} ||S_{n,N}^B - S_N^{B_0}||_{\tau_X} &\leq \max_{0\leq k\leq N} |S_n^P(\tau k/N|z, h_n) - S_T(\tau k/N|z, h_n)| \\ &\leq ||S_n^P - S_T||_{\tau_X} \to 0 \end{aligned}$$

as $n \to \infty$ almost surely. Consequently, $||S_{n,N}^B - S_T||_{\tau_X} \to 0$ and $n, N \to \infty$. □

Proof of Theorem 1 The pointwise normality of m_P follows from the functional delta method. We make the same assumptions made by Iglesias-Pérez and González-Manteiga (1999), setting the left truncated variable to 0, which amounts to $(A1)$–$(A4)$. Throughout we use the notation $\int f = \int_{-\infty}^{\infty} f(u)du$ and $\int_b^a f = \int_b^a f(u)du$. Furthermore, let $\mathbb{E} = \{f : \int f^2 < \infty\}$, $\mathbb{E}^* = \{f : \int f < \infty$ and $1 - f$ is a proper distribution function$\}$ and $\mathbb{D} = \{f : f \in \mathbb{R}, f > 0, f' \geq -1, \int_0^{\infty} f^{-1} = \infty\}$, be metric spaces endowed with the norm $||\cdot||_{\tau_X}$ given in lemma 1.

Let $g_n = (nh_n)^{-1/2}$ and $G_n(t|z) = g_n^{-1}\{\hat{S}_n^P(t|z) - S_T(t|z)\}$. Recall that under $(A1)$–$(A4)$ $G_n \to G$ as $n \to \infty$, where G is a Gaussian process with covariance function $\Gamma(y, t|z)$ given in (10) (Iglesias-Pérez and González-Manteiga 1999). First we will show that $\int_t^{\tau_X} G_n \to \int_t^{\tau_X} G \in \mathbb{E}$. By the boundedness of $\hat{S}_n^P(\cdot|z)$ and $S_T(\cdot|z)$, G_n is tight for all $n < \infty$. Furthermore, G is tight since $\Gamma(y, t|z) < \infty$ for all $(y, t) \in [0, \tau_X) \times [0, \tau_X)$ and $z \in \mathcal{Z}$. By assumption $(A0)$,

$$\int_t^{\tau_X} G_n(u|z)du \leq (\tau_X - t) \sup_{u \in [0, \tau_X)} \{G_n(u|z)\} < \infty,$$

for all $n < \infty$. Similarly notice that the variability Gaussian Process $\int_t^{\tau_X} G$ is bounded

$$E\left\{\left(\int_t^{\tau_X} G\right)^2\right\} \leq (\tau_X - t)^2 \sup_{(y,t) \in [t, \tau_X) \times [t, \tau_X)} \{\Gamma(y, t|z)\} < \infty.$$

Consequently, $\int_t^{\tau_X} G_n$ and $\int_t^{\tau_X} G$ are tight and elements of $\mathbb{E}$. Combining the results, the continuous mapping theorem can be use to show that $\int_t^{\tau_X} G_n \to \int_t^{\tau_X} G$.

Consider the process $\phi(t; G) = \lim_{n \to \infty} g_n^{-1}\{m(t|z, \hat{S}_n^P) - m(t|z, S)\}$ where we use the functional notation for the MRL function $m(t|z, S)$. Note that, $\hat{S}_n^P(t|z) = S_T(t|z) + g_n G_n(t|z)$ and

$$\begin{aligned}
\phi(t; G) &= \lim_{n \to \infty} \frac{1}{g_n} \left\{ \frac{\int_t^{\tau_X} (S_T + g_n G_n)}{S_T(t|z) + g_n G_n(t|z)} - \frac{\int_t^{\tau_X} S_T}{S_T(t|z)} \right\} \\
&= \lim_{n \to \infty} \frac{1}{g_n} \left\{ \frac{g_n S_T(t|z) \int_t^{\tau_X} G_n - g_n G_n(t|z) \int_t^{\tau_X} S_T}{S_T(t|z)\{S_T(t|z) + g_n G_n(t|z)\}} \right\} \\
&= \frac{1}{S_T(t|z)} \left\{ \int_t^{\tau_X} G - G(t|z) m(t|z) \right\}.
\end{aligned}$$

Using the earlier results on the tightness of G and $\int_t^{\tau_X} G$, we have $\phi(t; G) \in \mathbb{E}^\dagger$ for all $t \in [0, \tau_X)$ and $z \in \mathcal{Z}$ wp1. Therefore, $m : \mathbb{E}^* \subset \mathbb{E} \to \mathbb{D}$ is Hadamard-differentiable at S_T tangentially to $\mathbb{E}^*$, with derivative $\phi(\cdot; G)$. By the functional delta method $\sqrt{nh_n}\{m_n^P(t|z, h_n) - m(t|z)\} \to^d \phi(t; G)$, where $\phi(t; G) \sim N(0, E_G\{\phi(t; G)^2\})$ with

$$
\begin{aligned}
E_G\{\phi(t;G)^2\} &= E_G\left[\frac{1}{S_T(t|z)^2}\left\{\int_t^{\tau_X} G - G(t|z)m(t|z)\right\}^2\right] \\
&= \frac{1}{S_T(t|z)^2} E_G\left\{\left(\int_t^{\tau_X} G\right)^2 - 2G(t|z)m(t|z)\left(\int_t^{\tau_X} G\right)\right. \\
&\quad \left. + G(t|z)^2 m(t|z)^2\right\} \\
&= \frac{1}{S_T(t|z)^2}\left[\int_t^{\tau_X}\int_t^{\tau_X} \Gamma(u,w|z)\,du\,dw\right. \\
&\quad \left. -2m(t|z)\int_t^{\tau_X} \Gamma(t,u|z)\,du + m(t|z)^2\Gamma(t,t|z)\right].
\end{aligned}
$$

The form of (11) can be found by the imputing the definition of $\Gamma(y,t|z)$, and some simplification. □

References

Abdous B, Berred A (2005) Mean residual life estimation. J Stat Plann Inference 132:3–19

Babu GJ, Canty AJ, Chaubey YP (2002) Application of Bernstein polynomials for smooth estimation of a distribution and density function. J Stat Plann Inference 105:377–392

Carnicer JM, Peña JM (1993) Shape preserving representations and optimality of the Bernstein basis. Adv Comput Math 1:173–196

Chaubey YP, Sen A (2008) Smooth estimation of mean residual life under random censoring. In: Balakrishnan N, Peña EA, Silvapulle MJ (eds) Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen. Inst Math Stat, Beachwood pp 35–49

Chaubey YP, Sen PK (1999) On smooth estimation of mean residual life. J Stat Plann Inference 75:223–236

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

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

Chen YQ, Cheng S (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

Dabrowska DM (1987) Nonparametric regression with censored survival time data. Scand J Stat 14(3):181–197

Dabrowska DM (1989) Uniform consistency of the kernel conditional Kaplan–Meier estimate. Ann Stat 17:1157–1167

Dabrowska DM (1992) Variable bandwidth conditional Kaplan–Meier estimate. Scand J Stat 19:351–361

Feller W (1965) An introduction to probability theory and its applications, vol. II. Wiley, New York

Gonzalez-Manteiga W, Cadarso-Suarez C (1994) Asymptotic properties of a generalized Kaplan–Meier estimator with some applications. J Nonparametr Stat 4:65–78

Györfi L, Kohler M, Krzyżak A, Walk H (2002) A distribution-free theory of nonparametric regression. Springer-Verlag, New York

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

Hayfield T, Racine JS (2008) Nonparametric econometrics: the np package. J Stat Softw 27:1–32

Hu Z, Follmann DA, Qin J (2010) Semiparametric dimension reduction estimation for mean response with missing data. Biometrika 97:305–319

Iglesias-Pérez C, González-Manteiga W (1999) Strong representation of a generalized product-limit estimator for truncated and censored data with some applications. J Nonparametr Stat 10:213–244

Kalbfleisch JD, Prentice RL (1980) The statistical analysis of failure time data. Wiley, New York-Chichester-Brisbane

Lin DY, Ying Z (1994) Semiparametric analysis of the additive risk model. Biometrika 81:61–71

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

Pintér M (2001) Consistency results in nonparametric regression estimation and classification. PhD thesis, Technical University of Budapest

Sen PK (1999) The mean residual life with a concomitant. In: Dixit UJ, Satam MR (eds) Statistical inference and design of experiments. Narosa Publishing House, New Delhi pp 71–91

Sheather SJ, Jones MC (1991) A reliable data-based bandwidth selection method for kernel density estimation. J R Stat Soc Ser B 53:683–690

Stone CJ (1977) Consistent nonparametric regression. Ann Stat 5:595–620

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

van der Vaart AW, Wellner JA (1996) Weak convergence and empirical processes. Springer-Verlag, New York

Yang G (1977) Life expectancy under random censorship. Stoch Process Appl 6:33–39