Journal of the Korean
Data & Information Science Society
2001, Vol. 12, No.2, pp. 125 ~ 143

Kernel Estimation of Hazard Ratio Based on Censored Data

Myong-Hui Choi[1] · In-Suk Lee · Jae-Kee Song[2]

Abstract

We, in this paper, propose a kernel estimator of hazard ratio with censored survival data. The uniform consistency and asymptotic normality of the proposed estimator are proved by using counting process approach. In order to assess the performance of the proposed estimator, we compare the kernel estimator with Cox estimator and the generalized rank estimators of hazard ratio in terms of MSE by Monte Carlo simulation. Two examples are illustrated for our results.

Key Words and Phrases: Proportional hazard model, Hazard ratio, Censored data, Kernel method.

1. INTRODUCTION

Being compared survival across treatment groups in a clinical trial, it is useful to have a descriptive measure of the difference in survival between groups. When the hazard functions in two groups are roughly proportional, the ratio of hazard functions has the interpretation of relative risk.

The Cox's proportional hazard model(here after will be abbreviated by PHM) (Cox, 1972) assumes that the hazard function for the survival time t of an individual with covariate vector z has the form

$$\alpha(t|z) = \alpha_0(t)\exp({\beta_0}'z), \quad t \geq 0,$$

where β_0 is a p-vector of unknown regression coefficients and $\alpha_0(t)$ is an unknown and unspecified baseline hazard function. In case that $p = 1$ and covariate z is an indicator for treatment group, the PHM becomes

$$\alpha_2(t) = \theta\alpha_1(t),$$

[1]Department of Information Communication, Daegu Polytechnic College, Daegu, 706-711, Korea.
[2]Department of Statistics, Kyungpook National University, Daegu, 702-701, Korea.

where the proportionality constant $\theta(=e^{\beta_0})$ is the relative risk, and $\alpha_1(t)$ and $\alpha_2(t)$ are the hazard functions in control and treatment group respectively. Cox(1972, 1975) use the partial likelihood function to estimate the regression coefficient β_0. (see Andersen and Gill, 1982).

Andersen (1983) introduces the generalized rank estimator of the hazard ratio as a new interpretation of the linear nonparametric two sample tests for censored data, and establishes asymptotic normality by using counting process and martingale theory. Dabrowska, Doksum and Song (1989) provide graphs, confidence procedures and tests that could be used with censored survival data to compare the hazard experience of a treatment group with that of a control group. In particular, they considered the relative change $\Delta(t)$ in a cumulative hazard function, which equals to $\theta - 1$ under PHM.

In this paper, we extend Dabrowska et al. (1989)'s estimator by smoothing via kernel method. A kernel estimator of hazard ratio is proposed and the asymptotic properties of the proposed estimator are derived using counting processes approaches in Section 2. Section 3 presents the results of a simulation study, comparing this estimator with others. Finally we illustrate the estimators of hazard ratio using the Veteran's Administration lung cancer data and Stanford heart transplant data.

2. KERNEL ESTIMATION OF HAZARD RATIO

Nelson(1972) and Aalen(1978) suggested the estimator for cumulative hazard function, as follows,

$$\widehat{\beta}_i(t) = \int_0^t \frac{dN_i(s)}{Y_i(s)}, \quad i = 1, 2, \tag{2.1}$$

where $N_i(t)$ are the numbers of deaths in the group i in the interval $[0, t]$ and $Y_i(t)$ are the numbers of the individuals at risk at time $t-$ in the group i.

Dabrowska et al(1989) develop a simultaneous confidence band for hazard ratio by using the ratio of the Nelson-Aalen estimators and provide a graphical procedure to check whether the PHM holds or not. A smoothed version of their estimator may be considered, which we now propose a kernel estimator for hazard ratio as follows.

$$\widehat{\theta}_{KER}(t) = \frac{\int_0^1 K(\frac{t-s}{b})\, d\widehat{\beta}_2(s)}{\int_0^1 K(\frac{t-s}{b})\, d\widehat{\beta}_1(s)}, \tag{2.2}$$

where K is a bounded function with integral 1, and $\widehat{\beta}_i(t)$, $i = 1, 2$, is the Nelson-Aalen estimator of the i-th group cumulative hazard function $\int_0^t \alpha_i(s)ds$, and b is the positive number, where plays role of the amount of smoothing. The optimal selection of b is very crucial.

Remark 1. Practical application of kernel method such as density estimation is crucially dependent on the choice of the smoothing parameter or bandwidth b. Although effective data analysis has often been by a subjective, trial-and-error approach to this choice, the usefulness of kernel method would be greatly enhanced if an efficent and objective method of using the data to determine the amount of smoothing could be agreed upon. Hence various data-driven methods for choosing the bandwidth have been proposed and studied. For details, see Park and Marron(1990) and references therein.

Theorem 1. Assume that the following conditions hold :
(i) $\alpha_i(t),\ i=1,2$, are continuous on $[0,1]$.
(ii) There exist funtions y_1, y_2 taking values in $(0,1)$ such that under the PHM

$$\sup_{t\in[0,1]}\left|\frac{Y_i^{(n)}(t)}{n}-y_i(t)\right| \xrightarrow{p} 0,\quad i=1,2,\quad \text{as}\quad n\to\infty.$$

Then $\widehat{\theta}_{KER}^{(n)}(t)$ is uniformly consistent for θ under the PHM, where $Y_i^{(n)}(t)$ and $\widehat{\theta}_{KER}^{(n)}(t)$ were superscripted by (n) to represent the dependency of n.
(the proof is given in the Appendix).

Note that in the assumption (ii), the functions $y_i(t)$ play role of a kind of survival functions of control and treatment groups.

Theorem 2. Assume that the conditions of Theorem 1 are satisfied and that the following conditions hold :
(i) $b_n \in (0,1/2)$ and $b_n \to 0$ as $n\to\infty$.
(ii) The kernel has support within $[-1,1]$ and is symmetric about zero.
(iii) $\frac{1}{b_n}\int_0^1 K^{(n)}(\frac{t-s}{b_n})\alpha_1(s)ds \xrightarrow{p} \alpha_1(t),\quad t\in[b_n, 1-b_n],\quad \text{as}\quad n\to\infty.$

Then the process $\sqrt{nb_n}(\widehat{\theta}_{KER}^{(n)}(t)-\theta)$ converges weakly to a Gaussian process mean zero with variance function $\sigma^2_{KER}(t)$ given by

$$\sigma^2_{KER}(t) \;=\; \frac{\theta^2}{\alpha_1(t)}\left(\frac{y_1(t)+\theta y_2(t)}{\theta y_1(t)y_2(t)}\right)\int_{-1}^{1}K^2(u)\,du. \tag{2.3}$$

(the proof is given in the Appendix).

Corollary 1. Assume that the condition (ii) of Theorem 1 and the condition (iii) of Theorem 2 are satisfied. Then $\widehat{\sigma}^2_{KER}(t)$ is uniformly consistent for $\sigma^2_{KER}(t)$ under the PHM, which is given by

$$\widehat{\sigma}^2_{KER}(t) \;=\; n(\widehat{\theta}_{KER}^{(n)}(t))^2 \times \frac{\int_{-1}^{1}K^2(u)\left(\dfrac{d[N_1(t-bu)+N_2(t-bu)]}{Y_1(t-bu)\widehat{\theta}Y_2(t-bu)}\right)}{\left(\dfrac{\int_{-1}^{1}K(u)dN_1(t-bu)}{Y_1(t-bu)}\right)^2}. \tag{2.4}$$

(the proof is given in the Appendix).

3. SIMULATION STUDY

In this section we compare the performance of the kernel estimator with others such as Cox and the generalized rank estimators in terms of MSE by Monte Carlo simulation. Crowley, Liu and Voelkel(1982) compare asymptotic variances for the maximum likelihood estimator, the Cox estimator and the generalized rank estimator of relative risk as a large sample measure of efficiency.

The simulation structure adopted here has $n = n_1 + n_2$, which is composed of the sample sizes of control group and treatment group respectively. Lifetimes of control and treatment groups are generated from an exponential distridution ($Exp(\lambda)$) with mean $\frac{1}{\lambda}$ and a Weibull distribution ($Weib(\lambda, \delta)$) with parameters δ and λ whose survival function is $exp(-\lambda t^{\delta})$. Both censoring times are generated from exponential distributions with parameters being calculated to achieve given fixed censoring rates. The values of λ are determined by censoring rates which are ranged from 10% to 30%.

To compare the performances of the kernel estimator in the case of PHM, we take $Exp(\lambda_{11})$ and $Exp(\lambda_{21})$ as the lifetime distributions of control and treatment groups, and $Exp(\lambda_{12})$ and $Exp(\lambda_{22})$ as their corresponding censoring distributions. Here λ_{12} and λ_{22} are determined so that censoring rates are ranged from 10% to 30%. The true hazard ratio $\theta(t)$ is $\frac{\lambda_{21}}{\lambda_{11}}$. Similarly for the case that the PHM does not hold (non–PMH), we take $Weib(\lambda_{11}, \delta_1)$ and $Weib(\lambda_{21}, \delta_2)$ as the lifetime distributions of control and treatment groups, and $Exp(\lambda_{12})$ and $Exp(\lambda_{22})$ as their corresponding censoring distributions. Also we choose λ_{12} and λ_{22} in the same way as in the PHM case. The true hazard ratio $\theta(t)$ is $\frac{\lambda_{21}}{\lambda_{11}}\frac{\delta_2}{\delta_1}t^{\delta_2-\delta_1}$.

For both cases, the values of estimates of hazard ratio and their MSE's are tabulated in Table 1 and 2. For each simulation run, there are 200 replications and the values of t at which estimates are evaluated, are selected the $5^{th}, 10^{th}, \cdots, 90^{th}$ and 95^{th} quantile points of the control group lifetime distribution. The Cox's estimator $\widehat{\theta}_{COX}(= e_0^{\widehat{\beta}})$ is obtained iteratively by the Newton-Raphson algorithm with the Mantel-Haenszel estimator of β_0 as an initial value. The weight functions proposed by Mantel-Haenszel, Gehan and Harrington-Fleming are used to obtain the generalized rank estimators $\widehat{\theta}_{MH}(t)$, $\widehat{\theta}_{GH}(t)$, $\widehat{\theta}_{HF}(t)$ (For details, see Gill and Schumacher, 1987).

Finally the following kernel function is used (Epanechnikov, 1969)

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2) & \text{if } |x| \leq 1 \\ 0 & \text{otherwise,} \end{cases}$$

and at each point the bandwidth is selected so that the MSE is minimized among 0.05(0.05)0.5 times of range of the generated survival times of the control group.

The results are summarized in Table 1 and 2. Table 1 provide estimates of $\theta(t)$ and their MSE's for the two selected models which hold the PHM, and Table 2 for non-PHM. From Table 1 and 2, the following conclusions may be extracted.
(a) In cases of PHM, the proposed kernel estimator $\widehat{\theta}_{KER}(t)$ seems to be better than the generalized rank estimators $\widehat{\theta}_{MH}(t)$, $\widehat{\theta}_{HF}(t)$ and $\widehat{\theta}_{GH}(t)$, and compatible with Cox estimator, $\widehat{\theta}_{COX}$ in the sense of MSE.
(b) In cases of non-PHM, $\widehat{\theta}_{KER}(t)$ works best among others.
(c) Although we are not tabulated the results due to limited space, $\widehat{\theta}_{KER}(t)$ seems to have smaller biases than other estimators in any case.

In most cases $\widehat{\theta}_{KER}(t)$ seems to have on the whole smaller biases, MSE's and their standard deviations of MSE's than $\widehat{\theta}_{MH}(t)$, $\widehat{\theta}_{HF}(t)$, and $\widehat{\theta}_{GH}(t)$. Even, in case of PHM, $\widehat{\theta}_{KER}(t)$ seems to be as good as Cox estimator which is known to be best asymptotically.

Table 1. Estimates of $\theta(t)$ and their MSE's under the PHM

(1) $\lambda_{11} = 1.0, \ \lambda_{12} = 0.111, \ \lambda_{21} = 1.5, \ \lambda_{22} = 0.167$

censoring rate = 10 %

n	t	0.288	0.357	0.431	0.598	0.693	0.799	0.916	1.050	1.204
	KER	1.537	1.558	1.536	1.532	1.499	1.496	1.448	1.508	1.382
n_1	COX	1.539	1.539	1.539	1.539	1.539	1.539	1.539	1.539	1.539
	GGH	1.774	1.709	1.650	1.658	1.618	1.609	1.608	1.626	1.612
30	GMH	1.769	1.697	1.651	1.640	1.607	1.593	1.599	1.609	1.577
	GHF	1.770	1.702	1.649	1.647	1.610	1.598	1.594	1.615	1.591
	MSE1	0.183	0.254	0.273	0.204	0.198	0.188	0.162	0.197	0.185
n_2	MSE2	0.188	0.188	0.188	0.188	0.188	0.188	0.188	0.188	0.188
	MSE3	1.091	0.965	0.563	0.663	0.412	0.427	0.364	0.409	0.334
30	MSE4	1.031	0.884	0.561	0.576	0.348	0.361	0.288	0.367	0.267
	MSE5	1.502	0.914	0.555	0.610	0.369	0.382	0.314	0.378	0.287
	KER	1.533	1.518	1.530	1.503	1.501	1.472	1.501	1.510	1.461
n_1	COX	1.533	1.533	1.533	1.533	1.533	1.533	1.533	1.533	1.533
	GGH	1.634	1.659	1.622	1.593	1.554	1.573	1.557	1.570	1.542
50	GMH	1.631	1.659	1.617	1.589	1.548	1.566	1.557	1.563	1.543
	GHF	1.632	1.658	1.619	1.590	1.550	1.568	1.556	1.565	1.541
	MSE1	0.121	0.099	0.126	0.116	0.117	0.115	0.133	0.152	0.146
n_2	MSE2	0.112	0.112	0.112	0.112	0.112	0.112	0.112	0.112	0.112
	MSE3	0.513	0.460	0.370	0.290	0.219	0.220	0.176	0.216	0.187
50	MSE4	0.497	0.462	0.347	0.259	0.194	0.201	0.160	0.179	0.144
	MSE5	0.502	0.458	0.355	0.270	0.201	0.205	0.163	0.189	0.157

Table 1. (continued)

(2) $\lambda_{11} = 1.5$, $\lambda_{12} = 0.167$, $\lambda_{21} = 1.0$, $\lambda_{22} = 0.111$

censoring rate = 10%

n	t	0.149	0.192	0.287	0.399	0.462	0.532	0.611	0.7	0.803
	KER	0.688	0.668	0.712	0.668	0.680	0.704	0.679	0.677	0.724
n_1	COX	0.704	0.704	0.704	0.704	0.704	0.704	0.704	0.704	0.704
	GGH	0.842	0.732	0.712	0.726	0.706	0.673	0.716	0.699	0.704
30	GMH	0.837	0.737	0.715	0.721	0.697	0.671	0.709	0.694	0.702
	GHF	0.839	0.734	0.713	0.723	0.701	0.672	0.712	0.696	0.702
	MSE1	0.038	0.051	0.047	0.041	0.047	0.048	0.047	0.054	0.058
n_2	MSE2	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
	MSE3	0.674	0.225	0.146	0.127	0.114	0.082	0.112	0.076	0.053
30	MSE4	0.653	0.232	0.142	0.111	0.109	0.082	0.098	0.068	0.049
	MSE5	0.663	0.228	0.143	0.117	0.110	0.081	0.103	0.070	0.050
	KER	0.678	0.671	0.695	0.684	0.690	0.699	0.694	0.681	0.672
n_1	COX	0.687	0.687	0.687	0.687	0.687	0.687	0.687	0.687	0.687
	GGH	0.747	0.766	0.746	0.712	0.712	0.671	0.706	0.690	0.688
50	GMH	0.747	0.768	0.745	0.711	0.710	0.670	0.701	0.686	0.688
	GHF	0.747	0.767	0.745	0.711	0.711	0.670	0.703	0.688	0.687
	MSE1	0.026	0.026	0.025	0.026	0.026	0.030	0.031	0.028	0.031
n_2	MSE2	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
	MSE3	0.262	0.208	0.117	0.054	0.068	0.052	0.051	0.057	0.040
50	MSE4	0.261	0.205	0.116	0.054	0.065	0.050	0.047	0.053	0.033
	MSE5	0.261	0.206	0.116	0.054	0.066	0.050	0.048	0.054	0.035

Table 2. Estimates of $\theta(t)$ and their MSE's under the non-PHM

(1) λ_{11}=1.0, δ_1=0.5, λ_{12}=0.067, λ_{21} = 1.15, δ_2=2.0, λ_{22} = 0.49

censoring rate = 30%

n	t	0.127	0.186	0.261	0.481	0.638	0.839	1.102	1.450
	$\theta(t)$	0.209	0.368	0.613	1.532	2.342	3.539	5.322	8.028
	KER	0.266	0.373	0.589	1.503	2.082	3.207	4.411	5.916
n_1	COX	0.989	1.009	0.988	0.891	0.930	0.902	0.928	0.970
	GGH	0.048	0.076	0.147	0.270	0.347	0.465	0.540	0.587
30	GMH	0.051	0.083	0.166	0.327	0.453	0.660	0.839	1.001
	GHF	0.050	0.080	0.159	0.304	0.409	0.576	0.702	0.799
	MSE1	0.031	0.030	0.060	0.247	0.610	1.503	4.567	16.734
n_2	MSE2	0.723	0.525	0.235	0.455	2.051	7.025	19.372	49.866
	MSE3	0.030	0.094	0.235	1.613	4.008	9.490	22.925	55.415
30	MSE4	0.029	0.092	0.222	1.479	3.613	8.358	20.198	49.454
	MSE5	0.030	0.093	0.227	1.532	3.771	8.837	21.422	52.319
	KER	0.247	0.370	0.645	1.494	2.085	3.160	4.607	5.958
n_1	COX	1.109	1.025	1.021	1.033	1.078	0.994	1.042	1.045
	GGH	0.047	0.081	0.127	0.261	0.364	0.451	0.539	0.575
50	GMH	0.051	0.088	0.144	0.319	0.474	0.644	0.851	0.988
	GHF	0.049	0.085	0.137	0.296	0.430	0.562	0.709	0.792
	MSE1	0.010	0.019	0.035	0.143	0.327	0.957	3.106	9.833
n_2	MSE2	0.941	0.475	0.209	0.278	1.697	6.506	18.367	48.778
	MSE3	0.029	0.086	0.242	1.629	3.929	9.560	22.912	55.570
50	MSE4	0.028	0.083	0.228	1.490	3.516	8.425	20.057	49.605
	MSE5	0.028	0.084	0.233	1.544	3.680	8.896	21.328	52.401

Table 2. (continued)

(2) $\lambda_{11} = 1.15,\ \delta_1=2.0,\ \lambda_{12}=0.49,\ \lambda_{21} = 1.0,\ \delta_2=0.5,\ \lambda_{22} = 0.067$

censoring rate = 30%

n	t	0.612	0.721	0.833	0.893	0.955	1.023	1.284	1.415
	$\theta(t)$	0.454	0.355	0.286	0.258	0.233	0.210	0.149	0.129
	KER	0.551	0.395	0.322	0.296	0.289	0.258	0.214	0.217
n_1	COX	1.145	1.118	1.114	1.211	1.032	1.085	1.136	1.169
	GGH	3.594	2.889	2.446	2.404	2.251	2.246	1.954	2.009
30	GMH	2.723	2.137	1.703	1.643	1.478	1.435	1.141	1.144
	GHF	3.015	2.394	1.958	1.903	1.741	1.712	1.418	1.447
	MSE1	0.052	0.030	0.023	0.022	0.022	0.018	0.016	0.028
n_2	MSE2	0.574	0.664	0.768	1.028	0.729	0.841	1.043	2.220
	MSE3	13.540	8.475	5.577	5.430	2.647	5.079	3.905	4.114
30	MSE4	6.998	4.275	2.456	2.268	1.803	1.795	1.129	1.165
	MSE5	8.836	5.508	3.370	3.181	4.793	2.706	1.862	1.981
	KER	0.505	0.367	0.306	0.285	0.246	0.237	0.198	0.185
n_1	COX	1.026	0.975	1.002	1.041	1.029	0.998	0.970	0.985
	GGH	3.357	2.753	2:550	2.416	2.282	2.092	1.931	1.887
50	GMH	2.594	2.017	1.788	1.632	1.506	1.327	1.121	1.082
	GHF	2.853	2.262	2.042	1.895	1.769	1.582	1.396	1.355
	MSE1	0.028	0.016	0.012	0.0135	0.009	0.0089	0.009	0.0097
n_2	MSE2	0.357	0.413	0.554	0.644	0.686	0.695	0.697	0.766
	MSE3	10.445	6.548	5.844	5.171	4.636	3.963	3.529	3.426
50	MSE4	5.667	3.157	2.578	2.100	1.781	1.379	1.031	0.981
	MSE5	7.082	4.139	3.506	2.967	2.588	2.085	1.702	1.635

4. ILLUSTRATION

In this section the hazard ratio estimates $\widehat{\theta}_{KER}(t)$, $\widehat{\theta}_{COX}$, $\widehat{\theta}_{MH}(t)$, $\widehat{\theta}_{GH}(t)$ and $\widehat{\theta}_{HF}(t)$ are plotted for two real data which are available in Kalbfleisch and Prentice(1980). We use the Nelson-Aalen estimator of the cumulative hazard function whether each data set satisfies the PHM or not.

The data in the first example are the time to death of males with advanced inoperable lung cancer and randomized to either a standard or test chemotheraphy. The observations are shown in Table 3, where * stands for censored data. Only 9 of the 137 survival times are censored. The plot of cumulative hazard functions shows that the test group looks like to be more risky proportional to the standard group.

Figure 1 and 2 gives the Cox and generalized rank estimators,and kernel estimator of $\theta(t)$. The Cox estimator equals 1 and the kernel estimator and the generalized rank estimators wiggle rapidly for $t \leq 160$ but almost constant for $t > 160$. Therefore the five estimators of $\theta(t)$ show that the lung cancer data doesn't hold PHM generally, but locally holds PHM for $t > 160$.

Table 3. Veteran's Administration Lung Cancer Data

	72	411	228	126	118	10	82	110	314	100*
	42	8	114	25*	11	30	384	4	54	13
Standard	123*	97*	153	59	117	16	151	22	56	21
Group	18	139	20	31	52	287	18	51	122	27
	54	7	63	392	10	8	92	35	117	132
	12	162	3	95	177	162	216	553	278	12
	260	200	156	182*	143	105	103	250	100	
	999	112	87*	231*	242	991	111	1	587	389
	33	25	357	467	201	1	30	44	283	15
Test	25	103*	21	13	87	20	7	24	99	8
Group	99	61	25	95	80	51	29	24	18	83*
	31	51	90	52	73	8	36	4	8	7
	140	186	84	19	45	80	52	164	19	53
	15	43	340	133	111	231	378	49		

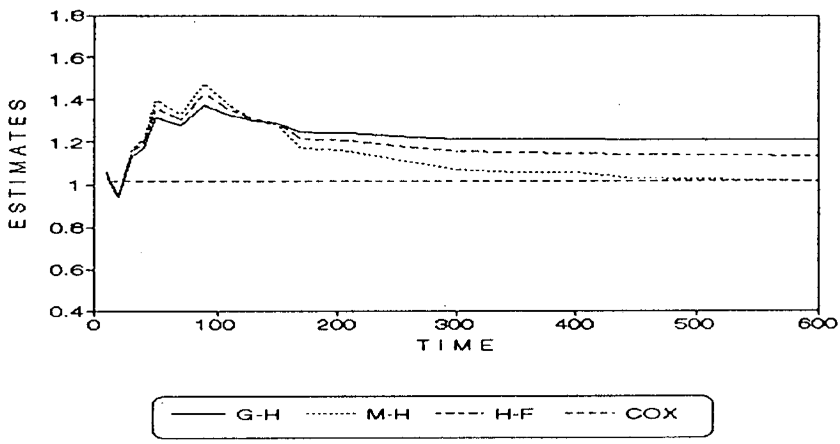

Figure 1. Cox and Generalized Rank Estimates of the Veteran's data

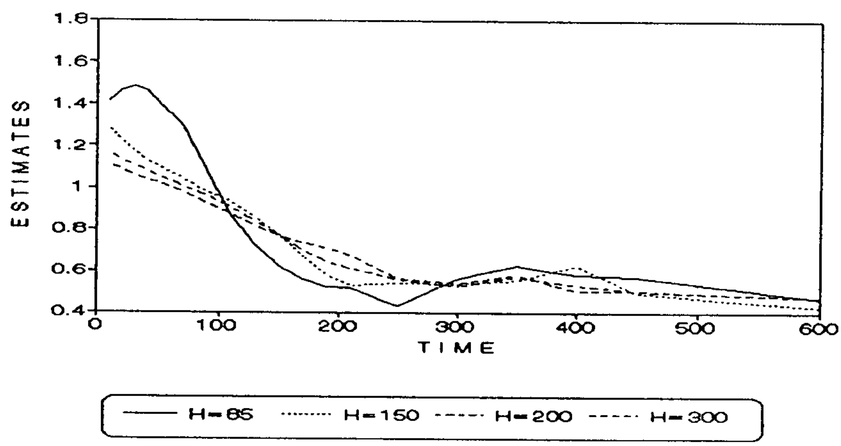

Figure 2. Kernel Estimates of the Veteran's data

In the second example we study the effect of the waiting time on the posttransplant survivals in the heart transplant data. Two groups in Table 4 are given the 69 transplanted patients with 45 recorded deaths and used two levels of waiting time. One group is the patients with waiting time up to 20 days and the other longer than 20. Figure 3 and 4 provides the generalized rank estimators and kernel estimator of $\theta(t)$ for the heart transplant data. The kernel estimator decreases and the generalized rank estimator increases slightly as time t increases, but both estimators almost constant. So the four estimators of $\theta(t)$ show that the heart transplant data holds PHM.

From Figures 1-4, the kernel estimator is smoothed by choosing the larger size of bandwidth and the generalized rank estimators seem to have the similar values.

Table 4. Stanford Heart Transplant Data

	5	16	16	17	28	30	43	45	51
Sample 1	53	58	61	66	68	68	77	78	81
	165	180*	397*	445*	596*	733	852	995	1032
	39	39*	72	72	80	90	96	100	109*
	110	131*	153	186	188	207	219	265*	285
Sample 2	285	308	334	340*	342	370*	482*	515*	545*
	583	620*	670*	675	841*	915*	941*	979	1141*
	1321*	1386	1407*	1571*	1586*	1799*			

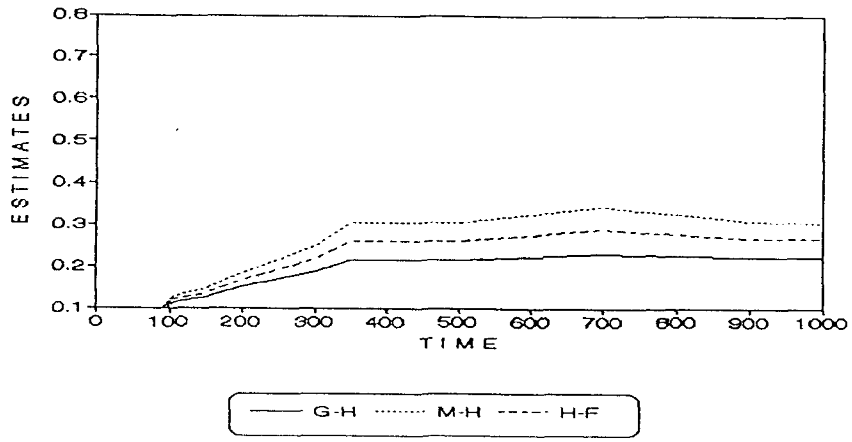

Figure 3. Cox and Generalized Rank Estimates of the Stanford data

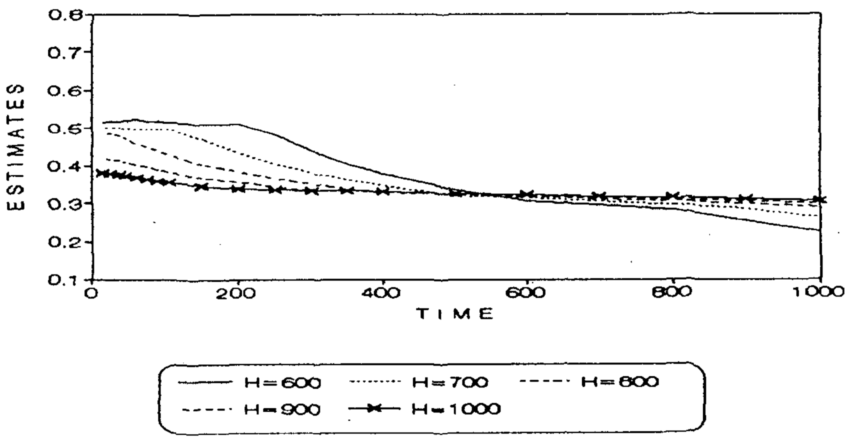

Figure 4. Kernel Estimates of the Stanford data

APPENDIX

A.1. Uniform consistency of $\widehat{\theta}_{KER}(t)$

Lemma 1. (Lenglart's inequality; Andersen & Gill, 1982) Let M be a local square integrable martingale. Then for all $\delta, \eta > 0$

$$P\left\{\sup_{t\in[0,1]} |M(t)| > \eta\right\} \le \frac{\delta}{\eta} + P\{< M, M > (1) > \delta\}.$$

Proof of Theorem 1. From the definition (2.2) of the kernel estimator $\widehat{\theta}^{(n)}_{KER}(t)$, we have, under the PHM

$$\begin{aligned}
\widehat{\theta}^{(n)}_{KER}(t) - \theta &= \frac{\int_0^1 K^{(n)}\left(\frac{t-s}{b_n}\right)\frac{dN_2^{(n)}(s)}{Y_2^{(n)}(s)}}{\int_0^1 K^{(n)}\left(\frac{t-s}{b_n}\right)\frac{dN_1^{(n)}(s)}{Y_1^{(n)}(s)}} - \theta \\
&= \frac{\int_0^1 K^{(n)}\left(\frac{t-s}{b_n}\right)\left(\frac{dM_2^{(n)}(s)}{Y_2^{(n)}(s)} - \theta\,\frac{dM_1^{(n)}(s)}{Y_1^{(n)}(s)}\right)}{\int_0^1 K^{(n)}\left(\frac{t-s}{b_n}\right)\frac{dN_1^{(n)}(s)}{Y_1^{(n)}(s)}},
\end{aligned}$$

where the last equality follows from the facts that $dN_i^{(n)}(s) = dM_i^{(n)}(s) + \alpha_i(s)Y_i^{(n)}(s)$ and $\alpha_2(s) = \theta\alpha_1(s)$ under the PHM. Here $K/Y_i, i = 1, 2$, is interpreted as 0 whenever $Y_i = 0, i = 1, 2$. Hence $\widehat{\theta}^{(n)}_{KER}(t) - \theta$ can be represented by a stochastic integral as follows :

$$\widehat{\theta}^{(n)}_{KER}(t) - \theta = \int_0^1 B_2^{(n)}(s)\, dM_2^{(n)}(s) - \int_0^1 B_1^{(n)}(s)\, dM_1^{(n)}(s),$$

where

$$B_1^{(n)}(s) = \left(\int_0^1 K^{(n)}\left(\frac{t-s}{b_n}\right)\frac{dN_1^{(n)}(s)}{Y_1^{(n)}(s)}\right)^{-1}\frac{K^{(n)}\left(\frac{t-s}{b_n}\right)}{Y_1^{(n)}(s)}$$

and

$$B_2^{(n)}(s) = \left(\int_0^1 K^{(n)}\left(\frac{t-s}{b_n}\right)\frac{dN_1^{(n)}(s)}{Y_1^{(n)}(s)}\right)^{-1}\frac{K^{(n)}\left(\frac{t-s}{b_n}\right)}{Y_2^{(n)}(s)}$$

are the stochastic integrals. Since $B_1^{(n)}$ and $B_2^{(n)}$ are the predictable processes and $M_1^{(n)}$ and $M_2^{(n)}$ are the martingales.

By using Lemma 1, we have, for all $\delta, \eta > 0$,

$$P\Big(\sup_{t\in[0,1]} |\widehat{\theta}^{(n)}_{KER}(t) - \theta| > \eta\Big) \le \frac{\delta}{\eta^2} + P\Big(< \widehat{\theta}^{(n)}_{KER} - \theta > (1) > \delta\Big)$$

$$= \frac{\delta}{\eta^2} + P\bigg(\Big(\int_0^1 [B_2^{(n)}(s)]^2 d < M_2^{(n)} > (s) + \int_0^1 [B_1^{(n)}(s)]^2 d < M_1^{(n)} > (s)\Big) > \delta\bigg).$$

By using $\frac{Y_i^{(n)}(t)}{n} \xrightarrow{p} y_i(t) > 0$, for each t, we complete the proof.

A.2. Asymptotic property of $\widehat{\theta}_{KER}(t)$

Lemma 2. (Martingale CLT; Andersen & Gill, 1982) Let $p \ge 1$ be fixed, and consider a sequence $N^{(n)}$ of k_n-variate counting processes with intensity processes $\Lambda^{(n)}$, and a sequence $H^{(n)}$ of $p \times k_n$-matrices of predictable processes, such that the stochastic integrals

$$U_j^{(n)}(t) = \int_0^t \sum_{h=1}^{k_n} H_{jh}^{(n)}(s)\{dN_h^{(n)}(s) - \Lambda_h^{(n)}(s)\, ds\}, \quad j = 1, \cdots, p,$$

are well defined. If, as $n \to \infty$,

$$< U_j^{(n)}, U_l^{(n)} > (t) \longrightarrow C_{jl}(t), \quad j, l = 1, \cdots, p, \quad t \in [0,1], \qquad (A.1)$$

where C is $p \times p$ matrix of continuous functions on [0,1] forming the covariance function of a p-variate Gaussian martingale $U^{(\infty)}$ with $U^{(\infty)}(0) = 0$, and if for all $\epsilon > 0$, as $n \to \infty$,

$$\int_0^1 \sum_{h=1}^{k_n} [H_{jh}^{(n)}(t)]^2 \Lambda_h^{(n)}(t) I\{|H_{jh}^{(n)}(t)| > \epsilon\}\, dt \xrightarrow{p} 0, \; j = 1, \cdots, p, \qquad (A.2)$$

then

$$U^{(n)} \xrightarrow{\mathcal{D}} U^{(\infty)}, \quad \text{as } n \to \infty, \quad \text{in } D([0,1]^p).$$

Proof of Theorem 2. Let $Z^{(n)}(t) \equiv \sqrt{nb_n}(\widehat{\theta}^{(n)}_{KER}(t) - \theta)$. Then, the process $Z^{(n)}(t)$ is simply

$$Z^{(n)}(t) = \frac{\sqrt{nb_n}\int_0^1 K^{(n)}\left(\frac{t-s}{b_n}\right)\left(\frac{dM_2^{(n)}(s)}{Y_2^{(n)}(s)} - \theta\frac{dM_1^{(n)}(s)}{Y_1^{(n)}(s)}\right)}{\int_0^1 K^{(n)}\left(\frac{t-s}{b_n}\right)\frac{dN_1^{(n)}(s)}{Y_1^{(n)}(s)}}$$

$$= \int_0^1 H_2^{(n)}(s)\, dM_2^{(n)}(s) - \int_0^1 H_1^{(n)}(s)\, dM_1^{(n)}(s)$$

where

$$H_i^{(n)}(s) = \sqrt{nb_n}B_i^{(n)}(s), \quad i = 1, 2,$$

and $B_1^{(n)}(s)$, $B_2^{(n)}(s)$ are defined in the proof of Theorem 1. Since $H_1^{(n)}(s)$ and $H_2^{(n)}(s)$ are the predictable processes and $M_1^{(n)}(s)$ and $M_2^{(n)}(s)$ are the martingales, $Z^{(n)}(t)$ is the stochastic integral with respect to $M_1^{(n)}(t)$ and $M_2^{(n)}(t)$.

Now, we need check two conditions (A.1) and (A.2) to apply Lemma 2. Since

$$\{|H_i^{(n)}(s)| > \epsilon\} = \left\{\left|\frac{1}{\sqrt{n}}\left(\int_0^1 K^{(n)}\left(\frac{t-s}{b_n}\right)\frac{dN_i^{(n)}(s)}{Y_i^{(n)}(s)}\right)^{-1}\frac{K^{(n)}(\frac{t-s}{b_n})}{Y_i^{(n)}(s)/n}\right| > \epsilon\right\}$$

and the condition (ii) of Theorem 1, we have

$$I\{|H_i^{(n)}(s)| > \epsilon\} \xrightarrow{p} 0 \text{ uniformly on } [0,1], \quad i = 1, 2.$$

That shows that the condition (A.2) of Lemma 2 is satisfied.

Next, from the definition of the variance process $< Z^{(n)} > (t)$, we have

$$\begin{aligned} < Z^{(n)} > (t) &= \frac{\theta^2}{\left(\int_{-1}^1 K^{(n)}(u)\frac{dN_1^{(n)}(t-b_n u)}{Y_1^{(n)}(t-b_n u)}\right)^2} \\ &\times \int_{-1}^1 K^{(n)}(u)^2\left(\frac{n}{Y_1^{(n)}(t-b_n u)} + \frac{n}{\theta Y_2^{(n)}(t-b_n u)}\right)\alpha_1(t-b_n u)\,du \\ &\xrightarrow{p} \frac{\theta^2}{\alpha_1(t)}\left(\frac{1}{\theta y_2(t)} + \frac{1}{y_1(t)}\right)\int_{-1}^1 K(u)^2 du, \end{aligned}$$

because of the conditions of Theorem 2 and (i), (ii) of Theorem 1. Hence, $< Z^{(n)} > (t) \xrightarrow{p} \sigma_{KER}^2(t)$. Therefore, by Lemma 2, $Z^{(n)}(t)$ converges in distribution to $N(0, \sigma_{KER}^2(t))$. So, we complete the proof.

A.3. Uniform consistancy of $\widehat{\sigma}_{KER}(t)$

Proof of Corollary 1. For convenience, denote $\widehat{\theta}_{KER}^{(n)}(t)$ by $\widehat{\theta}$. Furthermore, by Lemma 1, we have

$$\frac{1}{b_n}\int_0^1 K^{(n)}\left(\frac{t-s}{b_n}\right)\frac{dN_1^{(n)}(s)}{Y_1^{(n)}(s)} \xrightarrow{p} \alpha_1(t), \quad \text{as} \quad n \to \infty. \tag{A.3}$$

From the (2.3) and (2.4) , we have

$$\sup_{t\in[0,1]} |\widehat{\sigma}_{KER}^{(n)^2}(t) - \sigma_{KER}^2(t)|$$

$$\leq \sup_{t\in[0,1]} \left| \widehat{\theta}^2 \frac{\int_{-1}^{1} K^{(n)^2}(u) \left(\dfrac{n}{\widehat{\theta} Y_2^{(n)}(t-b_n u)} + \dfrac{n}{Y_1^{(n)}(t-b_n u)} \right) \alpha_1(t-b_n u) du}{\left(\int_{-1}^{1} K^{(n)}(u) \dfrac{dN_1^{(n)}(t-b_n u)}{Y_1^{(n)}(t-b_n u)} \right)^2} \right.$$

$$\left. - \frac{\widehat{\theta}^2 \int_{-1}^{1} K^{(n)^2}(u) \left(\dfrac{1}{\theta y_2(t)} + \dfrac{1}{y_1(t)} \right) \alpha_1(t)\, du}{\left(\int_{-1}^{1} K^{(n)}(u) \dfrac{dN_1^{(n)}(t-b_n u)}{Y_1^{(n)}(t-b_n u)} \right)^2} \right|$$

$$+ \sup_{t\in[0,1]} \left| \left[\left(\frac{\widehat{\theta}}{\int_{-1}^{1} K^{(n)}(u) \dfrac{dN_1^{(n)}(t-b_n u)}{Y_1^{(n)}(t-b_n u)}} \right)^2 - \left(\frac{\theta}{\alpha_1(t)} \right)^2 \right] \right.$$

$$\left. \times \left(\frac{1}{\theta y_2(t)} + \frac{1}{y_1(t)} \right) \alpha_1(t) \int_{-1}^{1} K^{(n)^2}(u) du \right|$$

$$= (I) + (II).$$

It is sufficient to show that $(I) \xrightarrow{p} 0$ and $(II) \xrightarrow{p} 0$. Using Theorem 1 and Theorem 2, we have

$$\int_{-1}^{1} K^{(n)^2}(u) \left(\frac{\alpha_1(t-b_n u)}{\theta Y_2^{(n)}(t-b_n u)/n} - \frac{\alpha_1(t)}{\theta y_2(t)} \right) du$$

$$+ \int_{-1}^{1} K^{(n)^2}(u) \left(\frac{\alpha_1(t-b_n u)}{Y_1^{(n)}(t-b_n u)/n} - \frac{\alpha_1(t)}{y_1(t)} \right) du \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Therefore $(I) \xrightarrow{p} 0$ is satisfied.
Also,

$$(II) = \sup_{t\in[0,1]} \left| \left(\frac{\widehat{\theta}}{\int_{-1}^{1} K^{(n)}(u) \dfrac{dN_1^{(n)}(t-b_n u)}{Y_1^{(n)}(t-b_n u)}} - \frac{\theta}{\alpha_1(t)} \right) \right.$$

$$\times \left(\frac{\widehat{\theta}}{\int_{-1}^{1} K^{(n)}(u) \dfrac{dN_1^{(n)}(t-b_n u)}{Y_1^{(n)}(t-b_n u)}} + \frac{\theta}{\alpha_1(t)} \right)$$

$$\left. \times \left(\frac{1}{\theta y_2(t)} + \frac{1}{y_1(t)} \right) \alpha_1(t) \int_{-1}^{1} K^{(n)^2}(u)\, du \right| .$$

By Theorem 1 and (A.3), we have

$$\frac{\widehat{\theta}}{\int_{-1}^{1} K^{(n)}(u)\dfrac{dN_1^{(n)}(t-b_nu)}{Y_1^{(n)}(t-b_nu)}} \xrightarrow{p} \frac{\theta}{\alpha_1(t)}.$$

Hence $(II) \xrightarrow{p} 0$ as $n \to \infty$. Therefore we complete the proof.

References

1. Aalen, O. O. (1978). Nonparametric Inference for a Family of Counting Processes, *The Annals of Statistics*, 6, 701-726.

2. Andersen, P. K. (1983). Comparing Survival Distributions via Hazard Ratio Estimates, *Scandinavian Journal of Statistics*, 10, 77-85.

3. Andersen, P. K. and Borgan, O. (1985). Counting Process Models for Life History Data, a Review, *Scandinavian Journal of Statistics*, 12, 97-158.

4. Andersen, P. K. and Gill, R. D. (1982). Cox's Regression Model for Counting Processes : A Large Sample Study, *The Annals of Statistics*, 10, 1100 - 1120.

5. Begun, J. M. and Reid, N. (1983). Estimating the Relative Risk with Censored data, *Journal of the American Statistical Association*, 78, 337-341.

6. Choi, M. H. (1992). Estimation of Hazard Ratio with Censored Survival Data , *Ph. D. Dissertation Kyungpook National University, Department of Statistics.*

7. Choi, M. H. (1994). Small Sample Study of Kernel Hazard Ratio Estimator, *Journal of Statistical Theory & Methods*, 5(2), 59-74.

8. Cox, D. R. (1972). Regression Models and Life Tables (with Discussion), *Journal of The Royal Statistical Society*, Ser. B, 34, 187-220.

9. Cox, D. R. (1975). Partial Likelihood, *Biometrika*, 62, 269-279.

10. Crowley, J., Liu, P.Y., and Voelkel, G. (1982). Estimation of the Ratio of Hazard Functions, *IMS Lecture Notes-Monograph Series*, 2, 46-73.

11. Dabrowska, D. M., Doksum, K. A. and Song, J. K. (1989). Graphical Comparison of Cumulative Hazard for two Populations, *Biometrika*, 76, 763-773.

12. Epanechnikov, V. A. (1969). Nonparametric estimation of multidimensional probability density, *Theory of Probability and its Applications*, 14, 153-158.

13. Fleming, T, R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*, Now York: John Wiley.

14. Gill, R. D. (1980). *Censoring and Stochastic Integrals*, Mathematical Centre Tracts 124, Mathematisch Centrum: Amsterdam.

15. Gill, R. D. and Schumacher, M. (1987). A Simple Test of the Proportional Hazards Assumption, *Biometrika*, 74, 289-300.

16. Harrington, D. P. and Fleming, T. R. (1982). A Class of Rank Test Procedures for Censored Survival Data, *Biometrika*, 69, 553-566.

17. Hastie, T. J. and Tibshirani, R. J. (1986). Generalized Additive Models (with Discussion), *Statistical Science*, 1, 297-318.

18. Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time data*, New York: John Wiley.

19. Kalbfleisch, J. D. and Prentice, R. L. (1981). Estimation of the Average Hazard Ratio, *Biometrika*, 68, 105 -12 .

20. Nelson, W. (1972). Theory and Applications of Hazard Plotting for Censored Failure Data, *Technometrics*, 14, 945-966.

21. O' Sullivan, F. (1986). Relative Risk Estimation, *Technical Report No. 76. University of California, Berkeley.*

22. Park, B. U. and Marron, J. S. (1990). Comparison of Data-Driven Bandwidth Selectors, *Journal of the American Statistical Association*, 85, 66-72.

23. Ramlau-Hansen, H. (1983). Smoothing Counting Process Intensities by means of Kernel Functions, *The Annals of Statistics*, 11, 453-466.

24. Silverman, B. W. (1985). *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

25. Zhang, B. (1996). Some Asymptotic Result for Kernel Density Estimation under Random Censorship, *Bernoull*, 2, 183-198.