

# Kernel Estimation of the Baseline Function in the Cox Model

*Lule Basha,  
Fatmir Hoxha,*

Department of Applied Mathematics, Faculty of Natural Science,  
University of Tirana, Tirana, Albania

Doi: 10.19044/esj.2019.v15n6p105      [URL:http://dx.doi.org/10.19044/esj.2019.v15n6p105](http://dx.doi.org/10.19044/esj.2019.v15n6p105)

---

## Abstract

Survival analysis is the analysis of time-to-event data. Two important functions in the analysis of survival data are the survival function and the hazard function. The Kaplan-Meier method is widely used to estimate the survival function. One of the objectives of the analysis of survival data might be to examine whether survival times are related to other features. A popular regression model for the analysis of survival data is the Cox proportional hazard regression model. The most commonly used approaches, for the baseline survival function, are the Breslow and Kalbfleisch-Prentice methods. These methods provide a step function estimate of the survivor function and in many instances a continuous estimate would be preferable. For these reason, in this paper we proposed a kernel smoothing technique for baseline estimator, based on Kalbfleisch-Prentice method. We start with kernel smoothing of baseline hazard function, based on Kalbfleisch-Prentice estimator and epanechnikov kernel, than we use it to calculate the baseline survival function. To evaluate the usefulness of the kernel estimator of the baseline function, in the case of right censoring, based on Kalbfleisch-Prentice estimator we conduct simulation studies across a range of conditions, by varying the sample size and censoring rate. We compare it with the smoothing of the Breslow estimator regarding bias.

---

**Keywords:** Kernel smoothing of Breslow estimator, Kernel smoothing of Kalbfleisch Prentice estimator, Survival analysis, Cox model, Simulations

## Introduction

Survival analysis is generally defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest, or known as the analysis of time-to-event data. The outcome is often referred to as a failure time, survival time, or event time. In survival

analysis exist the possibility of not observing the event of interest for some individuals, this is called censoring. These incomplete observations cannot be ignored, but need to be handled differently. Censoring is an important issue in survival analysis, representing a particular type of missing data. Censoring that is random and non informative is usually required in order to avoid bias in a survival analysis. Survival methods correctly incorporate information from both censored and uncensored observations in estimating important model parameters.

Two important functions in the analysis of survival data are the survival function and the hazard function. They are key concepts for describing the distribution of event times. The survival function gives, for every time, the probability of surviving (or not experiencing the event) up to that time. The hazard function gives the potential that the event will occur, per time unit, given that an individual has survived up to the specified time. There are parametric, nonparametric and semiparametric approaches to estimate the survival function and the hazard function. The Kaplan–Meier (or product-limit) estimator (Kaplan and Meier, 1958) is a non-parametric estimator of the survival function and Nelson-Aalan estimator is a non-parametric estimator of the cumulative hazard function. These two estimators are the most useful methods in survival analysis.

One of the objectives of the analysis of survival data might be to examine whether survival times are related to other features. A number of models are available to analyze the relationship of a set of predictor variables with the survival time. A popular regression model for the analysis of survival data is the Cox proportional hazard regression model (Cox, 1972). The Cox regression model provides useful and easy to interpret information regarding the relationship of the hazard function to predictors (Choi et al., 2014; Mao and Wang, 2010). The regression parameter and the baseline function are the two unknown parameters in this model. The regression coefficients can be estimated using the partial likelihood without specifying the baseline hazard function. The partial likelihood estimator can be obtained through Newton-Raphson method easily when there are no ties among exact failure time in the recorded data. If ties are presented, Breslow (1974), Efron (1977), and Kalbfleisch and Prentice (1973) have proposed several approximated partial likelihood functions, for handling ties in the Cox proportional hazard model. In particular, and in contrast with parametric models, it makes no assumptions about the shape of the baseline hazard function. This is why the Cox proportional hazards model is referred to as a semi-parametric method, that is, a method in which survival times are assumed to be related to the explanatory variables in a particular way, but no assumptions are made on the overall shape of the survival times, that is the shape of the hazard function need not be specified.

After obtaining the estimators, we may interest in estimating the baseline survival function. In the discussion followed by Cox's paper, all these estimators reduce to the product limit estimate (Kaplan and Meier, 1958) when there are no covariates. In previous works, more attention has been paid to the estimation of the regression parameter than to the estimation of the baseline function. Oakes (1972) suggested a step-function estimate instead of the point-wise estimate for baseline hazard function. To estimate the baseline function, the Breslow estimator (Breslow, 1974) or the Kalbfleisch-Prentice (KP) estimator (Kalbfleisch and Prentice, 1973; 1980) can be used. The Breslow estimator uses the profile likelihood approach by extending the Nelson-Aalan estimator (Lin, 2007; Breslow and Wellner, 2007; Huang and Strawderman, 2006). Kalbfleisch-Prentice (1973) proposed a step function estimate, where the baseline hazard function is assumed to be a constant between convenient (but arbitrary) subdivisions of the time scale, which is analogous to the Kaplan-Meier estimator.

Another known estimator of the baseline function is a kernel estimator, introduced by Ramlau-Hansen (1983a; 1983b), which is obtained by smoothing the Breslow estimator of the cumulative baseline function. Guilloux et al., (2016) introduced an adaptive kernel estimation of the baseline function in the Cox model with high-dimensional covariates. Selingerova et al., (2016) proposed kernel estimator of the conditional hazard function, in the Cox model. In this work we propose a method by smoothing the Kalbfleisch-Prentice estimator. We compare the smoothing of Breslow estimator with the proposed estimator, running a series of simulations. Some specified models are used to see the finite sample performance of them.

The paper is organized as follows. Section 2 introduces the smoothing of the Kalbfleisch-Prentice estimator function for right-censored data, based on kernel techniques. The finite sample properties of the new estimator and the comparison with the smoothing of Breslow estimator are investigated in Section 3 and Section 4 concludes.

## Methodology

Let  $Y$  be a variable of interest with density  $f$  and distribution function  $F$ , and we denote by  $C_1, \dots, C_n$  a nonnegative random right censoring variable with continuous distribution function  $G$ . We assume that  $Y$  is independent of  $C$ . In the case when we consider the random right censoring, the variable is not completely observed. One can only observe  $(T_i, \delta_i)$  where  $T_i = \min(X_i, C_i)$  and  $\delta_i = I(X_i \leq C_i)$  with  $I(\cdot)$  being the indicator function, such that is equal to 1 when observation is complete and 0 when it is censored. The survival function can be estimated using the well-known Kaplan-Meier estimator (Kaplan and Meier, 1958), given as:

$$\hat{S}(t) = 1 - \hat{F}(t) = \prod_{i: T_i \leq t} \left( 1 - \frac{1}{\sum_{j=1}^n 1_{\{T_j \geq T_i\}}} \right)^{\delta_i} \quad (1)$$

The Kaplan-Meier estimator is the simplest way of computing the survival probabilities and graphical presentation.

The Cox proportional hazard model, introduced by Cox (1972), is a regression model that specifies the conditional hazard function of the failure time for a given set of covariates. The hazard function is then defined by

$$h(t | \mathbf{X}) = h_0(t) \exp(\beta^T \mathbf{X}) \quad (2)$$

where  $\mathbf{X} = (X_1, \dots, X_p)^T$  a  $p$ -dimensional vector of covariates is,  $\beta = (\beta_1, \dots, \beta_p)^T$  is the vector of regression coefficients and  $h_0(t)$  is the baseline hazard function. The Cox proportional hazards model assumes that the hazard function at time  $t$  for a given covariate vector is the product of an arbitrary baseline hazard function and an exponential function of the linear combination of the covariates. The Cox model is a semiparametric model, because the baseline hazard function is left unspecified. If the model is correct, then the hazard at time  $t$  for an individual with covariate vector  $\mathbf{X} = x^* = (x_1^*, \dots, x_p^*)^T$  is

$$h(t | \mathbf{X} = x^*) = h_0(t) \exp(\beta^T x^*) \quad (3)$$

The primary method of analysis in estimating the regression coefficients is called partial likelihood method. It formed the basis of the Cox (1972) analysis of proportional hazards model, and was discussed further in Cox (1975). Efron (1977) and Oakes (1977) developed asymptotic efficiency formulae for maximum partial likelihood estimators within Cox's (1972) regression model, while Tsiatis (1981) was the first to prove consistency of such estimators.

The full likelihood function of the data  $(t_i, \delta_i, x_i)$ , conditional on  $x_1, \dots, x_p$ , and considering that the censoring time is noninformative, can be written as:

$$L(\beta | t, \delta, x) = \prod_{i=1}^n [f(t_i | x_i)]^{\delta_i} [S(t_i | x_i)]^{1-\delta_i} = \prod_{i=1}^n [h(t_i | x_i)]^{\delta_i} S(t_i | x_i) \quad (4)$$

The partial likelihood function of the data  $(t_i, \delta_i, x_i)$  is

$$L(\beta) = \prod_{i=1}^k \frac{\exp(\beta x_{(i)})}{\sum_{j \in R_{(i)}} \exp(\beta x_j)} \quad (5)$$

where  $R_{(t_i)}$  is the set of patients who are still at risk of experiencing the event at time  $t_i$ . Partial likelihood has also been used by Peto and Peto (1972) for constructing asymptotically efficient rank test statistics in the two-sample problem with censored survival data.

One of the primary quantities desired from a survival analysis point of view is estimated survival curve. If no model is used to fit survival data, a survival curve can be estimated using a Kaplan-Meier method. When a Cox model is used to fit survival data, survival curves can be obtained that adjust for the explanatory variables used as predictors (Kleinbaum and Klein, 2005). These are called adjusted survival curves. For a random individual with covariates  $\mathbf{X} = \mathbf{x}^* = (x_1^*, \dots, x_p^*)^T$ , the conditional survival function of time can be estimated as:

$$\hat{S}(t | X = \mathbf{x}^*) = \hat{S}_0(t)^{\exp(\hat{\beta}^T \mathbf{x}^*)} \tag{6}$$

where  $\hat{S}_0(t)$  is the estimated baseline survival function and  $\hat{\beta}$  is the regression coefficients that can be estimated using the partial likelihood without specifying the baseline hazard function (Cox, 1972; 1975). Because of the relationship of hazard to survival

$$\hat{S}_0(t) = \exp\{-\hat{H}_0(t)\} \tag{7}$$

we can write

$$\hat{S}(t | X = \mathbf{x}^*) = e^{-\exp(\hat{\beta}^T \mathbf{x}^*) \hat{H}_0(t)} \tag{8}$$

where  $\hat{H}_0(t)$  is an estimator for the cumulative baseline hazard function. Then the estimate of conditional survival function can be taken from estimating either the estimated baseline survival function, or the cumulative baseline hazard function. To estimate the baseline survival function, the Breslow estimator (Breslow, 1974), or the Kalbfleisch Prentice estimator (Kalbfleisch and Prentice, 1973; 1980) can be used.

**The Breslow estimator and kernel smoothing**

To obtain the baseline hazards function, Breslow (1972) starts by the full likelihood function in equation (4) with  $\beta$  replace by  $\hat{\beta}$ . Breslow proposed a nonparametric maximum likelihood estimator for the cumulative baseline hazard function, which in the situation of no ties between the observed event times can be written as:

$$\hat{H}_{0,BR}(t) = \sum_{i:t_{(i)} < t} \hat{h}_0(t_{(i)}) = \sum_{j:t_j \leq t} \left\{ \frac{\delta_j}{\sum_{k \in R(t_j)} \exp(\hat{\beta}^T x_k)} \right\} \tag{9}$$

Thus, the Breslow survival function estimator for a subject with covariate vector can be derived as:

$$\hat{S}(t | X = x^*) = \hat{S}_0(t)^{\exp(\hat{\beta}^T x^*)} = e^{-\exp(\hat{\beta}^T x^*) \hat{H}_0(t)} \tag{10}$$

Another known estimator of the baseline function is a kernel estimator, introduced by Ramlau-Hansen (1983a; 1983b), which is obtained by smoothing the Breslow estimator of the cumulative baseline hazard function. Kernel estimator of baseline hazard function has the form:

$$\hat{h}_{0,BRK}(t) = \frac{1}{b} \int K\left(\frac{t-u}{b}\right) d\hat{H}_0(u) = \frac{1}{b} \sum_{i=1}^n K\left(\frac{t-T_i}{b}\right) \frac{\delta_i}{\sum_{j=1}^n I(T_j \geq T_i) e^{\hat{\beta} x_j}} \tag{11}$$

where  $K$  is a kernel function,  $0 < b \equiv b_n$  is a bandwidth sequence. Then, the smoothing baseline survival function estimate is given by substituting in equation (10) the cumulative baseline hazard function taken from equation (11). The choice of the bandwidth in kernel estimation is crucial, in particular when one is interested in establishing non-asymptotic adaptive inequalities. Ramlau-Hansen (1981) has suggested the cross-validation method to select the bandwidth but without any theoretical guarantees.

**The Kalbfleisch-Prentice estimator and the proposed kernel smoothing**

The Kalbfleisch-Prentice estimator, proposed by Kalbfleisch and Prentice (1973; 1980), for survival functions uses the discrete failure time to approach a continuous function. Assume that baseline survival function has only jump points on the  $k$  distinct failure times  $t_{(1)}, \dots, t_{(k)}$ . Then, replacing

$S(t | x)$  by  $S_0(t)^{\hat{\beta}^T x}$  to the full likelihood function (Weng, 2007) we have

$$\begin{aligned} &L(S_0(t_{(1)}), \dots, S_0(t_{(k)}) | t, x) \\ &= \prod_{i=1}^n [S(t_i | x_i) - S(t_i + \varepsilon | x_i)]^{\delta_i} [S(t_i + \varepsilon | x_i)]^{1-\delta_i} \\ &= \prod_{i=1}^n [S_0(t_i)^{\exp(\hat{\beta} x_i)} - S_0(t_i + \varepsilon)^{\exp(\hat{\beta} x_i)}]^{\delta_i} [S_0(t_i + \varepsilon)^{\exp(\hat{\beta} x_i)}]^{1-\delta_i} \tag{12} \\ &= \prod_{1 \leq i \leq k} \prod_{j \in D_i} (S_0(t_{(i)})^{\exp(\hat{\beta} x_j)} - S_0(t_{(i+1)})^{\exp(\hat{\beta} x_j)}) \prod_{j \in C_i} S_0(t_{(i+1)})^{\exp(\hat{\beta} x_j)} \end{aligned}$$

Here,  $D_i$  is the set of individuals who failed at time  $t_i$  and  $C_i$  is the set of individuals censored in  $[t_{(i)}, t_{(i+1)})$ ,  $i = 0, \dots, k$ . Let  $\alpha_i = P(T > t_{(i)} | T \geq t_{(i)}, x = 0)$  denote the conditional survival probability at time  $t_i$  for a baseline subject. The Kalbfleisch-Prentice baseline survival function can be estimated as:

$$\hat{S}_{0,KP}(t) = \prod_{i:t_i \leq t} \hat{\alpha}_i^{\delta_i} \tag{13}$$

This leads to the following likelihood function

$$L = \prod_{1 \leq i \leq k} \left[ \prod_{j \in D_i} \left[ \left( \prod_{t_{(m)} < t_{(i)}} \alpha_m \right)^{\exp(\hat{\beta}x_j)} - \left( \prod_{t_{(m)} \leq t_{(i)}} \alpha_m \right)^{\exp(\hat{\beta}x_j)} \right] \right] \left[ \prod_{j \in C_i} \left( \prod_{t_{(m)} \leq t_{(i)}} \alpha_m \right)^{\exp(\hat{\beta}x_j)} \right] \tag{14}$$

$$= \prod_{1 \leq i \leq k} \left[ \prod_{j \in D_i} \left( 1 - \alpha_i^{\exp(\hat{\beta}x_j)} \right) \right] \left[ \prod_{j \in R(t_{(i)})-D_i} \alpha_i^{\exp(\hat{\beta}x_j)} \right]$$

Taking the estimated regression coefficients from the partial likelihood and differentiating the logarithm of the last equation with respect to  $\alpha_i$  gives the maximum likelihood estimate of  $\alpha_i$ . The maximum likelihood estimate of the contribution  $\hat{h}_0(t_{(i)})dt_{(i)} = 1 - \hat{\alpha}_i$  to the hazard at  $t = t_{(i)}$  is given by

$$\sum_{m \in D_j} \frac{\exp(\hat{\beta}x_m)}{(1 - \alpha_i^{\exp(\hat{\beta}x_m^*)})} = \sum_{k \in R(t_j)} \exp(\hat{\beta}x_k) \tag{15}$$

If there are no ties, the solution  $\hat{\alpha}_i$ , an estimate of  $\alpha_i$  is

$$\hat{\alpha}_i = \left[ 1 - \frac{\exp(\hat{\beta}x_i)}{\sum_{k \in R(t_i)} \exp(\hat{\beta}x_k)} \right]^{\exp(-\hat{\beta}x_i)} \tag{16}$$

Accordingly, the Kalbfleisch-Prentice estimated survival function for a subject with covariates is:

$$\hat{S}_{KP}(t | X = x^*) = \left[ \prod_{i:t_i \leq t} \hat{\alpha}_i^{\delta_i} \right]^{\exp(\hat{\beta}x^*)} \tag{17}$$

Equation (17) is a step function estimate of the survivor function and in many instances a continuous estimate would be preferable, especially for suggesting a parametric form for  $h_0(t)$ . Kalbfleisch and Prentice focused on the probability of survival,  $\alpha_0 = 1 - h_0$ , rather than on that of failure,  $h_0$ . For these reason, in this paper we proposed a kernel smoothing technique for baseline estimator, based on Kalbfleisch-Prentice method. We start with kernel smoothing of baseline hazard function, based on Kalbfleisch-Prentice estimator and epanechnikov kernel, than we use it to calculate the cumulative baseline hazard function.

$$\hat{h}_{0,KPK}(t) = \frac{1}{b} \sum_{i=1}^n K\left(\frac{t-T_i}{b}\right) \left( 1 - \left( 1 - \frac{\exp(\hat{\beta}x_i)}{\sum_{k \in R(t_i)} \exp(\hat{\beta}x_k)} \right)^{\exp(-\hat{\beta}x_i)} \right) \quad (18)$$

For the kernel we have used the Epanechnikov kernel proposed by (Muller and Wang, 1994) with boundary correction, given by

$$K(t) = \frac{3}{4}(1-t^2), \quad |t| \leq 1 \quad (19)$$

Then, the smoothing baseline survival function estimator can be taken by substituting in equation (10) the cumulative baseline hazard function taken from the baseline hazard function of equation (18). For randomly censored survival data, Marron and Padgett (1987) have shown that the cross-validation method gives the optimal bandwidth for estimating the density: the ratio between the integrated squared error for the cross-validation bandwidth and the infimum of the integrated squared error for any bandwidth almost surely converges to 1. The bandwidth selected by cross-validation is defined by:

$$\hat{b}^{\hat{\beta}} = \arg \min_b \left\{ E \int_0^{\tau} \{ \hat{\alpha}_b^{\hat{\beta}}(t) \}^2 dt - 2 \sum_{i \neq j} \frac{1}{b} K\left(\frac{T_i - T_j}{b}\right) \frac{\delta_i}{\bar{Y}(T_i)} \frac{\delta_j}{\bar{Y}(T_j)} \right\} \quad (20)$$

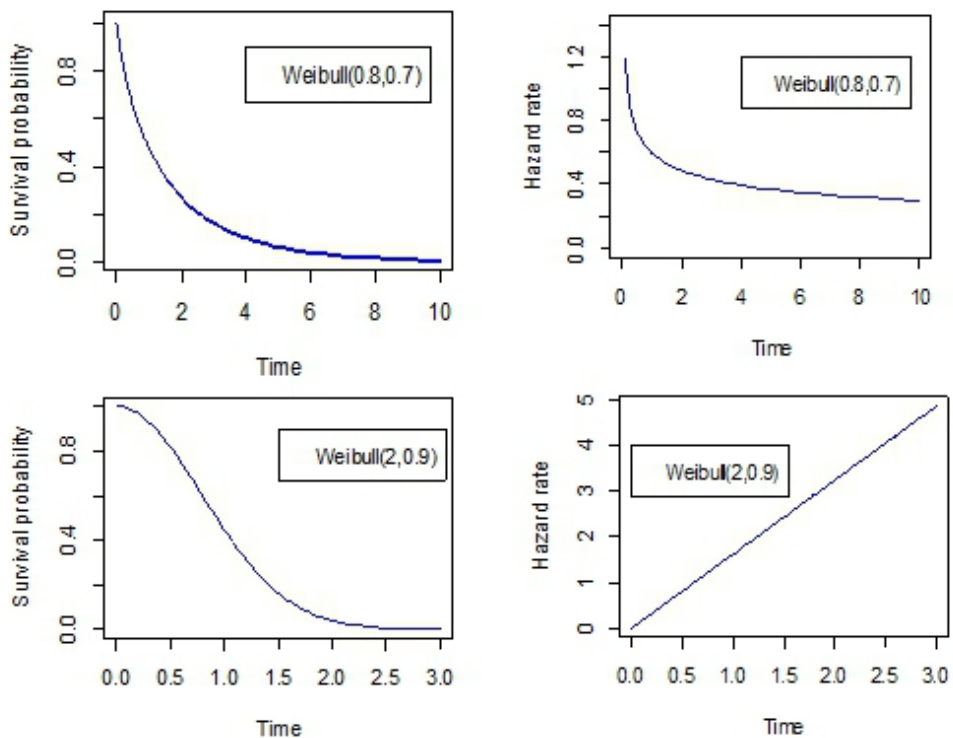
where  $\bar{Y} = \sum_{i=1}^n I_{\{T_i \geq t\}}$ .

### Simulation Results

The aim of this section is to illustrate the behavior of the kernel estimator of the baseline function in the case of right censoring, based on Kalbfleisch-Prentice estimator (KPK)(the proposed method) and to compare it with the smoothing of the Breslow estimator (BRK), with a bandwidth selected by cross-validation introduced by Ramlau-Hansen (1983b). We compare the performance of the proposed method with that of the smoothing of the Breslow estimator. The comparison is based on Bias. Xia et al., (2018) compare the performance of Kalbfleisch-Prentice and Breslow estimators regarding bias, mean squared error and relative mean squared error. In most situations in their study, the Kalbfleisch Prentice estimator results in less bias and smaller mean squared error than the Breslow estimator. Their differences are especially clear at the tail of the distribution.

Some specified models are used to see the finite sample performance. We present two scenarios under which to evaluate the relative performance of the smoothing of Breslow estimator and the kernel smoothing of KP estimator, considering two Weibull distributions. For the Weibull distributions, we include both cases of increasing and decreasing hazard functions. The survival and hazard functions of two scenarios are presented in Figure 1.



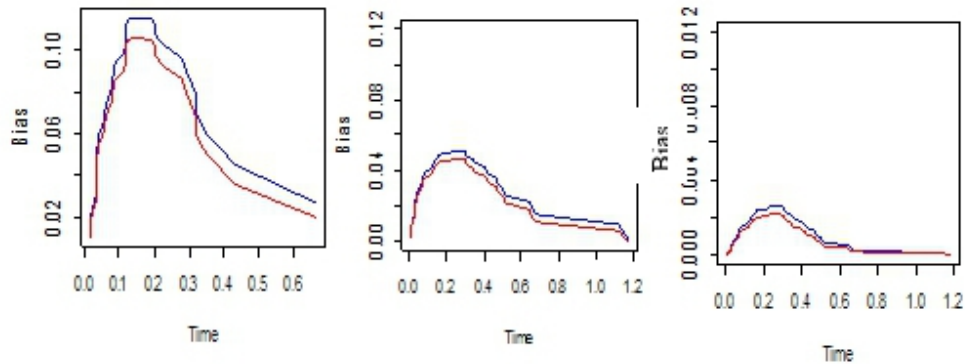


**Figure 1.** Weibull survival probability and hazard rate with shape parameters  $a = 0.8; 2$  and scale parameter  $b = 0.7; 0.9$ .

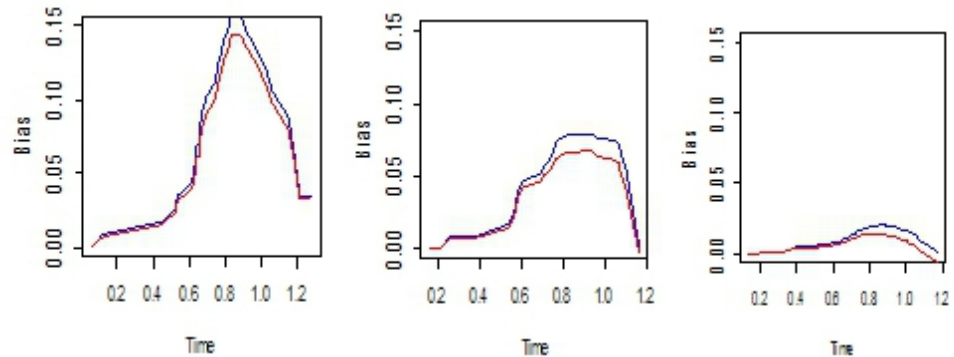
The survival time is generated using the Cox proportional model with one covariate  $x$ ,  $x \sim Unif(0, 1)$ , and for the regression coefficient  $\beta$  we set to be 0.6. The censoring time  $C_i$  for  $i = 1, \dots, n$  are simulated independently from the survival times via an exponential distribution. The distribution of censoring time  $C$  is deliberately calibrated to obtain the desired censoring rate  $r$ . The simulation settings used to study the effect of sample size and the censoring rate. The scenarios are: for each example baseline survival distribution, we have taken three different sample sizes  $n = 30, 60, 90$ , and two different censoring rate  $r = 10\%$  and  $r = 40\%$ . We assume that  $x^* = 0.9$ . The associated baseline function has the form  $h_0(t) = ab^a t^{a-1}$ , where  $a$  and  $b$  stand for the parameters in  $W(a, b)$ , Weibull distribution. The estimators of the baseline hazard function are both constructed with the Epanechnikov kernel. We compare the performance according to the bias. Let  $S(t | X = x^*)$  denote the true survival function for a subject with covariate vector  $X = x^*$  and  $N$  denote the number of simulations, which is 1000. The bias is calculated as follows:

$$Bias = \frac{1}{N} \sum_{i=1}^N \hat{S}_i(t | X = x^*) - S(t | X = x^*) \quad (21)$$

The two cases considered for Weibull distribution are with shape parameter  $a=2$  and scale parameter  $b=0.9$ ; and with shape parameter  $a=0.8$  and scale parameter  $b=0.7$ . The Weibull distribution with  $a = 2$  corresponds to an increasing hazard, and that with  $a = 0.8$  corresponds to a decreasing hazard. To this end, we run 1000 simulations and for every generated data set we calculate the estimators at the point  $t = 0.8$  for Weibull (2, 0.9) and  $t = 2$  for Weibull (0.8, 0.7). We first study the case for Weibull(2, 0.9) and the bias plots are presented in Figure 2 and Figure 3.



**Figure 2.** Bias for survival estimates of a Weibull baseline survival distribution with  $a = 2$ ,  $b = 0.9$ ,  $r=10\%$  and  $x^* = 0.9$ , the kernel smoothing of KP estimator (red curve) and the smoothing of Breslow estimator (blue curve)

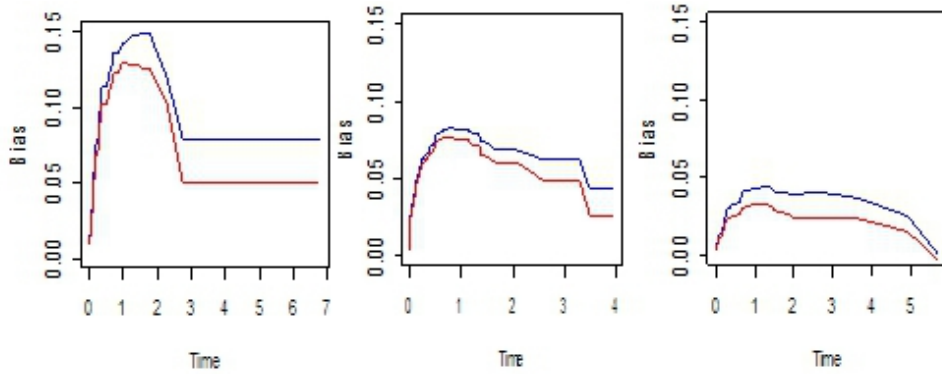


**Figure 3.** Bias for survival estimates of a Weibull baseline survival distribution with  $a = 2$ ,  $b = 0.9$ ,  $r=40\%$  and  $x^* = 0.9$ , the kernel smoothing of KP estimator (red curve) and the smoothing of Breslow estimator (blue curve).

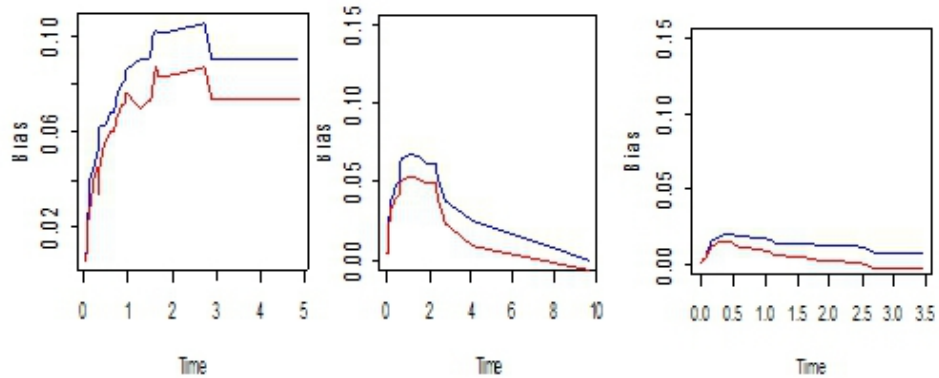
As expected, with both procedures, the Bias decreases when the sample size increases. Also the Bias increases when the censoring rate increases, and the sample size remain constant. Generally, the magnitude of the BRK estimator is larger than that of the KPK estimator. For all scenarios where the censoring rate is 10%, the bias reaches a plateau after  $t = 0.6$ . The

BRK estimator always overestimates the survival probabilities, particularly for small sample sizes.

The bias plots for Weibull(0.8, 0.7) are presented in Figure 4 and Figure 5.



**Figure 4.** Bias for survival estimates of a Weibull baseline survival distribution with  $a = 0.8$ ,  $b = 0.7$ ,  $r=10\%$  and  $x^* = 0.9$ , the kernel smoothing of KP estimator (red curve) and the smoothing of Breslow estimator (blue curve)



**Figure 5.** Bias for survival estimates of a Weibull baseline survival distribution with  $a = 0.8$ ,  $b = 0.7$ ,  $r=40\%$  and  $x^* = 0.9$ , the kernel smoothing of KP estimator (red curve) and the smoothing of Breslow estimator (blue curve).

In general, bias decreases as  $n$  increase, or  $r$  decreases. The magnitude of the bias for the BRK estimator is almost always larger than that for the KPK estimator. For all scenarios under Weibull (0.8, 0.7) the bias reaches a plateau after  $t = 3$ .

**Table 1:** Bias for BRK and KPK estimators, for Weibull (2, 0.9) and Weibull (0.8, 0.7) densities, with two censoring rates  $r = (10\%, 40\%)$  and three sample size  $n=30,60,90$ .

		Weibull(2,0.9) $t=0.8$		Weibull(0.8,0.7) $t=2$	
		BIAS BRK	for KPK	BIAS BRK	for KPK
<b>n=30</b>	<b>r=10%</b>	0.03	0.02	0.12	0.1
	<b>r=40%</b>	0.15	0.13	0.12	0.08
<b>n=60</b>	<b>r=10%</b>	0.018	0.014	0.07	0.054
	<b>r=40%</b>	0.07	0.06	0.074	0.059
<b>n=90</b>	<b>r=10%</b>	0.0009	0.001	0.047	0.032
	<b>r=40%</b>	0.02	0.012	0.028	0.019

Table 1 shows the results for the bias, with the two methods. The results are obtained with 1000 replications. Firstly the results of the table demonstrate that the bias for KPK estimator is significantly reduced with the increasing of the sample size. Secondly, when the degree of censoring increases, KPK estimator remains significantly better compared with the BRK estimator, even though the Bias increases as expected. We calculate the estimators at the point  $t = 0.8$  for Weibull (2, 0.9) and  $t = 2$  for Weibull (0.8, 0.7).

## Conclusion

In this paper, we developed a kernel smoothing estimator for baseline survival function, conditional on covariates, based on Kalbfleisch and Prentice method. This estimator is obtained by starting with a kernel smoothing of baseline hazard function, based on Kalbfleisch-Prentice estimator and epanechnikov kernel, we use it to calculate the cumulative baseline hazard function and then to estimate the baseline survival function.

The simulation results confirm the bias reduction property compared to that of kernel smoothing of Breslow estimator. The scenarios were made using two weibull distribution, Weibull (2, 0.9) and Weibull (0.8, 0.7), with two censoring rates  $r = (10\%, 40\%)$  and three sample size  $n=30, 60, 90$ . In general, bias decreases as  $n$  increase, or  $r$  decreases. The proposed estimator remains significantly better compared with the smoothing of Breslow estimator.

## References:

1. Breslow N., (1972) Discussion of Professor Cox's paper. Journal of the Royal Statistical Society, Series B 34: 216-217.
2. Breslow N. E., and Wellner J. A., (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scand J Statist* 34: 86–102

3. Choi S., Huang X., Chen Y. H., (2014). A class of semiparametric transformation models for survival data with a cured proportion. *Lifetime Data Analysis* 20: 369–386.
4. Cox D., (1972) Regression models and life-tables. *Journal of the Royal Statistical Society, Series B* 34: 187-220.
5. Cox D., (1975) Partial likelihood. *Biometrika* 62: 269-276.
6. Efron B., (1977). The efficiency of Cox's likelihood function for censored data. *J. Am. Statist. Assoc.* 72, 557-565.
7. Guilloux A., Lemler S., and Taupin M. (2016) Adaptive kernel estimation of the baseline function in the Cox model with high-dimensional covariates. HAL Id: hal-01171775.
8. Huang X., and Strawderman R. L. (2006). A note on the Breslow survival estimator. *Journal of Nonparametric Statistics* 18: 45–56.
9. Kalbfleisch J., and Prentice R., (1973) Marginal likelihoods based on Cox's regression and life model. *Biometrika* 60: 267-278.
10. Kalbfleisch J., and Prentice R., (1980) *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York.
11. Kaplan E., and Meier P., (1958) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, vol. 53, pp. 457–481.
12. Kleinbaum D., and Klein M., (2005) *Survival Analysis. Statistics for Biology and Health*, Springer, New York, NY, USA, 2nd edition.
13. Lin D.Y. (2007) *Lifetime Data Anal* 13: 471. <https://doi.org/10.1007/s10985-007-9048-y>
14. Marron, J., and Padgett, W., (1987) Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples. *The Annals of Statistics* 15, 1520–1535.
15. Mao M., and Wang J.L. (2010). Semiparametric efficient estimation for a class of generalized proportional odds cure models. *Journal of the American Statistical Association* 105: 302–311.
16. Muller H., and Wang J., (1994) Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, Vol. 50, No. 1, pp.61–76.
17. Oakes D., (1972) Comment on D. R. Cox (1972) paper. *J. R. Statist. Soc. B*, 34, 208
18. Oakes D., (1977) The asymptotic information in censored survival data. *Biometrika* 64, 441-8.
19. Peto R., and Peto J., (1972). Asymptotically efficient rank invariant procedures (with discussion). *J. R. Statist. Soc. A* 135, 185-206.
20. Ramlau-Hansen H., (1981) Udglætning med kernefunktioner i forbindelse med tælleprocesser: Del 1. Forsikringsmatematisk Laboratorium, Københavns universitet.

21. Ramlau-Hansen, H. (1983a) Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics* 11, 2, 453–466.
22. Ramlau-Hansen H., (1983b) The choice of a kernel function in the graduation of counting process intensities. *Scandinavian Actuarial Journal*, 165–182.
23. Selingerová I., Doleželová H., Horová I., Katina S., Zelinka J., (2016). Survival of Patients with Primary Brain Tumors: Comparison of Two Statistical Approaches. *PLOS ONE*. doi:10.1371/journal.pone.0148733
24. Tsiatis A., (1981). A large sample study of Cox's regression model. *Ann. Statist.* 9, 109-21.
25. Weng Y., (2007). Baseline survival function estimators under proportional hazards assumption. Institute of Statistics, National University of Kaohsiung, Kaohsiung, Taiwan, 811 R.O.C.
26. Xia F., Ning J., and Huang X., (2018) Empirical Comparison of the Breslow Estimator and the Kalbfleisch Prentice Estimator for Survival Functions. *Journal of Biometrics and Biostatistics* 9: 392. doi: 10.4172/2155-6180.1000392