Journal of the Korean Data & Information Science Society
2018, **29**(2), 513–521

http://dx.doi.org/10.7465/jkdi.2018.29.2.513
한국데이터정보과학회지

Comparison of Kendall's tau estimators for bivariate censored data†

Jeong Jae Park[1] · Hyemi Choi[2]

[1]Department of Statistics, Chonbuk National University
[2]Department of Statistics (Institute of Applied Statistics), Chonbuk National University

Received 11 January 2018, revised 26 February 2018, accepted 5 March 2018

Abstract

Kendall's τ is the important association measure and commonly used in biomedical application. Under censoring the estimation of τ has been challenging. Recently Hsieh (2010) replaced a censored event-time by a imputation with the conditional information and calculated the τ estimator based on the imputed data. We propose modified Hsieh's approach of the Kendall's tau statistic in the presence of univariate right censoring by implementing with the nonparametric bivariate survival function estimators in Lin and Ying (1993) and Wang and Wells (1997) with good performance under univariate (right)-censoring. Through simulation study the proposed estimators are compared with other practical estimators $\hat{\tau}_{MO}$ (Oakes, 2008) and $\hat{\tau}_{LRB}$ (Lakhal *et al*., 2009). In the case of data under heavy censoring or with weak association, the modified Hsieh estimators show better performance compared to $\hat{\tau}_{LRB}$ and $\hat{\tau}_{MO}$ and seem comparable in other cases of the simulation. An illustrative analysis is also given.

Keywords: Bivariate survival function, Kendall's tau, univariate-censoring.

1. Introduction

Kendall's τ is the important rank-based association measure along with spearman correlation coefficient. See, for example, Baek *et al*. (2015) and Lee and Ahn (2013). With a pair of independent realizations (X_i, Y_i) and (X_j, Y_j) from (X, Y), it is defined as

$$\tau = E[\text{sgn}((X_i - X_j)(Y_i - Y_j))]$$

where $\text{sgn}(x) = 1$ for $x > 0$, -1 for $x < 0$ and 0 for $x = 0$. It is also interpreted in terms of the probabilities of observing the concordant and discordant pairs. When X and Y are independent, the probabilities of concordance and discordance are same and implies $\tau = 0$. Kendall's τ is a function of the copula parameter so that its estimator naturally

† This research was supported by basic science research program through the national research foundation of Korea (NRF) funded by the ministry of education (No. 2012R1A1A3010532).
[1] Graduate student, Department of Statistics, Chonbuk National University, Jeollabuk-do 54896, Korea.
[2] Corresponding author: Professor, Department of Statistics (Institute of Applied Statistics), Chonbuk National University, Jeollabuk-do 54896, Korea. E-mail: hchoi@jbnu.ac.kr

yield estimators of the copula parameter in semi-parametric models such as Clayton (1978) and Frank (1979). Moreover, due to its rank invariant property and powerful asymptotic properties (Hoeffding, 1948), it is commonly used in survival analysis. Survival analysis is a widely used statistical tool in biostatistics. Censoring is commonly encountered in analysis of lifetime data (cf. Baek *et al.*, 2015; Lee *et al.*, 2015). However, under censoring the estimation of τ has been challenging because there are some pairs that are not comparable because of censoring. A pair is comparable if its concordance-discordance status can be established using the information available in the censored sample.

In this paper, we focus on the nonparametric estimation of τ under univariate-censoring, that is, when only $(\tilde{X}, \tilde{Y}, \delta^X, \delta^Y)$ are observable, where $(\tilde{X}, \tilde{Y}) = (\min(X, C), \min(Y, C))$, $\delta^X = I(X < C)$ and $\delta^Y = I(Y < C)$ are censoring indicators, C is a censoring random variable independent of (X, Y). Many authors studied the nonparametric estimation of Kendall's tau in the presence of censoring. Among these, Oakes (1982), Oakes (2008) and Lakhal *et al.* (2009) suggested some estimators using only comparable pairs under bivariate censoring. The estimator of Oakes (1982) perform poorly when τ is not zero. Oakes (2008) suggested a renormalized estimator of Oakes (1982) and proved that it is consistent when (X, Y) follow Clayton's model. Lakhal *et al.* (2009) considered the inverse probability censoring weight and modified the estimator introduced by Oakes (1982) using a Horvitz-Thompson-type correction for the pairs that are not comparable, which is shown to be consistent and asymptotically normally distributed. Wang and Wells (2000) derived the estimator of τ using the definition of τ as an integral expression of the bivariate survival function. More recently Hsieh (2010) proposed alternative methods which replace a censored event-time by an imputation with the conditional information from the estimated bivariate survival function of Dabrowska (1988) and calculate the τ estimator of Oakes (1982) based on the imputed data. Its consistency and asymptotic normality were established along with a good performance in finite samples.

We are interested in inference for the Kendall's tau in the presence of univariate right censoring. Univariate-censoring is often realistic when the bivariate survival times are, say, observed for two different events on the same individual. Examples include times to severe visual loss on the left and right eyes, and times of cancer detection in the left and right breasts. Such cases are mainly considered here. The proposed estimators in this paper are the modification of Hsieh (2010) by implementing with the nonparametric bivariate survival function estimators under univariate-censoring in place of the bivariate survival function estimator suggested by Dabrowska (1988) under general bivariate censoring. We adopt the bivariate survival function estimators under univariate-censoring in Lin and Ying (1993) and Wang and Wells (1997) since they have considerably simpler forms for practical applications under univariate-censoring (Dabrowska,1988; Prentice and Cai, 1992) and desirable asymptotic properties such as strong consistency and asymptotic normality as well as good finite sample performance. We refer to them as Lin-Ying survival function and Wang-Wells survival function.

The rest of the paper is organized as follows. In Section 2, we briefly review the Kendall's tau estimators for the right censored data. Section 3 presents the modified Hsieh's estimators with Lin-Ying survival function and Wang-Wells survival function. In Section 4, we numerically compare the biases and the mean square errors of the proposed estimators with those of the existing τ estimators (Oakes, 2008; Lakhal *et al.*, 2009). We go on to apply the methods to analyzing the Diabetic Retinopathy Study dataset considered in Huster *et al.*

(1989). Some concluding remarks are given in Section 5.

2. Previous estimators of Kendall's tau under censoring

Let $\{(X_i, Y_i) : i = 1, \ldots, n\}$ be independent replications of (X, Y) with the bivariate survival function $S(x, y)$. Then a Kendall's τ is defined as

$$E[\mathrm{sgn}((X_i - X_j)(Y_i - Y_j))],$$

which is equivalent to $P[(X_i-X_j)(Y_i-Y_j) > 0]-P[(X_i-X_j)(Y_i-Y_j) < 0]$. The (i, j)th pair is called concordant if $(X_i-X_j)(Y_i-Y_j) > 0$ and discordant if $(X_i-X_j)(Y_i-Y_j) < 0$. Thus τ is interpreted as the difference between concordance and discordance rates and $-1 \leq \tau \leq 1$. If X and Y are independent, then $\tau = 0$. In the complete data the representation of τ as

$$E[(2I(X_1 - X_2 > 0) - 1)(2I(Y_1 - Y_2 > 0) - 1)]$$

immediately gives us an estimator of τ

$$\hat{\tau} = \frac{1}{\binom{n}{2}} \sum_{i<j} a_{ij} b_{ij}$$

for $a_{ij} = 2I(X_i - X_j > 0) - 1, b_{ij} = 2I(Y_i - Y_j > 0) - 1$.

We focus on the nonparametric estimation of τ under univariate-censoring, that is, when only $(\widetilde{X}, \tilde{Y}, \delta_X, \delta_Y)$ are observable, where $(\tilde{X}, \tilde{Y}) = (\min(X, C), \min(Y, C))$, $\delta^X = I(X < C)$ and $\delta^Y = I(Y < C)$ are censoring indicators, C is a censoring random variable independent of (X, Y) with survival function $G(\cdot)$. Here and in the sequel, $I(\cdot)$ denotes the indicator function, $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. Suppose we observe i.i.d. copies of $(\tilde{X}, \tilde{Y}, \delta^X, \delta^Y)$, which we denote by $\{(\tilde{X}_i, \tilde{Y}_i, \delta_i^X, \delta_i^Y) : i = 1, \ldots, n\}$.

Among estimators proposed under censoring, we briefly review the Kendall's tau estimators suggested by Oakes (1982) and (2008), Lakhal *et al.* (2009), which are often used in applications. They are numerically compared in later section with our proposed estimators. Oakes (1982) noted that some pairs are comparable in spite of censoring such as a pair $(\tilde{X}_i, \delta_i^X)$ and $(\tilde{X}_j, \delta_j^X)$ with $\tilde{X}_i < \tilde{X}_j$, $\delta_i^X = 1$ and $\delta_j^X = 0$ and extended the Kendall's estimator to the censoring case using comparable pairs only as follows.

$$\hat{\tau}_O = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} L_{ij} a_{ij} b_{ij},$$

where $L_{ij} = 1$ if the (i, j)th pair is comparable or zero if not. However, some information provided by censored data is ignored so that it performs poorly when $\tau \neq 0$.

Oakes (2008) suggested a renormalized estimator τ,

$$\hat{\tau}_{MO} = \frac{C - D}{C + D},$$

where C is the number of concordant pairs and D the number of discordance, which is proved to be consistent when (X, Y) follow Clayton's model.

Lakhal *et al.* (2009) considered the inverse probability censoring weight and modified the estimator in Oakes (1982) using a Horvitz-Thompson-type correction for the pairs that are not comparable. The idea is to correct for censored subjects by giving extra weight to subjects who are not censored.

$$\hat{\tau}_{LRB} = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \frac{L_{ij} a_{ij} b_{ij}}{\hat{p}_{ij}},$$

where $\hat{p}_{ij}$ is the estimator of the selection probabilities,

$$p_{ij} = P\left\{\tilde{X}_i \wedge \tilde{X}_j < C, \tilde{Y}_i \wedge \tilde{Y}_j < C\right\}^2 = \left\{G((\tilde{X}_i \wedge \tilde{X}_j) \vee (\tilde{Y}_i \wedge \tilde{Y}_j))\right\}^2$$

and the survival function of C, $G(\cdot)$, can be estimated by the Kaplan-Meier method based on $\{(\tilde{X}_k \vee \tilde{Y}_k, 1 - \delta_k^X \delta_k^Y) : k = 1, \ldots, n\}$. It has been established that $\hat{\tau}_{LRB}$ is consistent and asymptotically normally distributed, and outperforms $\hat{\tau}_{MO}$ (Lakhal *et al.*, 2009). Our modified Hsieh's estimators with Lin-Ying survival function and Wang-Wells survival function are presented in the following section.

3. Modified Hsieh's approach for univariate-censored data

More recently Hsieh (2010) proposed alternative methods which replace a censored event-time by a imputation with the conditional information obtained from the bivariate survival function. He computed the conditional "mean", "median" or "mode" of the lifetime X and Y for four possible cases of observed censorings, $(\delta_i^X, \delta_i^Y) = \{(1,1), (1,0), (0,1), (0,0)\}$. For example, if the ith observation $\{x_i, y_i, \delta_i^X = 1, \delta_i^Y = 1\}$, we know that $X_i = x_i, Y_i = y_i$. Suppose we have the ith observed variables $\{x_i, y_i, \delta_i^X = 0, \delta_i^Y = 1\}$. Then we know that $X_i > x_i, Y_i = y_i$ and X_i follows the conditional density $f_X(x|X > x_i, Y = y_i)$. Hence the value X_i could be imputed by "mean", "median" or "mode" of $f_X(x|X > x_i, Y = y_i)$. The imputed values are summarized for four possible censoring cases in Table 1.

Hsieh (2010) applied Oakes (1982) to obtain an estimator of τ based on the imputed data $\{(\tilde{X}_i^{type}, \tilde{Y}_i^{type}) : i = 1, \ldots, n\}$ (type="mean", "median" or "mode"), denoted by $\hat{\tau}_{mean}, \hat{\tau}_{median}, \hat{\tau}_{mode}$;

$$\hat{\tau}_{type} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} a_{ij}^{type} b_{ij}^{type}, \tag{3.1}$$

where a_{ij}^{type} and b_{ij}^{type} are the a_{ij} and b_{ij} for the imputed data.

Table 3.1 The imputation of $\tilde{X}$ and $\tilde{Y}$: Values of $\tilde{X}_i^{type}$ and $\tilde{Y}_i^{type}$

(δ_i^X, δ_i^Y)	$\tilde{X}_i^{type}$	$\tilde{Y}_i^{type}$
(1,1)	$\tilde{X}_i$	$\tilde{Y}_i$
(0,1)	type of $f_X(x\|X > x_i, Y = y_i)$	$\tilde{Y}_i$
(1,0)	$\tilde{X}_i$	type of $f_Y(y\|X = x_i, Y > y_i)$
(0,0)	type of $f_X(x\|X > x_i, Y = y_i)$	type of $f_Y(y\|X = x_i, Y > y_i)$

However, the conditional densities $f_X(x|X \in A, Y \in B)$ and $f_Y(y|X \in A, Y \in B)$ in Table 3.1 are unknown so that they should be estimated to implement Hsieh's method.

Noting that they depend on $S(x, y)$, the bivariate survival function of (X, Y), Hsieh (2010) estimated those conditional densities by using the nonparametric bivariate survival function estimator in Dabrowska (1988) under general bivariate censoring. Dabrowska (1988, 1989) presented a nice representation of the bivariate distribution function in terms of the three component bivariate hazard vector and proposed the nonparametric estimator of bivariate survival function based on this representation. But it has considerably complex forms for practical application under univariate-censoring. In the univariate-censoring case, Lin and Ying (1993) and Wang and Wells (1997) developed the nonparametric estimators of the bivariate survival function with practically simpler forms and strong asymptotic properties as well as good finite sample performance. The Wang-Wells estimator takes a very similar structure to the Lin-Ying estimator except for the difference in estimating the survival function of the censoring time as given in the following.

(i) Lin-Ying bivariate survival function estimator $\hat{S}_L(x, y)$: It is based on the representation

$$S(x, y) = \frac{P(\tilde{X} \geq x, \tilde{Y} \geq y)}{G(x \vee y)}$$

due to the independence between (X, Y) and C.

$$\hat{S}_L(x, y) = \frac{\hat{H}(x, y)}{\hat{G}(x \vee y)},$$

where $\hat{H}(x, y)$ is the empirical survival function

$$\frac{1}{n}\sum_{i=1}^{n} I(\tilde{X}_i \geq x, \tilde{Y}_i \geq y)$$

and $\hat{G}(\cdot)$ is the Kaplan-Meier estimator using the data $\{(\tilde{X}_i \vee \tilde{Y}_i, 1 - \delta_i^X \delta_i^Y) : i = 1, \ldots, n\}$.

(ii) Wang-Wells bivariate survival function estimator $\hat{S}_W(x, y)$: It is based on the alternative representation

$$S(x, y) = \frac{P(\tilde{X} \geq x, \tilde{Y} \geq y)}{G(x) \wedge G(y)}.$$

$$\hat{S}_W(x, y) = \frac{\hat{H}(x, y)}{\hat{G}_X(x) \wedge \hat{G}_Y(y)},$$

where $\hat{G}_X(\cdot)$ and $\hat{G}_Y(\cdot)$ are the Kaplan-Meier estimators using the data $(\tilde{X}_i, 1 - \delta_i^X)$ and $(\tilde{Y}_i, 1 - \delta_i^Y)$ $(i = 1, \ldots, n)$, respectively.

As seen above, Lin-Ying estimator and Wang-Wells estimator have considerably simpler forms for practical applications than other bivariate function estimators and natural generalization of the univariate Kaplan-Meier estimator. In the absence of censoring, both reduce to the usual empirical survival function. Furthermore, they have desirable asymptotic properties such as strong consistency and asymptotic normality as well as good finite sample performance. Therefore, we use Lin-Ying estimator and Wang-Wells estimator for the estimation of conditional densities in Table 1, which describes the imputed value in the censored data. All pairs become comparable by imputation so that we apply $\hat{\tau}$ for the complete data to estimate the Kendall's tau.

The modified estimators (3.1) based on Lin-Ying estimator and Wang-Wells estimator are denoted as for $type = mean, med, mode$

$$\hat{\tau}_{L,type} = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} a_{L,ij}^{type} b_{L,ij}^{type} \quad \text{and} \quad \hat{\tau}_{W,type} = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} a_{W,ij}^{type} b_{W,ij}^{type},$$

where $a_{L,ij}^{type}, b_{L,ij}^{type}$ and $a_{W,ij}^{type}, b_{W,ij}^{type}$ are the a_{ij} and b_{ij} for the imputed data based on the Lin-Ying and the Wang-Wells, respectively.

In Section 4, we numerically investigate the performance of the proposed modification, $\hat{\tau}_{L,type}$ and $\hat{\tau}_{W,type}$ (type=mean, median, mode) through extensive numerical studies. The proposed estimators are compared with two existing practical estimators, $\hat{\tau}_{MO}$ (Oakes, 2008) and $\hat{\tau}_{LRB}$ (Lakhal *et al.*, 2009).

4. Numerical studies

4.1. Simulation

In the study, we set the association parameter τ=0.2, 0.5, and 0.8, and consider Clayton model (1978) for the joint distribution of (X, Y) with exponential marginals with mean 1, i.e. the joint survival function of (X, Y)

$$\begin{aligned} S(x, y) &= \{S_1^{-\alpha}(x) + S_2^{-\alpha}(y) - 1\}^{-1/\alpha} \\ &= (e^{\alpha x} + e^{\alpha y} - 1)^{-1/\alpha} \end{aligned}$$

with $\alpha > 0$. The Kendall's τ is expressed as $\tau = \alpha/(2+\alpha)$.

The univariate-censoring times $C_1, C_2, \ldots, C_n$ follow uniform distribution over the interval $(0, \theta)$. We assume independent censoring scheme, i.e. $C_1, C_2, \ldots, C_n$ are independent of $(X_1, Y_1), \ldots, (X_n, Y_n)$. We choose θ in such way that the censoring fractions are about 10%, 30%, and 60%, respectively. We generate 1000 datasets with a size of $n = 50, 100$ and 200.

The performances of estimators are evaluated in terms of the empirical bias and mean-square-error (MSE) for each given different τ defined as

$$\widehat{\text{Bias}} = \frac{1}{r}\sum_{j=1}^{r} \hat{\tau}_j - \tau \quad \text{and} \quad \widehat{\text{MSE}} = \frac{1}{r}\sum_{j=1}^{r} (\hat{\tau}_j - \tau)^2 \quad \text{for } r = 1000 \text{ replications.}$$

The simulation results are summarized in Table 4.1. The results of $\hat{\tau}_{L,mode}$ and $\hat{\tau}_{W,mode}$ are not included because they perform poorly compared to other types for all the models. In addition, we also consider the lognormal bivariate survival distribution but the result shows very similar tendency in the performance of estimators. Thus we don't present the result of the lognormal case. In most cases of sample size and tau value considered for simulation with a 10% and a 30% censoring fractions, the estimator $\hat{\tau}_{LRB}$ outperform other estimators. We note comparable or better performance of $\hat{\tau}_{L,med}$ and $\hat{\tau}_{W,mean}$ than $\hat{\tau}_{LRB}$ under light and moderate censoring (10% and 30%), while $\hat{\tau}_{LRB}$ tends to underestimate τ regardless of the association strength and its performance becomes poor under heavy censoring (60%).

The underestimation property of $\hat{\tau}_{LRB}$ has been observed in various models by Hsieh (2010) as well as Lakhal *et al*. (2009). As expected, there is more bias under heavy censoring than under light censoring in most cases. Under heavy censoring (60%), the modified Hsieh estimators show better performances compared to $\hat{\tau}_{LRB}$. Specifically $\hat{\tau}_{W,med}$ is the best in terms of bias for the case of $\tau = 0.2$, $\hat{\tau}_{L,med}$ for $\tau = 0.5$ and $\hat{\tau}_{L,mean}$ for $\tau = 0.8$ in our simulation models with a 60% censoring rate. Among the modified Hsieh estimators, we see bias smaller in "mean" type estimator than "median" type estimator of a high association ($\tau = 0.8$) regardless of censoring fraction. It is seen that *median* type tends to underestimate for relatively large τ, and *mean* type outperforms. The performance of $\hat{\tau}_{MO}$ shows a poor performance in all cases.

Table 4.1 The estimation of Kendall's tau. In each cell of the estimator column the first number is the empirical bias and the number in the parenthesis the mean squared error of $\hat{\tau}$ based on 1,000 replications. CR stands for censoring rate.

CR(%)	τ	n	$\hat{\tau}_{L,mean}$	$\hat{\tau}_{L,med}$	$\hat{\tau}_{W,mean}$	$\hat{\tau}_{W,med}$	$\hat{\tau}_{MO}$	$\hat{\tau}_{LRB}$
10	0.2	50	0.0149 (0.0101)	0.0034 (0.0096)	0.0013 (0.0098)	-0.0052 (0.0099)	0.0199 (0.0114)	-0.0020 (0.0102)
		100	0.0122 (0.0049)	0.0010 (0.0048)	-0.0003 (0.0048)	-0.0064 (0.0047)	0.0185 (0.0055)	-0.0016 (0.0049)
		200	0.0120 (0.0024)	-0.0007 (0.0023)	0.0010 (0.0023)	-0.0049 (0.0023)	0.0198 (0.0028)	0.0005 (0.0023)
	0.5	50	0.0111 (0.0069)	-0.0042 (0.0070)	-0.0016 (0.0068)	-0.0153 (0.0073)	0.0358 (0.0082)	-0.0013 (0.0067)
		100	0.0100 (0.0037)	-0.0077 (0.0037)	-0.0013 (0.0036)	-0.0155 (0.0040)	0.0362 (0.0049)	0.0004 (0.0036)
		200	0.0089 (0.0018)	-0.0098 (0.0019)	-0.0017 (0.0017)	-0.0173 (0.0020)	0.0020 (0.0357)	0.0007 (0.0017)
	0.8	50	0.0050 (0.0023)	-0.0155 (0.0028)	-0.0001 (0.0023)	-0.0362 (0.0039)	0.0287 (0.0034)	-0.0004 (0.0022)
		100	0.0053 (0.0017)	-0.0229 (0.0019)	-0.0008 (0.0015)	-0.0373 (0.0025)	0.0279 (0.0028)	0.0001 (0.0015)
		200	0.0047 (0.0012)	-0.0271 (0.0014)	-0.0015 (0.0010)	-0.0383 (0.0019)	0.0019 (0.0270)	-0.0002 (0.0010)
30	0.2	50	0.0505 (0.0135)	0.0171 (0.0117)	0.0125 (0.0118)	-0.0072 (0.0119)	0.0624 (0.0183)	-0.0135 (0.0141)
		100	0.0481 (0.0660)	0.0120 (0.0055)	0.0084 (0.0057)	-0.0124 (0.0063)	0.0612 (0.0094)	-0.0067 (0.0069)
		200	0.0475 (0.0040)	0.0090 (0.0030)	0.0099 (0.0031)	-0.0120 (0.0038)	0.0624 (0.0058)	-0.0015 (0.0036)
	0.5	50	0.0385 (0.0099)	-0.0129 (0.0098)	0.0109 (0.0093)	-0.0376 (0.0117)	0.1070 (0.0192)	-0.0123 (0.0106)
		100	0.0401 (0.0067)	-0.0180 (0.0065)	0.0100 (0.0057)	-0.0402 (0.0081)	0.1076 (0.0156)	-0.0060 (0.0064)
		200	0.0394 (0.0046)	-0.0221 (0.0045)	0.0096 (0.0036)	-0.0414 (0.0062)	0.1079 (0.0135)	-0.0014 (0.0037)
	0.8	50	0.0201 (0.0066)	-0.0476 (0.0083)	0.0137 (0.0063)	-0.0921 (0.0126)	0.0762 (0.0137)	-0.0122 (0.0059)
		100	0.0224 (0.0067)	-0.0657 (0.0074)	0.0113 (0.0058)	-0.0959 (0.0106)	0.0748 (0.0141)	-0.0078 (0.0051)
		200	0.0253 (0.0066)	-0.0771 (0.0076)	0.0113 (0.0054)	-0.0969 (0.0099)	0.0748 (0.0140)	-0.0044 (0.0044)
60	0.2	50	0.1815 (0.0292)	0.0758 (0.0198)	0.1333 (0.0216)	0.0206 (0.0230)	0.1797 (0.0466)	-0.0410 (0.0497)
		100	0.1926 (0.0228)	0.0706 (0.0128)	0.1297 (0.0137)	0.0101 (0.0190)	0.1853 (0.0284)	-0.0312 (0.0360)
		200	0.1929 (0.0193)	0.0649 (0.0088)	0.1266 (0.0096)	0.0054 (0.0153)	0.1857 (0.0210)	-0.0205 (0.0239)
	0.5	50	0.1340 (0.0289)	-0.0010 (0.0299)	0.1241 (0.0259)	-0.0523 (0.0336)	0.2583 (0.0638)	-0.0645 (0.0435)
		100	0.1478 (0.0259)	-0.0185 (0.0240)	0.1229 (0.0213)	-0.0619 (0.0305)	0.2609 (0.0604)	-0.0532 (0.0337)
		200	0.1542 (0.0254)	-0.0268 (0.0213)	0.1206 (0.0192)	-0.0691 (0.0290)	0.2612 (0.0592)	-0.0399 (0.0252)
	0.8	50	0.0432 (0.0244)	-0.1103 (0.0339)	0.0794 (0.0312)	-0.1855 (0.0424)	0.1421 (0.0503)	-0.1094 (0.0251)
		100	0.0561 (0.0282)	-0.1447 (0.0321)	0.0764 (0.0327)	-0.1968 (0.0411)	0.1419 (0.0535)	-0.0945 (0.0229)
		200	0.0702 (0.0319)	-0.1674 (0.0336)	0.0766 (0.0332)	-0.2015 (0.0410)	0.1419 (0.0544)	-0.0824 (0.0190)

4.2. Real data analysis

As an illustrative example, we analyzed a subset of data with 197 high-risk patients from the Diabetic Retinopathy Study, which was considered in Huster *et al.* (1989). This study was conducted to investigate the effectiveness of laser photocoagulation in delaying the onset of blindness in diabetic retinopathy patients. One eye of each patient was randomly selected for treatment and the other eye was observed without treatment. The bivariate survival times in this data set are the time to blindness in months for the treated and untreated eyes. Since both eyes of a patient are observed at the same time, the survival times are subject to univariate-censoring, caused by death, droptout, or end of the study. The data consists of 197 patients with censoring rates 73% and 49% for the treated and untreated eyes, respectively, i.e. overall censoring rate 61% . One interesting question of the DRS study was whether the survival times for the treated eye of a patient is related to the survival time of the untreated eye.

We estimated τ to measure the association between survival times of the treated and untreated eyes. To check the performance of the proposed method, we apply the bootstrap approach with 1000 replications. The estimated τ values are reported in Table 4.2 along with 95% confidence interval obtained by normal approximation. The estimated τ values are all positive. $\hat{\tau}_{LRB}$ is the smallest and $\hat{\tau}_{L,med}$ is the second smallest in the DRS. Using $\hat{\tau}_{LRB}$, the smallest value, the 95% confidence interval (0.1460,0.1590). As we noted above, the underestimation property of $\hat{\tau}_{LRB}$ implies that the association of the survival times of two eyes is positive and significantly different from 0 as Huster *et al*. (1989) studied the association parametrically. Taking into account the censoring rate of this data, we may consider the modified Hsieh estimtors as more reliable.

Table 4.2 The estimation of Kendall's tau for Diabetic Retinopathy Study

	$\hat{\tau}_{L,mean}$	$\hat{\tau}_{L,med}$	$\hat{\tau}_{W,mean}$	$\hat{\tau}_{W,med}$	$\hat{\tau}_{MO}$	$\hat{\tau}_{LRB}$
	0.3445	0.2252	0.3433	0.2776	0.3606	0.1525
95% CI	(0.3341, 0.3550)	(0.2139, 0.2364)	(0.3335, 0.3531)	(0.2673, 0.2879)	(0.3488, 0.3724)	(0.1460, 0.1590)

5. Discussion

We proposed a nonparametric estimator of Kendall's tau under univariate-censoring. The estimators suggested in this paper are motivated by the idea of Hsieh (2010) and implemented with the nonparametric bivariate survival function estimator of Lin and Ying (1993) and Wang and Wells (1997) under univariate-censoring, which is a very simple compared to Dabrowska's and shows a good performance. We numerically compare the empirical bias and mean square error of our proposed modification with other existing estimators of τ. Although the estimator $\hat{\tau}_{LRB}$ outperform other estimators in the cases of strong association under light or moderate censoring, the modified Hsieh type estimators may be comparable to $\hat{\tau}_{LRB}$. Especially in the case of data under heavy censoring or with weak association, the modified Hsieh estimators show better performance compared to $\hat{\tau}_{LRB}$ so that they might be recommendable under heavy censoring at least in our simulation model. We noted that $\hat{\tau}_{LRB}$ tends to underestimate the τ as the censoring rate becomes heavier, while $\hat{\tau}_{L,mean}$ and $\hat{\tau}_{W,mean}$ tend to overestimate. This issue will be carefully considered as our future work.

References

Baek, H., Han, J. and Lee, H. (2015). Analysis of mortality after death of spouse in relation to duration of bereavement and dependence relation between married couple -using married couples data from survivor's pension of National Pension Service. *Journal of the Korean Data & Information Society*, **26**, 931-946.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141-151.

Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *The Annals of Statistics*, **16**, 1475-1489.
Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *The Annals of Statistics*, **17**, 1157-1167.
Frank, M. J. (1979). On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematicae*, **19**, 194-226.
Hoeffding, W. (1948). A class of statistics with asymptotic normal distribution. *The Annals of Mathematical Statistics*, **19**, 293-325.
Hsieh, J. J. (2010). Estimation of Kendall's tau from censored data. *Computational Statistics & Data Analysis*, **54**, 1613-1621.
Huster, W. J., Brookmeyer, R. and Self, S. G. (1989). Modelling paired survival data with covariates. *Biometrika*, **45**, 145-156.
Lee, W. J. and Ahn, J. Y. (2013). On the application of multivariate Kendall's tau and its interpretation. *The Korean Journal of Applied Statistics*, **26**, 495-509.
Lee, S., Shim, B. Y. and Kim, J. (2015). Estimation of hazard function and hazard change-point for the rectal cancer data. *Journal of the Korean Data & Information Society*, **26**, 1225-1238.
Lakhal, L., Rivest, L. P. and Beaudoin, D. (2009). IPCW estimator for Kendall's tau under bivariate censoring. *The International Journal of Biostatistics*, **5**, 1121-1141.
Lin, D. Y. and Ying, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, **80**, 573-581.
Oakes, D. (1982). A concordance test for independence in the presence of censoring. *Biometrics*, **38**, 451-455.
Oakes, D. (2008). On consistency of Kendall's tau under censoring. *Biometrika*, **95**, 997-1001.
Prentice, R. L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, **79**, 495-512.
Wang, W. and Wells, M. T. (1997). Nonparametric estimators of the bivariate survival function under simplified censoring conditions. *Biometrika*, **84**, 863-880.
Wang, W. and Wells, M. T. (2000). Estimation of Kendall's tau under censoring. *Statistica Sinica*, **10**, 1199-1215.