Nonparametric Estimation of the Conditional Mean Residual Life Function Based on Censored Data

Alexander C. McLain and Sujit K. Ghosh

NC State Department of Statistics Tech Report# 2627

Abstract

The conditional mean residual life (MRL) function is the expected remaining life given a set of predictors. In this work we consider two nonparametric estimators of the conditional MRL function when the lifetime variable is subject to right censoring. We discuss some theoretical difficulties with current semiparametric models for MRL function estimation methods. We derive the asymptotic consistency of both estimators proposed. The numerical properties of the proposed estimators are investigated via simulation study, and compared to semiparametric estimators in varying settings. We apply the proposed methods to the World Bank's 1997 cross-section Vietnam Living Standards Survey, to demonstrate the applicability of the MRL function to cost data.

Key Words: Covariates; Local averaging; Mean residual life function; Right censoring; Smoothing.

1. Introduction

In survival and reliability analysis, the hazard function has long been studied and used to make inference based on time to event data. There exist many efficient semiparametric hazard based analysis techniques that determine the size and significance of explanatory variables on survival time (see Therneau and Grambsch, 2000). One shortcoming of the hazard function is that it's interpretation as the "instantaneous rate of failure" is conceptually difficult to understand and may not be a relevant measure in practice. The mean residual life (MRL) function of a survival time T (such that $T > 0$ a.e.), denoted by $m(t)$, is the expected remaining lifetime given survival up to time t. That is,

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

where $S(t) = \Pr(T > t)$ is the survival function and $0/0 \equiv 0$. The relation of $S(\cdot)$ and $m(\cdot)$ is given by the inversion formula as

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

From (1) and (2) it follows that $m(\cdot)$ uniquely determines $S(\cdot)$ and vice versa. Furthermore, it can be shown that the hazard function, denoted by $\lambda(\cdot)$, can be written as

$$\lambda(t) = -\partial \log\{S(t)\}/\partial t = \{m'(t) + 1\}/m(t),$$

where $m'(t) \equiv \partial m(t)/\partial t$.

The MRL function is especially useful when the tail behavior of the distribution is of interest. In medical studies the tail of the distribution (i.e., the people that survive for very long periods) is not of interest as much as the immediate rate of failure, hence the hazard rate is more desirable. In reliability and actuarial studies, however, there are many circumstances where the tail of the distribution must be accounted for. 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 up to t^*. In this case a summary of the conditional distribution, rather than the immediate rate of failure, is of primary interest.

This paper proposes two methods for nonparametrically estimating the conditional MRL function, $m(t|\boldsymbol{z})$, when censoring and covariates ($\boldsymbol{z}$) are present. Considerable research has been done on nonparametric estimation of the conditional survivor and cumulative hazard functions (c.f. Dabrowska, 1987, 1989, 1992; Gonzalez-Manteiga and Cadarso-Suarez, 1994), but little has been done for the MRL function. There has been a recent surge in 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 is discussed in section 2, there are drawbacks to this framework. Previous methods of nonparametric estimation of the MRL function have been proposed by Yang (1977); Chaubey and Sen (1999); Abdous and Berred (2005); Chaubey and Sen (2008), none of which have done so in the presence of covariates. Stute (1996) derives the asymptotic distribution of a nonparametric estimate of the MRL function for units with covariates in a set, A, for example. This model uses the method proposed Yang (1977) by excluding observations with covariates not in A, and is noteworthy as a theoretic example.

The MRL function is unique in the properties it must hold so that a valid probability distribution is generated. Nonparametric estimation procedures of the MRL function are appealing since they can fit under these requirements. The Characterization theorem of Hall and Wellner (1981) provides necessary and sufficient conditions, such that $m(\cdot)$ is a MRL function of a continuous nonnegative variable T. That is, $F(\cdot)$ as defined in (2) is a proper continuous distribution function if and only if $m(\cdot)$ satisfies:

(a) $m(t) \geq 0$ for all $t \geq 0$, and continuous;

(b) $m(t) + t$ is nondecreasing in t;

(c) if there exist a τ such that $m(\tau) = 0$ then $m(t) = 0$ for all $t \geq \tau$, otherwise, $\int_0^\infty m(t)^{-1} dt = \infty$,

(Hall and Wellner, 1981). Hereafter these will be referred to as conditions (a) – (c).

A MRL function not satisfying (a) for some t^* will result in a negative survival function at t^*. Similarly a MRL function that satisfies (a), but does not satisfy (b) at t^* will have a negative hazard and probability density function at t^*. Condition (c) is needed to ensure that $F(\cdot)$ is a proper distribution function (i.e. $F(\infty) = 1$). Clearly, if $F(\cdot)$ is not a proper distribution function $m(t) = \infty$ for all t. For convenience for the remainder of the article we will assume that $m(\cdot)$ is a differentiable function, though this condition can be relaxed for our estimation method.

The paper proceeds as follows. In section 2 we will discuss some of the current semiparametric methods of analysis used to estimate a conditional MRL function. This section focusses on the theoretical limitations of some of the popular semiparametric models. Section 3 introduces two methods for nonparametrically estimating the conditional MRL function under random right censoring. This section contains asymptotic properties of the proposed estimators. The properties of the proposed estimators are investigated in practical settings via simulation study in section 4. This section includes an efficiency comparison between our proposed estimates versus two of the semiparametric methods from section 2. The MRL function has a nice application to cost data as noted in Chen (2007). In section 5 this application is highlighted by analyzing the World Bank's 1997 cross-section Vietnam Living Standards Survey (VLSS). Proofs and regularity conditions are deferred to the appendix.

2. Semiparametric Estimation of the Conditional MRL Function

Consider the case where it is of interest to estimate $m(\cdot|\boldsymbol{z})$ in light of some explanatory variables ($\boldsymbol{z}$) and a possible censoring mechanism. Let, $\boldsymbol{z} = (z_1, \ldots, z_q)$ denote a q-dimensional vector of covariates, and C denote the censoring variable with distribution function $F_C(\cdot)$.

In the presence of covariates and censoring the proportional MRL model (Oakes and Dasu, 1990) has been considered, where

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

with $m_0^{pr}(\cdot)$ being a baseline MRL function, and $\boldsymbol{\beta}$ is a q-dimensional vector of regression coefficients. Note that $m^{pr\prime}(t|\boldsymbol{z}) = m_0^{pr\prime}(t)\exp(\boldsymbol{\beta}^T\boldsymbol{z})$, 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 cause problems is if the baseline population follows a standard Lognormal distribution, with MRL function $m^{LN}(\cdot)$. It can be shown that $m^{LN\prime}(0) = -1$, hence (3) must be fit under the restriction that $\boldsymbol{\beta}^T\boldsymbol{z} < 0$ for all $\boldsymbol{z}$. Methods for analyzing (3) have been proposed in Maguluri and Zhang (1994), Chen and Cheng (2005), and Chen et al. (2005). These methods focus more so, or exclusively, on estimation of $\boldsymbol{\beta}$ and do not remedy noncompliance of condition (b) of the characterization theorem.

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

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

with $m_0^a(t)$ being a baseline MRL function, and $\boldsymbol{\gamma}$ is a q-dimensional vector of regression coefficients. Estimation of this model must be done subject to the restraint that $m^a(t|\boldsymbol{z}) \geq 0$ for all $\boldsymbol{z}$ and $t \geq 0$. This restraint 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. Intuitively it appears that $m_0^a(t) \to 0$ as $t \to \infty$ would be the most likely scenario in medical and reliability studies.

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 $\boldsymbol{\gamma} \geq 0$ will guarantee that condition (a) holds. The author recommends that if $\gamma_i < 0$ for some i, then we define $z_i' = M_i - z_i$ where M_i is large. Then (4) is reparameterized replacing $\gamma_i z_i$ with $\gamma_i' z_i'$ for all i such that $\gamma_i < 0$. The problem with this method is that estimation procedures for (4) do not have the constraint that $m_0^a(t) > 0$. In section 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 $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ depend on the unknown baseline MRL function.

A semiparametric model that satisfies conditions (a) and (b) of the characterization theorem is presented in Liu and Ghosh (2008). They propose a proportional scaled MRL model given as

$$m^s(t|\boldsymbol{z}) = m_0^s\{t \exp(-\boldsymbol{\alpha}^T \boldsymbol{z})\} \exp(\boldsymbol{\alpha}^T \boldsymbol{z}). \tag{5}$$

This model will result in a proper MRL function for any $\boldsymbol{\alpha}$ and $\boldsymbol{z}$. It can be shown that (5) will hold if and only if $S(t|\boldsymbol{z}) = S_0\{t \exp(\boldsymbol{\alpha}^T \boldsymbol{z})\}$ for a baseline survival function S_0 that corresponds to m_0^s, hence the specification of (5) is equivalent to an accelerated failure time model. Such a model could be fit using accelerated failure models methods (cf. Robins and Tsiatis, 1992) then use the transformation in (1) to obtain $m^s(t|\boldsymbol{z})$. A difficulty of (5) is the interpretation of $\boldsymbol{\alpha}$ in terms of m^s.

Construction of a semiparametric estimators that satisfy conditions (a)–(c) 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

$$m^t(t|\boldsymbol{z}) = g\{m_0(t) + \beta^T \boldsymbol{z}\}, \tag{6}$$

that includes (3) and (4) as special cases. In page 814, the authors note "For the class of transformed mean residual life models (6)... 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 (6) with a general transformation function g."

3. Nonparametric Estimation of the Conditional MRL Function

For the remainder let $F_T(\cdot|\boldsymbol{z})$, $S_T(\cdot|\boldsymbol{z})$, and $m(\cdot|\boldsymbol{z})$ denote the conditional cumulative distribution, survival, and MRL function of a nonnegative random variable T, given $\boldsymbol{Z} = \boldsymbol{z}$. Let $S_C(\cdot|\boldsymbol{z})$ denote the conditional survival function of the censoring variable, given $\boldsymbol{Z} = \boldsymbol{z}$. For the remainder we will assume the censoring to be conditionally independent of the survival time given $\boldsymbol{z}$, with S_C and S_T containing no common jumps. In the case of censoring $X = T \wedge C$ will be observed, with the indicator variable $\delta = I_{(T \leq C)}$. The observed data set consists of n independent and identically distributed replicates of $(X_i, \delta_i, \boldsymbol{z}_i)$, for $i = 1, \ldots, n$. The survival time of the observed data is denoted by $S_X(\cdot|\boldsymbol{z}) = S_C(\cdot|\boldsymbol{z})S_T(\cdot|\boldsymbol{z})$. Let, τ_X, be such that, $\tau_X = \inf\{x > 0 : S_X(x|\boldsymbol{z}) = 0 \text{ for all } \boldsymbol{z}\}$, and and similarly for τ_T and τ_C.

We use a local averaging method to construct a nonparametric regression procedure for estimating $m(\cdot|\boldsymbol{z})$. That is, $E(T|\boldsymbol{z})$ will be estimated by the average of the T_i where $\boldsymbol{z}_i$ is "close" to $\boldsymbol{z}$. There are numerous estimates available to measure how "close" $\boldsymbol{z}_i$ is to $\boldsymbol{z}$ such as the *partitioning*, *k–nearest neighbor*, and *Nadaraya–Watson kernel* estimates. For the remainder we 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 Györfi et al., 2002, for some related results).

Let $K : \mathbb{R}^q \to \mathbb{R}$ be a q-dimensional kernel function and

$$W_{ni}(\boldsymbol{z}|h_n) \quad = \quad \frac{K\left(\frac{\boldsymbol{z}-\boldsymbol{z}_i}{h_n}\right)}{\sum_{j=1}^{n} K\left(\frac{\boldsymbol{z}-\boldsymbol{z}_j}{h_n}\right)}$$

where h_n is the bandwidth. It is widely known that the choice of 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.

To ensure that our nonparametric method produces an estimate $\hat{m}(\cdot|\boldsymbol{z})$ that satisfies the conditions of the characterization theorem we start with an estimate of the survivor function then use (1). Gonzalez-Manteiga and Cadarso-Suarez (1994), following the work of Beran (1981) and Dabrowska (1987), give the following generalized product-limit estimator,

$$S_n^P(t|\boldsymbol{z}, h_n) = I_{(t \le X_{(n)})} \prod_{\{i: X_{(i)} \le t\}} \left\{ \frac{\sum_{j=i+1}^{n} W_{n(j)}(\boldsymbol{z}|h_n)}{\sum_{j=i}^{n} W_{n(j)}(\boldsymbol{z}|h_n)} \right\}^{\delta_{(i)}} \tag{7}$$

where $X_{(i)}$ denotes the ith order statistic, $\{\delta_{(i)}, W_{n(i)}(\boldsymbol{z}|h_n)\}$ denote the corresponding censoring indicator and weight of $X_{(i)}$. This estimator reduces to the Kaplan–Meier estimator when $W_{n(i)}(z|h_n) = 1/n$ for all i. Note that, like previous authors, we set $S_n(t|z) = 0$ for all

$t > X_{(n)}$ to ensure (1) is finite, and that condition (c) holds.

Due to the volatility of MRL estimates, a smoothed estimate may be preferred. For this we propose a smoothed survivor function by using Bernstein polynomials,

$$\begin{aligned} S^B_{n,N}(t|\boldsymbol{z}, h_n, \tau) &\equiv \sum_{k=0}^{N} S^P_n \left(\frac{\tau k}{N} \middle| \boldsymbol{z}, h_n \right) \binom{N}{k} \left(\frac{t}{\tau} \right)^k \left(1 - \frac{t}{\tau} \right)^{N-k} \\ &= \sum_{k=0}^{N} S^P_n \left(\frac{\tau k}{N} \middle| \boldsymbol{z}, h_n \right) \psi_{k,N} \left(\frac{t}{\tau} \right), \end{aligned} \tag{8}$$

where $\tau > X_{(n)}$ and $\psi_{k,N}(\cdot)$ are the so-called Bernstein basis functions. This smooth estimate of the survival function was used by Babu et al. (2002) in the absence of censoring and covariates. 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 (c.f. Carnicer and Peña, 1993).

From these two estimates of the survivor function we introduce two corresponding estimates of the MRL function,

$$m^P_n(t|\boldsymbol{z}, h_n) = \int_t^\infty \frac{S^P_n(u|\boldsymbol{z}, h_n)}{S^P_n(t|\boldsymbol{z}, h_n)} du, \tag{9}$$

and

$$m^B_{n,N}(t|\boldsymbol{z}, h_n, \tau) = \int_t^\infty \frac{S^B_{n,N}(u|\boldsymbol{z}, h_n)}{S^B_{n,N}(t|\boldsymbol{z}, h_n)} du, \tag{10}$$

which shall be referred to as the 'GPLE' and 'Bernstein' MRL estimates, respectively.

Lemma 1. *(Consistency) Under the assumptions* $(A0)$*, and* $(A2)ii$ *given in the appendix, if*

$h_n \to 0$ and $nh_n^q \to \infty$,

$$(I) \quad ||m_n^P - m||_{\tau_T} \to 0 \quad as \quad n \to \infty$$
$$(II) \quad ||m_{n,N}^B - m||_{\tau_T} \to 0, \quad as \quad n, N \to \infty$$

for all $\boldsymbol{z} \in I$, where $||f||_{\tau_T} \equiv \sup_{t<\tau_T} |f(t)|$ is the supremum norm and I is as defined in the appendix (see (A2))..

The proof of Lemma 1 can be found in the appendix.

Let $T_{(1)}, \ldots, T_{(M)}$ denote the ordered and distinctly observed survival times from the data. Then the GPLE estimate evaluated at the observed survival times has the following closed form,

$$m_n^P(T_{(m)}|\boldsymbol{z}, h_n) = S_n^P(T_{(m)}|\boldsymbol{z}, h_n)^{-1} \sum_{l=m+1}^{M} (T_{(l)} - T_{(l-1)}) S_n^P(T_{(l-1)}|\boldsymbol{z}, h_n), \; m = 1, \ldots, M,$$

where $0/0 \equiv 0$. Furthermore for $T_{(m-1)} < t < T_{(m)}$, $m_n^P(t|\boldsymbol{z}, h_n)$ can be expressed as,

$$m_n^P(t|\boldsymbol{z}, h_n) = \{T_{(m)} - t\} + m_n^P(T_{(m)}|\boldsymbol{z}, h_n) \frac{S_n^P(T_{(m)}|\boldsymbol{z}, h_n)}{S_n^P(T_{(m-1)}|\boldsymbol{z}, h_n)}, \; m = 1, \ldots, M.$$

It is also straightforward to show that the Bernstein MRL function has the closed form

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

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 $B_q(a,b) = \int_0^q t^{a-1}(1-t)^{b-1}dt$ can be computed easily using the beta distribution function. Hence, both estimators have a convenient closed from expression that can be easily computed.

Before we state the asymptotic normality of m_n^P, we refer to a result shown by Gonzalez-

Manteiga and Cadarso-Suarez (1994) (or Pérez and Manteiga, 1999, for left truncated data). Under $(A1)$–$(A4)$, given in the appendix, $\sqrt{nh_n}\{S_n^P(\cdot|\boldsymbol{z},h_n)-S(\cdot|\boldsymbol{z})\}\rightarrow^d G(\cdot|\boldsymbol{z})$ as $n\rightarrow\infty$ where $G(\cdot|\boldsymbol{z})$ is a mean zero Gaussian process with covariance function

$$\begin{aligned}\Gamma(y,t|\boldsymbol{z}) &= S(y|\boldsymbol{z})S(t|\boldsymbol{z})\frac{(\int K^2)}{q(\boldsymbol{z})}\left\{\int_0^{y\wedge t}\frac{dH(u|\boldsymbol{z})}{C^2(u|\boldsymbol{z})}\right\} \\ &\equiv S(y|\boldsymbol{z})S(t|\boldsymbol{z})\frac{(\int K^2)}{q(\boldsymbol{z})}D(y\wedge t|\boldsymbol{z})\end{aligned} \tag{11}$$

where $q(\cdot)$ denotes the marginal density of $\boldsymbol{Z}$, $H(t|\boldsymbol{z})=\Pr(X\leq t,\delta=1|\boldsymbol{Z}=\boldsymbol{z})$, and $S_X(t|\boldsymbol{z})=\Pr(X\geq t|\boldsymbol{Z}=\boldsymbol{z})$.

Theorem 1. *(Asymptotic Normality) Under regularity conditions* $(A0)$–$(A4)$*, given in the appendix, for* $\boldsymbol{z}\in I$ *and* $t\in[0,\tau_X)$ *if* $nh_n^5\rightarrow 0$ *and* $(\log n)^3/(nh_n)\rightarrow 0$ *then*

$$\sqrt{nh_n}\{m_n^P(t|\boldsymbol{z},h_n)-m(t|\boldsymbol{z})\}\rightarrow^d \quad N(0,\Psi(t|\boldsymbol{z})) \tag{12}$$

with

$$\begin{aligned}\Psi(t|\boldsymbol{z}) &= \frac{(\int K^2)}{S(t|\boldsymbol{z})^2q(\boldsymbol{z})}\left[\int_t^\infty S(u|\boldsymbol{z})B_t(u|\boldsymbol{z})du+\int_t^\infty S(u|\boldsymbol{z})^2m(u|\boldsymbol{z})D(u|\boldsymbol{z})du\right. \\ &\quad \left.-m(t|\boldsymbol{z})^2S(t|\boldsymbol{z})^2D(t|\boldsymbol{z})\right].\end{aligned} \tag{13}$$

where $B_t(u|\boldsymbol{z})=\int_t^u S(w|\boldsymbol{z})D(w|\boldsymbol{z})dw$*.*

The proof of this theorem is contained in the appendix.

4. Simulation Studies

To evaluate the performance of our method in practical sample sizes simulation studies were performed. In the simulation studies $m^a(\cdot)$, $m^B(\cdot)$ with $N=n/4$, $m^{B2}(\cdot)$ with $N=$

$n/2$, $m^P(\cdot)$, and $m^{pr}(\cdot)$ were estimated. The main purpose of the simulations studies was to compare the efficiency of the estimation methods discussed and presented herein. The proportional scaled MRL was not included in the simulations, since it is not as popular as the additive and proportional models are. The estimators $m^{pr}(\cdot)$, and $m^a(\cdot)$ were fit using the methods proposed in Chen and Cheng (2005) and Chen and Cheng (2006), respectively. These methods also produce an estimate of the regression parameter, which is not reported since the estimate of $m(\cdot)$ is the parameter of interest.

There were two settings used to generate the data. The first corresponds to an additive model with $m(t|z) = t+1+\beta z$. The second simulation is a proportional model with $m(t|z) = \exp\{-t+\beta z\}$. The censoring variable C was generated from a exponential distribution with rate parameter that resulted in 30% of the observations being censored. Two simulations were ran for each setting, one with a categorical $\{0,1\}$ covariate, the other with a Uniform$(0,2)$ covariate. The β parameter was set to 1 and -1 in the additive and proportional settings, respectively, so that both satisfy conditions (a) – (c) and have valid MRL functions.

In the additive setting, $m_0^{a*}(t) = t+1$, which corresponds to a Pareto distribution. Recall that condition (c) requires that $\int_0^\infty m(t)^{-1}dt = \infty$, hence $m_0^{a*}(\cdot)$ increases as fast as possible (up to a constant) in the tails. As a result, this distribution can be viewed as having the most possible right skewness. The second setting uses $m_0^{pr*}(t) = e^{-t}$, which does not correspond to any major family of distributions. The reason this baseline MRL function is used here is because it tests the nonparametric estimators' ability to predict a nonlinear MRL function, and unlike the severe right skew of $m_0^{a*}(\cdot)$, $m_0^{pr*}(\cdot)$ is a decreasing MRL function and hence generates data that is much more compact and coincides with a typical 'nice' data set. By using these two situations we can compare the GPLE and Bernstein methods for different types of data. Two Bernstein methods were included to explore the effect that N had on the results. After some initial tests the values of $N = n/4$ and $n/2$ appeared to be good choices for a variety of distributions, but overall the Bernstein method was not sensitive to the value of N.

We refer to $Q_{0.5}$ as the median, $Q_{0.25}$ and $Q_{0.75}$ as the quartiles and $Q_{0.1}$, $Q_{0.9}$ as the deciles. The MRL function for $Q_{0.1}$, $Q_{0.25}$, $Q_{0.5}$, $Q_{0.75}$ and $Q_{0.9}$ of the distribution of X were predicted for $z = 1$ for both covariate types. The bias of the average estimate (BIAS), monte carlo standard error (σ_{MC}) and mean squared error (MSE) were calculated from 1000 estimates from a sample size of 200 to quantify the performance of each estimator. To fit the nonparametric models we used *biased cross-validation* for choosing the bandwidth h_n of a Gaussian kernel density estimator. After some initial tests the *biased cross-validation* method preformed superiorly to other off the shelf bandwidth selection methods (for a thorough review on bandwidth selection see Jones et al., 1996). Furthermore, the τ parameter was set to the maximum observation $X_{(n)}$ in each iteration for the Bernstein methods.

Approximate Location of Table 1

The results for the additive simulation are summarized in Tables 1 and 2. Since it is correctly specified it is expected that the semiparametric additive MRL model has a lower MSE than the nonparametric models, and this is mostly the case. For the binary covariate the GPLE model, for $Q_{0.1}$ and $Q_{0.25}$, and the Bernstein ($n/4$) model, for $Q_{.50}$–$Q_{.90}$, are competitive in MSE. 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 all points with the binary covariate. With the continuous covariate the proportional model stays competitive in terms of MSE due to the low variance of it's estimate, but has significant bias (more than $2\sigma_{MC}$ away from 0) for $Q_{.10}$, $Q_{.75}$ and $Q_{.90}$, and very close to significant bias for the other points.

Approximate Location of Table 2

With the continuous covariate the Bernstein $(n/4)$ model appears to perform the best out of the nonparametric estimators. It has the lowest MSE for all points, followed by the Bernstein $(n/2)$ with the second lowest and the GPLE has the highest. The results for the binary covariate are not as clear. For $Q_{.10}$ and $Q_{.25}$ the GPLE model performs the best in bias, variance and MSE, for $Q_{.50}$ all three perform similarly, and for the latter two points the Bernstein $(n/4)$ has the lowest variance and MSE. While the results for the binary covariate are not as clear as the continuous, it appears safe to say that the Bernstein $(n/4)$ model is the best choice in terms of the nonparametric estimators for this simulation setting. The setting with the continuous covariate really shows the robustness the Bernstein $(n/4)$ model has to the covariate type.

The results from the proportional simulation are contained in Tables 3 and 4. The proportional model perfoms the best overall for both covariate types. This is the case much more so than in the additive simulation (however, since the simulations were ran with such different data types this should not be taken as a comparison of semiparametric methods).

Approximate Location of Table 3

The additive model is greatly outperformed by the nonparametric methods for the binary covariate. The average estimate of m^a for $Q_{.75}$ and $Q_{.90}$ are less than zero. In fact, of the 1000 estimates 41.1%, 99.9%, and 100% of the estimates for $m^a(Q_{0.5}|1)$, $m^a(Q_{0.75}|1)$, and $m^a(Q_{0.9}|1)$, were less than zero, respectively. As a result the simulation was altered to use $Z' = 1 - Z$, and hence $\beta' = -\beta = 1$, 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 basically the same as in the first simulation. Furthermore, of the 1000 iterations 39.3%, 99.9%, and 100% of the estimates for $m^a(Q_{0.5}|0)$, $m^a(Q_{0.75}|0)$, and $m^a(Q_{0.9}|0)$, were less than zero, respectively. Clearly, the method for ensuring the positivity of m^a did not work.

With the continuous covariate the additive model has the highest bias, and MSE for $Q_{.10}$, $Q_{.25}$ and $Q_{.90}$, and has a lower MSE than the nonparametric methods for $Q_{.50}$, and $Q_{.75}$. Of the 1000 estimates for the additive model for $m(Q_{0.9}|1)$, 16% were less than 0. The nonparametric methods perform more consistently than the additive model in the continuous case, and better overall.

Approximate Location of Table 3

For the binary covariate, the GPLE model has a lower bias and MSE than both of the Bernstein models for $Q_{.10}$, $Q_{.25}$ and $Q_{.50}$, but for the latter time points the Bernstein models perform remarkably better than the GPLE method, and approximately equal to each other. 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 variance of the data. This results in lower variance of all the estimators, and hence the mean squared error reflects the estimators bias more so than the previous setting. For the most part the GPLE model has the lowest bias and highest variance, the Bernstein ($n/4$) has the highest bias and lowest 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. As a result, it is recommended that the Bernstein method be used, with $N = n/2$ for data with not much skew, while $n/4$ (and possibly $n/6$) are good choices when there is more skewness.

5. Data Analysis

A useful application of the MRL function is to cost data. For example $m(t|\boldsymbol{z})$ could express the average cost expenditures for an insurance company over the premium t. The MRL function $\mu(t_i|\boldsymbol{z}_i)$ gives the expected remaining payout, controlling for the predictor variables, for the ith individuals. This information can be compounded over all individuals to forecast remaining cost for the year. We bring attention to this application by analyzing data from the 1997 cross-section Vietnam Living Standards Survey (VLSS), which is part of the World Bank's Living Standards Measurement Study. The VLSS contains responses from 27,765 individuals on variables like total medical expenditure, gender, health insurance coverage, age, and others. There is no censoring in the study, but we create censored observations to show the consistency of our methods. The outcome of interest is the total medical expenditure measured in 1998 Vietnamese dong (VND), conditional on positive response.

Approximate Location of Figure 1

Figure 1 shows the estimates of the MRL function using the additive, proportional, and Bernstein method ($N = n/4$), separated by health insurance coverage. The nonparametric estimates show that there appears to be a difference between the two groups for until a cumulative expenditure of around 40, whereafter the two groups appear to follow a similar pattern. The proportional and additive models fail to pick up the difference in behavior between the two groups and their estimates become farther away from the nonparametric. The economic justification for the behaviors of the two groups becoming similar is that insurance companies have a cap at some value where they will no longer pay for medical services. As a result, for an individual with a cumulative expenditure of 10 VND we expect them to pay 11 VND more if they have insurance and around 15 VND more if they do not. However when they have a cumulative expenditure of 40 VND the insurance companies cap has most likely been met so we expect there to be no significance difference in the remaining payout.

6. Discussion

The models proposed in section 3 will predict MRL function for a given value of $\boldsymbol{z}$. These methods will not, however, give an estimate of the regression parameter like the additive and proportional means models do. For an estimate on the effectiveness or risk associated with a covariate there are many hazard based model that suffice. The benefit in using a MRL function type model over a hazard based model is the estimate of $m(t|\boldsymbol{z})$. This is a function that gives the most commonly used statistic, the average, based on the current cumulative observation and the individual explanatory variables.

The estimate of $m(\cdot|\boldsymbol{z})$ presented herein provides a dynamic (time-varying) estimate of the regression. For instance $m_j'(t|\boldsymbol{z}) \equiv \partial m(t|\boldsymbol{z})/\partial z_j$ provides a time varying estimate of the regression coefficient for the jth predictor. If we find that $m_j'(t|\boldsymbol{z})$ does not vary with t then it suggests that an additive model might be suitable for the given data. On the other hand we can also estimate $\partial \log\{m(t|\boldsymbol{z})\}/\partial z_j$ and if we find that it does not vary with time then a proportional mean model might be suitable for the observed data. In this way the nonparametric methods in section 3 is as a tool for checking the adequacy of a semi-parametric or parametric MRL models.

References

Abdous, B. and Berred, A. (2005), "Mean residual life estimation," *J. Statist. Plann. Inference*, 132, 3–19.

Babu, G. J., Canty, A. J., and Chaubey, Y. P. (2002), "Application of Bernstein polynomials for smooth estimation of a distribution and density function," *J. Statist. Plann. Inference*, 105, 377–392.

Beran, R. (1981), "Nonparametric regression with randomly censored survival data," Technical report, Dept. Statistics, Univ. California, Berkely.

Carnicer, J. M. and Peña, J. M. (1993), "Shape preserving representations and optimality of the Bernstein basis," *Adv. Comput. Math.*, 1, 173–196.

Chaubey, Y. P. and Sen, A. (2008), "Smooth estimation of mean residual life under random censoring," in *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*, Beachwood, OH: Inst. Math. Statist., vol. 1 of *Inst. Math. Stat. Collect.*, pp. 35–49.

Chaubey, Y. P. and Sen, P. K. (1999), "On smooth estimation of mean residual life," *J. Statist. Plann. Inference*, 75, 223–236, the Seventh Eugene Lukacs Conference (Bowling Green, OH, 1997).

Chen, Y. Q. (2007), "Additive expectancy regression," *J. Amer. Statist. Assoc.*, 102, 153–166.

Chen, Y. Q. and Cheng, S. (2005), "Semiparametric regression analysis of mean residual life with censored survival data," *Biometrika*, 92, 19–29.

— (2006), "Linear life expectancy regression with censored data," *Biometrika*, 93, 303–313.

Chen, Y. Q., Jewell, N. P., Lei, X., and Cheng, S. C. (2005), "Semiparametric estimation of proportional mean residual life model in presence of censoring," *Biometrics*, 61, 170–178.

Dabrowska, D. M. (1987), "Nonparametric regression with censored survival time data," *Scand. J. Statist.*, 14, 181–197.

— (1989), "Uniform consistency of the kernel conditional Kaplan-Meier estimate," *Ann. Statist.*, 17, 1157–1167.

— (1992), "Variable bandwidth conditional Kaplan-Meier estimate," *Scand. J. Statist.*, 19, 351–361.

Feller, W. (1965), *An Introduction to Probability Theory and its Applications*, VOL. II, New York: Wiley.

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

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

Hall, W. J. and Wellner, J. (1981), “Mean residual life,” in *Statistics and related topics (Ottawa, Ont., 1980)*, Amsterdam: North-Holland, pp. 169–184.

Jones, M. C., Marron, J. S., and Sheather, S. J. (1996), “A Brief Survey of Bandwidth Selection for Density Estimation,” *Journal of the American Statistical Association*, 91, 401–407.

Liu, S. and Ghosh, S. (2008), “Regression Analysis of Mean Residual Life Function,” Institute of Statistical Mimeo Series # 2613, North Carolina State University, Raleigh, NC 27695.

Maguluri, G. and Zhang, C.-H. (1994), “Estimation in the mean residual life regression model,” *J. Roy. Statist. Soc. Ser. B*, 56, 477–489.

Oakes, D. and Dasu, T. (1990), “A note on residual life,” *Biometrika*, 77, 409–410.

Pérez, C. I. and Manteiga, W. G. (1999), “Strong representation of a generalized product-limit estimator for truncated and censored data with some applications,” *J. Nonparametr. Statist.*, 10, 213–244.

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

Robins, J. and Tsiatis, A. A. (1992), “Semiparametric estimation of an accelerated failure time model with time-dependent covariates,” *Biometrika*, 79, 311–319.

Stone, C. J. (1977), “Consistent Nonparametric Regression,” *The Annals of Statistics*, 5, 595–620.

Stute, W. (1996), “Distributional Convergence under Random Censorship When Covariables Are Present,” *Scandinavian Journal of Statistics*, 23, 461–471.

Sun, L. and Zhang, Z. (2009), “A Class of Transformed Mean Residual Life Models With Censored Survival Data,” *Journal of the American Statistical Association*, 104, 803–815.

Therneau, T. M. and Grambsch, P. M. (2000), *Modeling survival data: extending the Cox model*, Statistics for Biology and Health, New York: Springer-Verlag.

Yang, G. (1977), “Life expectancy under random censorship,” *Stochastic Processes Appl.*, 6, 33–39.

Alexander McLain, Division of Epidemiology, Statistics and Prevention Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA.

E-mail: mclaina@mail.nih.gov

Appendix

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

(A0) $\tau_X = \tau_T < \infty$.

(A1) T, C, and $\boldsymbol{Z}$ 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 $I_\delta = [-\boldsymbol{z}_1 - \delta, \boldsymbol{z}_2 + \delta]^q \subseteq I = [-Z^*, Z^*]^q$ with $\delta > 0$ such that $0 < \inf\{q(\boldsymbol{z}) : \boldsymbol{z} \in I_\delta\} < \sup\{q(\boldsymbol{z}) : \boldsymbol{z} \in I_\delta\} < \infty$. Furthermore, for all $\boldsymbol{z} \in I_\delta$:

i. T and C are conditionally independent at $\boldsymbol{Z} = \boldsymbol{z}$.

ii. There exists a, $b \in \mathbb{R}$ such that $S_X(t|\boldsymbol{z}) \geq \theta > 0$ for all $t \in [a, b]$.

(A3) The first and all partial second derivatives with respect to T and $\boldsymbol{Z}$ of $q(\boldsymbol{z})$, $H(t|\boldsymbol{z})$ and $C(t|\boldsymbol{z})$ exist and are continuous for $(t, \boldsymbol{z}) \in [0, \tau_T) \times I_\delta$.

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

Proof of Lemma 1. Proof of (I) follows from the work of Stone (1977) and Pintér (2001) on consistency of nonparametric regression estimates. Pintér (2001) showed, under the conditions of Stone (1977), that $||S_n^P - S||_{\tau_T} \to 0$ as $n \to \infty$ almost surely. Since $\tau_T < \infty$ and S_T is continuous, $m(\cdot|\boldsymbol{z})$ is a bounded continuous function. As a result, $m(\cdot|\boldsymbol{z})$ is a Hadamard differentiable functional of $S(t|\boldsymbol{z})$, hence the same convergence properties can be claimed.

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^{B0} as

$$S_N^{B0}(t|\boldsymbol{z},\tau) \equiv \sum_{k=0}^{N} S_T\left(\frac{\tau k}{N}\middle|\boldsymbol{z}\right)\psi_{k,N}\left(\frac{t}{\tau}\right),$$

then Feller (1965) shows that if S_T is a bounded continuous function $||S_N^{B0} - S_T||_{\tau_T} \rightarrow 0$ as $N \rightarrow \infty$. As a result, since $||S_{n,N}^B - S_T||_{\tau_T} \leq ||S_{n,N}^B - S_N^{B0}||_{\tau_T} + ||S_N^{B0} - S_T||_{\tau_T}$ it suffices to show that $||S_{n,N}^B - S_N^{B0}||_{\tau_T} \rightarrow 0$. To see this notice that,

$$S_{n,N}^B(t|\boldsymbol{z},h_n,N,\tau) - S_N^{B0}(t|\boldsymbol{z},\tau) = \sum_{k=0}^{N}\left\{S_n^P\left(\frac{\tau k}{N}\middle|\boldsymbol{z},h_n\right) - S_T\left(\frac{\tau k}{N}\middle|\boldsymbol{z},h_n\right)\right\}\psi_{k,N}\left(\frac{t}{\tau}\right),$$

and hence $||S_{n,N}^B - S_N^{B0}||_{\tau_T} \leq \max_{0\leq k\leq N}|S_n^P\left(\frac{\tau k}{N}|\boldsymbol{z},h_n\right) - S_T\left(\frac{\tau k}{N}|\boldsymbol{z},h_n\right)| \leq ||S_n^P - S_T||_{\tau_T} \rightarrow 0$ as $n \rightarrow \infty$ almost surely. Therefore, $||S_{n,N}^B - S_T||_{\tau_T} \rightarrow 0$ as $n, N \rightarrow \infty$. $\square$

Proof of Theorem 1. The pointwise normality of m_P follows from the functional delta method. As a result, we make the same assumptions made by Pérez and Manteiga (1999), setting the left truncated variable to 0, which amounts to $(A1)$–$(A4)$. For the moment we write $m(t|\boldsymbol{z})$ as $m(t|S)$ to show the dependence of the MRLF on the survivor function, and to suppress the dependence on $\boldsymbol{z}$ for brevity. From $(A0)$ we have that $S(\cdot|\boldsymbol{z}) \in \mathbb{E} = \{f : \int f < \infty\}$, and $m(t|S) < \infty$ almost surely. First, we establish the existence and form of the Hadamard Derivative of m at S with respect to $||\cdot||_{\tau_T}$. In this situation, for $g_n = n^{-1/2}$ and $S_n = S + g_n G$, we wish to find $\phi(\cdot; G)$ such that

$$\phi(t;G) = \lim_{n\rightarrow\infty}\frac{1}{g_n}\{m(t|S_n) - m(t|S)\}. \tag{14}$$

where $||S_n - S||_{\tau_T} \to 0$. As a result, suppressing dummy arguments

$$\begin{aligned}\phi(t;G) &= \lim_{n\to\infty} \frac{1}{g_n}\left\{\frac{\int_t^\infty S + g_n G}{S(t)+g_nG(t)} - \frac{\int_t^\infty S}{S(t)}\right\} \\ &= \lim_{n\to\infty} \frac{1}{g_n}\left\{\frac{g_nS(t)\int_t^\infty G - g_nG(t)\int_t^\infty S}{S(t)\{S(t)+g_nG(t)\}}\right\} \\ &= \frac{1}{S(t)}\left\{\int_t^\infty G - G(t)m(t)\right\}. \end{aligned} \tag{15}$$

By the functional delta method we have $\sqrt{nh_n}\{m_n^P(t|\boldsymbol{z},h_n) - m(t|\boldsymbol{z})\} \to^d \phi(t;G)$, where $\phi(t;G) \sim N(0, E_G\{\phi(t;G)^2\})$. What is left to find is,

$$\begin{aligned}E_G\{\phi(t;G)^2\} &= E_G\left[\frac{1}{S(t)}\left\{\int_t^\infty G - G(t)m(t)\right\}\right] \\ &= \frac{1}{S(t)^2}E_G\left\{(\int_t^\infty G)(\int_t^\infty G) - 2G(t)m(t)(\int_t^\infty G) + G(t)^2m(t)^2\right\} \\ &= \frac{1}{S(t)^2}\left[\int_t^\infty\int_t^\infty E_G\{G(u)G(w)\}dudw\right. \\ &\quad \left. -2m(t)\int_t^\infty E_G\{G(t)G(u)\}du + m(t)^2E_G\{G(t)^2\}\right]. \end{aligned} \tag{16}$$

From the definition of $\Gamma(y,t|\boldsymbol{z})$ we have

$$\begin{aligned}(16) &= \frac{(\int K^2)}{S(t|\boldsymbol{z})^2q(\boldsymbol{z})}\left[\int_t^\infty\int_t^\infty S(u|\boldsymbol{z})S(w|\boldsymbol{z})D(u\wedge w|\boldsymbol{z})dudw\right. \\ &- \left. 2m(t|\boldsymbol{z})S(t|\boldsymbol{z})D(t|\boldsymbol{z})\int_t^\infty S(u|\boldsymbol{z})du + m(t|\boldsymbol{z})^2S(t|\boldsymbol{z})^2D(t|\boldsymbol{z})\right]. \end{aligned} \tag{17}$$

The 1st portion of the inner expression of (17) can be expressed as

$$\begin{aligned}&\int_t^\infty S(u|\boldsymbol{z})\left\{\int_t^u S(w|\boldsymbol{z})D(w|\boldsymbol{z})dw + D(u|\boldsymbol{z})\int_u^\infty S(w|\boldsymbol{z})dw\right\}du \\ &= \int_t^\infty S(u|\boldsymbol{z})B_t(u|\boldsymbol{z})du + \int_t^\infty S(u|\boldsymbol{z})^2m(u|\boldsymbol{z})D(u|\boldsymbol{z})du\end{aligned}$$

where $B_t(u|\boldsymbol{z}) = \int_t^u S(w|\boldsymbol{z})D(w|\boldsymbol{z})dw$. The 2nd and 3rd portions of the inner expression of (17) can be simplified to

$$-2m(t|\boldsymbol{z})^2 S(t|\boldsymbol{z})^2 D(t|\boldsymbol{z}) + m(t|\boldsymbol{z})^2 S(t|\boldsymbol{z})^2 D(t|\boldsymbol{z}) = -m(t|\boldsymbol{z})^2 S(t|\boldsymbol{z})^2 D(t|\boldsymbol{z}).$$

Putting everything together we have,

$$\begin{aligned} E_G\{\phi(t;G)^2\} &= \frac{(\int K^2)}{S(t|\boldsymbol{z})^2 q(\boldsymbol{z})} \left[\int_t^\infty S(u|\boldsymbol{z}) B_t(u|\boldsymbol{z}) du + \int_t^\infty S(u|\boldsymbol{z})^2 m(u|\boldsymbol{z}) D(u|\boldsymbol{z}) du \right. \\ &\quad \left. - m(t|\boldsymbol{z})^2 S(t|\boldsymbol{z})^2 D(t|\boldsymbol{z}) \right], \end{aligned}$$

which proves the theorem.

Table 1: Summarized results for the additive simulation with the binary covariate.

	$m(Q_{0.1}\|1)$			$m(Q_{0.25}\|1)$			$m(Q_{0.5}\|1)$			$m(Q_{0.75}\|1)$			$m(Q_{0.9}\|1)$		
	BIAS	σ_{MC}	MSE	BIAS	σ_{MC}	MSE	BIAS	σ_{MC}	MSE	BIAS	σ_{MC}	MSE	BIAS	σ_{MC}	MSE
$m^a(\cdot)$	0.264	**0.385**	0.217	0.235	**0.414**	0.227	0.154	**0.497**	**0.270**	**-0.064**	0.732	**0.540**	**-0.581**	1.191	**1.756**
$m^B(\cdot)$	0.290	0.407	0.250	0.237	0.436	0.246	0.104	0.515	0.276	-0.208	**0.723**	0.565	-0.824	**1.140**	1.979
$m^{B2}(\cdot)$	0.239	0.395	0.213	0.194	0.472	0.220	0.075	0.519	0.274	-0.216	0.752	0.613	-0.822	1.199	2.113
$m^P(\cdot)$	**0.191**	0.386	**0.185**	**0.153**	0.425	**0.204**	**0.051**	0.532	0.285	-0.215	0.798	0.683	-0.806	1.279	2.286
$m^{pr}(\cdot)$	0.221	0.513	0.311	0.209	0.557	0.354	0.138	0.680	0.481	-0.145	0.942	0.907	-0.945	1.397	2.845

Table 2: Summarized results for the additive simulation with the continuous covariate.

	$m(Q_{0.1}\|1)$			$m(Q_{0.25}\|1)$			$m(Q_{0.5}\|1)$			$m(Q_{0.75}\|1)$			$m(Q_{0.9}\|1)$		
	BIAS	σ_{MC}	MSE	BIAS	σ_{MC}	MSE	BIAS	σ_{MC}	MSE	BIAS	σ_{MC}	MSE	BIAS	σ_{MC}	MSE
$m^a(\cdot)$	-0.253	0.344	**0.182**	-0.292	0.390	**0.238**	-0.410	0.510	**0.428**	-0.713	0.825	**1.188**	**-1.410**	1.515	**4.285**
$m^B(\cdot)$	**-0.154**	0.519	0.293	**-0.217**	0.564	0.366	**-0.379**	0.695	0.627	-0.733	1.038	1.616	-1.467	1.734	5.159
$m^{B2}(\cdot)$	-0.211	0.511	0.305	-0.262	0.567	0.390	-0.403	0.714	0.672	-0.730	1.095	1.731	-1.446	1.864	5.565
$m^P(\cdot)$	-0.264	0.510	0.330	-0.299	0.581	0.426	-0.422	0.736	0.719	**-0.708**	1.177	1.887	-1.384	2.146	6.521
$m^{pr}(\cdot)$	-0.486	**0.241**	0.294	-0.550	**0.276**	0.379	-0.749	**0.356**	0.688	-1.267	**0.545**	1.903	-2.424	**0.897**	6.680

Table 3: Summarized results for the proportional simulation with the binary covariate. MSE* corresponds to the MSE×1000.

	$m(Q_{0.1}\|1)$			$m(Q_{0.25}\|1)$			$m(Q_{0.5}\|1)$		
	BIAS	σ_{MC}	MSE*	BIAS	σ_{MC}	MSE*	BIAS	σ_{MC}	MSE*
$m^a(\cdot)$	-0.042	0.028	2.53	-0.094	0.030	9.7	-0.21	0.042	45.8
$m^B(\cdot)$	0.024	0.027	1.32	0.028	0.028	1.5	0.033	0.032	2.0
$m^{B2}(\cdot)$	0.013	0.028	0.92	0.014	0.029	1.0	0.016	0.034	1.4
$m^P(\cdot)$	**0.001**	0.029	0.84	**0.001**	0.031	1.0	**0.0008**	0.039	1.5
$m^{pr}(\cdot)$	-0.005	**0.026**	**0.70**	-0.005	**0.023**	**0.6**	-0.007	**0.020**	**0.4**

	$m(Q_{0.75}\|1)$			$m(Q_{0.9}\|1)$		
	BIAS	σ_{MC}	MSE*	BIAS	σ_{MC}	MSE*
$m^a(\cdot)$	-0.360	0.057	133.0	-0.470	0.071	226.2
$m^B(\cdot)$	0.035	0.047	3.43	**0.010**	0.060	3.67
$m^{B2}(\cdot)$	0.016	0.055	3.31	-0.022	0.067	4.91
$m^P(\cdot)$	**0.007**	0.083	6.86	0.136	0.098	28.0
$m^{pr}(\cdot)$	-0.008	**0.018**	**0.38**	-0.014	**0.019**	**0.56**

Table 4: Summarized results for the proportional simulation with the continuous covariate. MSE* corresponds to the MSE×1000.

	$m(Q_{0.1}\|1)$			$m(Q_{0.25}\|1)$			$m(Q_{0.5}\|1)$		
	BIAS	σ_{MC}	MSE*	BIAS	σ_{MC}	MSE*	BIAS	σ_{MC}	MSE*
$m^a(\cdot)$	0.066	0.024	4.9	0.055	0.024	3.7	0.025	0.026	1.3
$m^B(\cdot)$	0.030	0.033	2.0	0.034	0.033	2.2	0.042	0.035	3.0
$m^{B2}(\cdot)$	0.020	0.033	1.5	0.024	0.033	1.7	0.030	0.036	2.2
$m^P(\cdot)$	0.012	0.035	1.3	0.014	0.035	1.4	0.019	0.039	1.9
$m^{pr}(\cdot)$	**-0.006**	**0.020**	**0.4**	**-0.007**	**0.020**	**0.5**	**-0.009**	**0.020**	**0.5**

	$m(Q_{0.75}\|1)$			$m(Q_{0.9}\|1)$		
	BIAS	σ_{MC}	MSE*	BIAS	σ_{MC}	MSE*
$m^a(\cdot)$	-0.036	0.036	2.5	-0.105	0.057	14.2
$m^B(\cdot)$	0.051	0.040	4.2	0.055	0.052	5.8
$m^{B2}(\cdot)$	0.038	0.042	3.3	0.042	0.059	5.2
$m^P(\cdot)$	0.027	0.050	3.2	0.038	0.081	7.9
$m^{pr}(\cdot)$	**-0.012**	**0.022**	**0.6**	**-0.017**	**0.026**	**1.0**

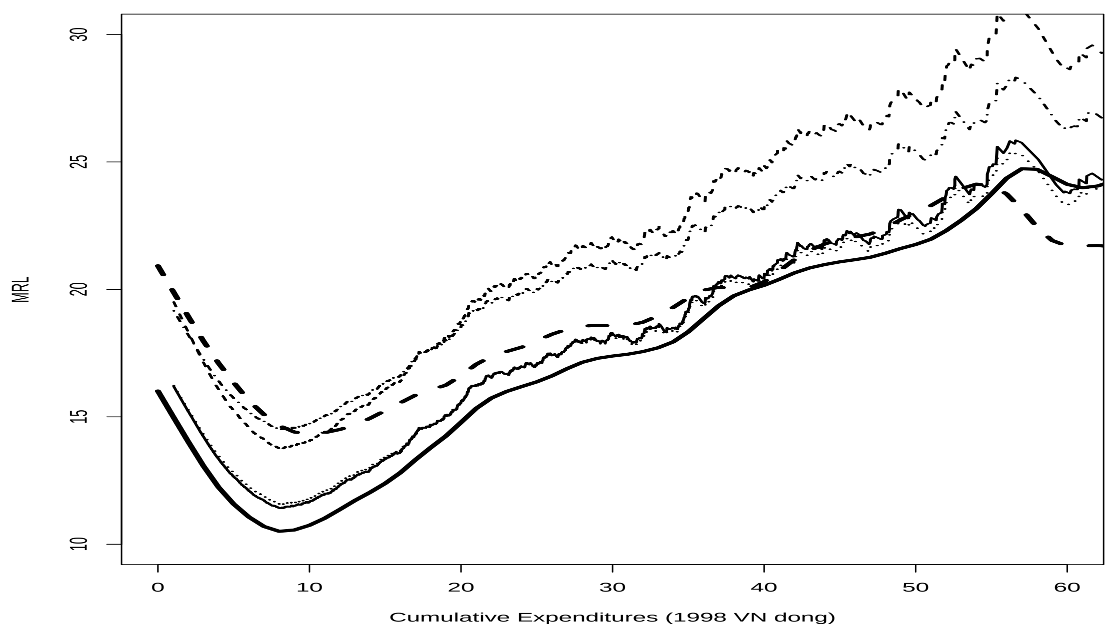

Figure 1: MRL estimation of the Vietnam medical expenditures where individuals with insurance have estimates for the additive (dotted), the proportional (solid) and the Bernstein (bold solid), and individuals without insurance have estimates for the additive (dot-dash), the proportional (dash) and the Bernstein (bold dash).