Nonparametric estimation and test of conditional Kendall’s tau under semi-competing risks data and truncated data

Jin-Jian Hsieh* and Wei-Cheng Huang

Department of Mathematics, National Chung Cheng University, Chia-Yi, Taiwan, ROC

(Received 12 January 2014; accepted 4 January 2015)

In this article, we focus on estimation and test of conditional Kendall’s tau under semi-competing risks data and truncated data. We apply the inverse probability censoring weighted technique to construct an estimator of conditional Kendall’s tau, $\tau_c$. Then, this study provides a test statistic for $H_0: \tau_c = \tau_0$, where $\tau_0 \in (-1, 1)$. When two random variables are quasi-independent, it implies $\tau_c = 0$. Thus, $H_0: \tau_c = 0$ is a proxy for quasi-independence. Tsai [10], and Martin and Betensky [8] considered the testing problem for quasi-independence. Via simulation studies, we compare the three test statistics for quasi-independence, and examine the finite-sample performance of the proposed estimator and the suggested test statistic. Furthermore, we provide the large sample properties for our proposed estimator. Finally, we provide two real data examples for illustration.

Keywords: conditional Kendall’s tau; IPCW; quasi-independence; semi-competing risks data; truncated data

1. Introduction

Let $(X, Y)$ be a pair of failure times which are possibly correlated. In many biomedical applications, bivariate survival analysis has attracted substantial attention due to its wide applications. In the past few years, applications have been extended to include multiple event times which have dependently censored or truncated relationship, such as semi-competing risks data and truncated data. For semi-competing risks data, there are two type events, terminal event and non-terminal event. The time to terminal event $Y$ may censor the time to non-terminal event $X$, but not vice versa. The paper by Tsai [10] considered the truncated data in which one variable $X$ truncates the other $Y$, and the two variables are possibly correlated. Therefore, we can observe the pair $(X, Y)$ only when $X < Y$.

*Corresponding author. Emails: jjhsieh@math.ccu.edu.tw, jinjian.hsieh@gmail.com

© 2015 Taylor & Francis
In bivariate survival analysis, we are often interested in association measurement between two variables X and Y. There is an association measure, Kendall’s tau, which measures the association between X and Y. Kendall’s tau is an usual association measurement in life-time models. However, when data become incomplete due to dependently censored or truncated, such as semi-competing risks data or truncated data, Kendall’s tau is non-identifiable without extra assumptions. There were some literatures discussed the association of (X, Y) for semi-competing risks data and truncated data under some proper assumptions, such as Fine et al. [2], Wang [11], and Lakhal et al. [6] for semi-competing risks data, and Chaieb et al. [1] for truncated data. All of these literatures considered a copula model to specify the dependence.

Tsai [10] proposed the conditional Kendall’s tau, τ_c, to measure the association of (X, Y) under comparable set for truncated data. Although τ_c cannot identify all the association information of (X, Y), it can reveal a part of the association information of (X, Y). Via the partial association information of (X, Y), τ_c, we can understand the relation of (X, Y) under the restriction due to incomplete data. By the same way, following Tsai’s definition, we also can define the conditional Kendall’s tau for semi-competing risks data. When X and Y are independent under the comparable set, we say that X and Y are quasi-independent, denoted as X ⊥ Q Y. Tsai [10] and Martin and Betensky [8] provided the statistical approaches to test the hypothesis, H_0 : τ_c = 0, which is a proxy for quasi-independence under truncated data.

In this article, we focus on estimation and test of conditional Kendall’s tau, τ_c, under semi-competing risks data and truncated data. τ_c can reveal the association information of (X, Y) under comparable set, which provides the association information of (X, Y) as more as possible under a restriction caused by incomplete data. For the testing problem, we consider the more general testing problem, H_0 : τ_c = τ_0, where τ_0 ∈ (−1, 1), instead of H_0 : τ_c = 0 considered by Tsai [10] and Martin and Betensky [8].

In Section 2, we introduce two data structures, semi-competing risks data and truncated data, and the definition of conditional Kendall’s tau. In Section 3, we use the IPCW technique to construct an estimator of conditional Kendall’s tau, derive the large sample properties, and propose a test statistic for conditional Kendall’s tau under semi-competing risks data and truncated data. Then, we examine the finite-sample performance of our proposed estimator and testing procedure in Section 4. In Section 5, we analyze two real data sets for illustration. Finally, we give the concluding remarks in Section 6.

2. Data and conditional Kendall’s tau

Let (X, Y) be the two life-time variables and C be an external right-censoring variable. We discuss the two data structures in the following.

2.1 Semi-competing risks data

Semi-competing risks data arise in the situation that one event is terminal, such as death, while the other is non-terminal, such as disease progression. Let X be the time to non-terminal event and Y be the time to terminal event. Y may censor X but Y is not censored by X. C is a right-censoring variable, which is independent of X and Y. Define the observed variables \( S = \min(X, Y, C) \), \( \delta_X = I(X \leq Y \wedge C) \), \( T = \min(Y, C) \), and \( \delta_Y = I(Y \leq C) \), where \( I(\cdot) \) is an indicator function. The observable data are \( (S, T, \delta_X, \delta_Y) \). For the study of leukemia patients receiving bone marrow transplants, let X be the time of leukemia relapse, Y be the time of death and C be the time of right-censoring. We know that when death occurs prior to relapse, we cannot observe X. But, if a patient relapses, we can still observe the death time Y if \( Y \leq C \). So the terminal event censors the non-terminal event, but not vice versa.
2.2 Truncated data

Tsai [10] considered the data structure in which one variable truncates the other and the two variables are possibly correlated. Let \( X \) be the left-truncated time and \( Y \) be the failure time. Thus, we can observe a pair \((X, Y)\) only if \( X < Y \). Such data are called truncated data. Let \( C \) be a right-censoring variable, and assume \( C \) and \( Y \) are conditional independent given \( X \). Define the observable variables \( T = \min(Y, C) \) and \( \delta = \mathbb{I}(Y \leq C) \). The observable data are \((X, T, \delta)\). For the Channing house retirement center example [5], subjects are observed only if they live long enough to enter the retirement house. Let \( X \) be the retirement house entry age, \( Y \) be the lifetime and \( C \) be the right-censoring time. Obviously, we observe a subject only when the subject enters the retirement house, that is, \( X < Y \).

2.3 Conditional Kendall’s tau

In bivariate survival analysis, we are often interested in association between two variables \( X \) and \( Y \). Kendall [4] proposed the Kendall’s tau to measure the dependency between \( X \) and \( Y \). Let \((X_1, Y_1)\) and \((X_2, Y_2)\) be two independent vectors with the same bivariate distribution of \((X, Y)\). If \((X_1 - X_2)(Y_1 - Y_2) > 0\), the pairs is said to be concordant, and discordant if \((X_1 - X_2)(Y_1 - Y_2) < 0\). Therefore, Kendall’s tau is defined as

\[
\tau = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0) = E[\text{sgn}((X_1 - X_2)(Y_1 - Y_2))],
\]

where \( \text{sgn}(x) = 1 \) for \( x > 0 \), \( \text{sgn}(x) = -1 \) for \( x < 0 \), and \( \text{sgn}(x) = 0 \) for \( x = 0 \). The value of Kendall’s tau is between \(-1\) and 1. Because of its rank invariant property, it is suitable for measuring the association in lifetime models. When \( X \) and \( Y \) are independent, the probability of concordance should be equal to the probability of discordance. That is, when \( X \) and \( Y \) are independent, it implies \( \tau = 0 \). Therefore, the null hypothesis, \( H_0 : \tau = 0 \), is a proxy for the null hypothesis, \( H_0' : X \perp Y \).

In some situations, data become incomplete due to dependently censored or truncated, such as semi-competing risks data or truncated data. For this kind data, the pairs \((X_1, Y_1)\) and \((X_2, Y_2)\) are comparable only in the comparable set, \( \Omega_{12} \). Thus, the Kendall’s tau is non-identifiable for semi-competing risks data and truncated data. Tsai [10] proposed the conditional Kendall’s tau for truncated data, which is defined as

\[
\tau_c = P((X_1 - X_2)(Y_1 - Y_2) > 0|\Omega_{12}) - P((X_1 - X_2)(Y_1 - Y_2) < 0|\Omega_{12}) = E[\text{sgn}((X_1 - X_2)(Y_1 - Y_2))|\Omega_{12}],
\]

where \( \Omega_{12} \) is the condition that \((X_1, Y_1)\) and \((X_2, Y_2)\) are comparable under incomplete data. For the semi-competing risks data, the comparable set \( \Omega'_{12} = \{\tilde{X}_{12} < \tilde{Y}_{12}\} \), where \( \tilde{X}_{12} = \min(X_1, X_2) \) and \( \tilde{Y}_{12} = \min(Y_1, Y_2) \). For the truncated data, the comparable set is \( \Omega'_{12} = \{\tilde{X}_{12} < \tilde{Y}_{12}\} \), where \( \tilde{X}_{12} = \max(X_1, X_2) \). \( \tau_c \) measures the association between \( X \) and \( Y \) under comparable set, \( \Omega_{12} \), and its value lies between \(-1\) and 1. Although \( \tau_c \) cannot provide all the association information of \((X, Y)\), it presents the partial association information of \((X, Y)\) as more as possible under a restriction due to incomplete data. When \( X \) and \( Y \) are independent under the condition \( \Omega_{12} \), we say that \( X \) and \( Y \) are quasi-independent, denoted as \( X \perp_Q Y \). Under \( X \perp_Q Y \), it can show that \( \tau_c = 0 \). Therefore, \( H_0 : \tau_c = 0 \) is a proxy for \( H_0' : X \perp_Q Y \).

In the presence of right-censoring, we cannot estimate \( \tau_c \) directly because the pairs \((X_1, Y_1)\) and \((X_2, Y_2)\) are comparable only in comparable and orderable set, \( A_{12} \), which is included in \( \Omega_{12} \).
Martin and Betensky [8] defined $\bar{\tau}_c$ as
\[
\bar{\tau}_c = P((X_1 - X_2)(Y_1 - Y_2) > 0|A_{12}) - P((X_1 - X_2)(Y_1 - Y_2) < 0|A_{12})
\]
\[
= E[\text{sgn}((X_1 - X_2)(Y_1 - Y_2))|A_{12}],
\]
where $A_{12}$ is the condition that $(X_1, Y_1)$ and $(X_2, Y_2)$ are comparable under incomplete data with right-censoring time. For the semi-competing risks data, the comparable and orderable set is $A_{12}^* = \{\tilde{X}_{12} < \tilde{Y}_{12} < \tilde{C}_{12}\}$, where $\tilde{C}_{12} = \min(C_1, C_2)$. For the truncated data, the comparable and orderable set is $A_{12}' = \{\tilde{X}_{12} < \tilde{Y}_{12} < \tilde{C}_{12}\}$. Under $H_0' : X \perp Q Y$, Martin and Betensky [8] showed that $\bar{\tau}_c = \tau_c = 0$.

3. The inference procedures

3.1 Estimation of conditional Kendall’s tau

Under truncated data, $\{(X_i, T_i, \delta_i) : i = 1, \ldots, n\}$, Martin and Betensky [8] estimated $\bar{\tau}_c$ by
\[
\hat{\tau}^i_{c,m} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sgn}((X_i - X_j)(T_i - T_j))I(A^i_{ij})}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(A^i_{ij})},
\]
which is valid in estimating $\tau_c$ only when $X \perp Q Y$. Therefore, we want to develop an appropriate estimator of conditional Kendall’s tau for general situation under semi-competing risks data and truncated data. Here, we use the technique, inverse probability censoring weighted (IPCW) [7], to modify the estimator.

Under semi-competing risks data with the presence of right-censoring variable, we can determine the concordance or discordance of two pairs only in comparable and orderable set, $A_{12}' = \{\tilde{X}_{12} < \tilde{Y}_{12} < \tilde{C}_{12}\}$. We will apply the survey sampling technique [3] to correct the bias for the estimation of $\tau_c$. We define a selection probability as
\[
p^i_{ij} = P(\tilde{Y}_{ij} < \tilde{C}_{ij}|\tilde{X}_{ij}, \tilde{Y}_{ij}) = (P(C > \tilde{Y}_{ij}))^2,
\]
where $p^i_{ij}$ is the selection probability for comparable $(i, j)$ pair. Let $S = \min(X, Y, C)$, $T = \min(Y, C)$, $\delta_X = I(X \leq Y \land C)$, and $\delta_Y = I(Y \leq C)$. By the IPCW technique, we propose an estimator of $\tau_c$ for semi-competing risks data as
\[
\hat{\tau}^i_c = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sgn}((\tilde{S}_i - \tilde{S}_j)(\tilde{T}_i - \tilde{T}_j))I(A^i_{ij})/\tilde{p}^i_{ij}}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(A^i_{ij})/\tilde{p}^i_{ij}},
\]
where $\tilde{p}^i_{ij} = (\hat{P}(C > \tilde{Y}_{ij}))^2$ and $\hat{P}(C > t)$ is the Kaplan–Meier estimator based on $(T_i, 1 - \delta_Y, i = 1, \ldots, n)$. Note that $\tilde{p}^i_{ij}$ corrects the selection bias introduced from censoring.

Under truncated data, Martin and Betensky [8] provided a consistent estimator of $\bar{\tau}_c$ as $\hat{\tau}^i_{c,m}$. But, the estimator is valid for the estimation of $\tau_c$ only when $X \perp Q Y$. Thus, we would like to develop an estimator of $\tau_c$ under truncated data for general situation. The method is similar to the previous procedure for semi-competing risks data. Define a selection probability as
\[
p^i_{ij} = P(\tilde{Y}_{ij} < \tilde{C}_{ij}|X_i, Y_i, \tilde{X}_{ij}, \tilde{Y}_{ij}),
\]
where $p^i_{ij}$ is the individual selection probability for comparable $(i, j)$ pair. Because $C > X$ is always true, let $C = X + D$ and assume $D$ is independent of $(X, Y)$. Because $D$ can be viewed
as the follow-up time after entering the study, which can be assumed to be independent of the truncated time and life time, the assumption is reasonable. Then, the \( p_{ij}' \) can be written as

\[
p_{ij}' = P(\tilde{Y}_{ij} < \tilde{C}_{ij}|X_i, X_j, \tilde{X}_{ij}, \tilde{Y}_{ij})
\]

\[
= P(C_i - X_i > \tilde{Y}_{ij} - X_i, C_j - X_j > \tilde{Y}_{ij} - X_j|X_i, X_j)
\]

\[
= P(D_i > \tilde{Y}_{ij} - X_i, D_j > \tilde{Y}_{ij} - X_j|X_i, X_j)
\]

\[
= P(D_i > \tilde{Y}_{ij} - X_i)P(D_j > \tilde{Y}_{ij} - X_j)
\]

\[
= P(D > \tilde{Y}_{ij} - X_i)P(D > \tilde{Y}_{ij} - X_j).
\]

By the IPCW technique, we can estimate \( \tau_c \) for truncated data by

\[
\hat{\tau}_c^t = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sgn}((X_i - X_j)(T_i - T_j))I(A_{ij}^t)/\hat{p}_{ij}' }{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(A_{ij}^t)/\hat{p}_{ij}' },
\]

(5)

where \( A_{ij}^t = \{ \tilde{X}_{ij} < \tilde{Y}_{ij} < \tilde{C}_{ij} \}, \hat{p}_{ij}' = \hat{P}(D > \tilde{Y}_{ij} - X_i)\hat{P}(D > \tilde{Y}_{ij} - X_j) \), and \( \hat{P}(D > t) \) is the Kaplan–Meier estimator based on \( (T_i - X_i, 1 - \delta_i, i = 1, \ldots, n) \).

### 3.2 Large sample properties

In this section, we provide the asymptotic properties of our proposed estimator. To derive the large sample properties, we assume the following conditions:

**C1:** For semi-competing risks data, assume all pairs \((X, Y)\) fall inside \([0, L_X] \times [0, L_Y]\) and \( P(C > L) > \epsilon_1 > 0 \), where \( L = \max(L_X, L_Y) \), for some constant \( \epsilon_1 \). It has that

\[
P(p_{ij}' > \epsilon_1^2 \text{ for all } (i, j)) = 1.
\]

**C2:** For truncated data, assume \( D = Y - X \) fall inside \([0, L_{XY}]\) and \( P(D > L_{XY}) > \epsilon_2 > 0 \), for some constant \( \epsilon_2 \). It has that

\[
P(p_{ij}' > \epsilon_2^2 \text{ for all } (i, j)) = 1.
\]

The conditions C1 and C2 avoid \( p_{ij}' (\ast = s \text{ or } t) \) to be zero, which avoid to appear zero in denominator for the asymptotic properties derivation. However, in practical applications, we do not need the conditions because \( \hat{p}_{ij}' \) and \( \hat{p}_{ij}' \) must be positive. In the following, we present a theorem for the asymptotic properties of our suggested estimator.

**Theorem 1** Under condition C1 for semi-competing risks data and C2 for truncated data, \( \hat{\tau}_c^* \) is a consistent estimator of \( \tau_c \) and \( \sqrt{n}(\hat{\tau}_c^* - \tau_c) \) is asymptotic normal with zero-mean, where \( \ast = s \text{ or } t \).

We present the detail proof in appendix of supplemental materials for semi-competing risks data. Similarly, the proof for truncated data is similar to semi-competing risks data and we omit it.

Since the asymptotic variance of \( \hat{\tau}_c^* \), \( \ast = s \text{ or } t \), is difficult to estimate, we use the jackknife approach to estimate the variance, which is shown as

\[
\hat{\sigma}^2 = \frac{n - 1}{n} \sum_{i=1}^{n} (\hat{\tau}_c^{* (i)} - \hat{\tau}_c^*)^2,
\]

where \( \hat{\tau}_c^{* (i)} \) is \( \hat{\tau}_c^* \) based on data deleting the \( i \)th subject and \( \hat{\tau}_c^* = n^{-1} \sum_{i=1}^{n} \hat{\tau}_c^{* (i)} \).
3.3 Testing for conditional Kendall’s tau

Tsai [10] suggested an approach to test whether $X$ and $Y$ are quasi-independent. Since $X \perp Q Y$ implies $\tau_c = 0$, $H_0 : \tau_c = 0$ is a proxy for $H_0' : X \perp Q Y$. He tested $H_0 : \tau_c = 0$ by $U$-statistic with exact variance formula. Martin and Betensky [8] also provided a test for $H_0 : \tau_c = 0$ by $U$-statistic with asymptotic variance. In Section 3.1, we propose an estimator for $\tau_c$, and provide the asymptotic properties in Section 3.2. In this section, we suggest a test statistic based on $\hat{\tau}_c$.

Our statistic can test not only $H_0 : \tau_c = 0$ but also $H_0 : \tau_c = \tau_0$, where $\tau_0 \in (-1, 1)$.

Under semi-competing risks data, our estimator of $\tau_c$ in Section 3.1 is $\hat{\tau}_c$, which is shown in (4). For $H_0 : \tau_c = \tau_0$, our test statistic is

$$T_s = \frac{\hat{\tau}_c - \tau_0}{\hat{\sigma}}$$

where $\hat{\sigma}$ is an estimator of standard deviation by the jackknife method. By the theorem in Section 3.2 and Theorem 2.1 in Shao and Tu [9], we can show that $T_s \overset{d}{\rightarrow} N(0, 1)$ under $H_0$. Therefore, $T_s^2 \overset{d}{\rightarrow} \chi_1^2$. We reject $H_0$ when $T_s^2 > \chi_1^2(\alpha)$, where $\alpha$ is the type I error.

Under truncated data, our estimator of $\tau_c$ is $\hat{\tau}_c^t$, as shown in Equation (5). For $H_0 : \tau_c = \tau_0$, our test statistic is

$$T_t = \frac{\hat{\tau}_c^t - \tau_0}{\hat{\sigma}}$$

where $\hat{\sigma}$ is an estimator of standard deviation by the jackknife method. By the theorem in Section 3.2 and Theorem 2.1 in Shao and Tu [9], we have that $T_t \overset{d}{\rightarrow} N(0, 1)$ and $T_t^2 \overset{d}{\rightarrow} \chi_1^2$ under $H_0$. Therefore, we reject $H_0$ when $T_t^2 > \chi_1^2(\alpha)$, where $\alpha$ is the significant level.

4. Simulation studies

In this section, we examine the finite-sample performance of our proposed estimator and testing procedure. We consider two data structures, semi-competing risks data and truncated data. Also, we use three different dependence structures to simulate, such as Clayton copula, Frank copula, and the beta frailty model. Furthermore, we compare our test statistic for quasi-independence with Tsai [10] and Martin and Betensky [8].

4.1 Semi-competing risks data

For semi-competing risks data, we consider three dependence structures, Clayton copula, Frank copula, and beta frailty model. For Clayton copula, the joint survival function of $X$ and $Y$ can be written as

$$P(X > x, Y > y) = (S_X(x)^{-\alpha} + S_Y(y)^{-\alpha} - 1)^{-1/\alpha}.$$  

For Frank copula, the joint survival function of $X$ and $Y$ is

$$P(X > x, Y > y) = \log_\alpha \left( 1 + \frac{\alpha S_X(x) - 1)(\alpha S_Y(y) - 1)}{\alpha - 1} \right),$$

where $S_X(x)$ and $S_Y(y)$ are marginal survival functions of $X$ and $Y$, and $\alpha$ is the corresponding association parameter. For the frailty model, the joint survival function of $X$ and $Y$ can be written as

$$P(X > x, Y > y) = \int (S_X(x)y'(S_Y(y))' f_r(r) dr,$$
Table 1. Estimation of $\tau_c$ under semi-competing data for Clayton copula and the beta frailty model.

| $\tau_c$ | 0.1  | 0.2  | 0.3  | 0.4  | 0.5  | 0.6  | 0.7  | 0.8  |
|----------|------|------|------|------|------|------|------|------|
| Clayton  |      |      |      |      |      |      |      |      |
| mean     | 0.0977 | 0.1922 | 0.2957 | 0.3939 | 0.4995 | 0.5983 | 0.6991 | 0.7985 |
| Bias     | -0.0023 | -0.0078 | -0.0043 | -0.0061 | -0.0005 | -0.0017 | -0.0009 | -0.0015 |
| SE       | 0.0902 | 0.0842 | 0.0827 | 0.0720 | 0.0637 | 0.0557 | 0.0401 | 0.0307 |
| SE_j     | 0.0904 | 0.0869 | 0.0814 | 0.0748 | 0.0657 | 0.0559 | 0.0441 | 0.0321 |
| MSE      | 0.0082 | 0.0076 | 0.0066 | 0.0056 | 0.0043 | 0.0031 | 0.0019 | 0.0010 |
| C.P.     | 0.940  | 0.954  | 0.937  | 0.958  | 0.957  | 0.946  | 0.955  | 0.947  |
| Cen_X    | 16.38  | 16.65  | 16.72  | 16.46  | 16.68  | 16.82  | 16.54  | 16.67  |
| Frailty  |      |      |      |      |      |      |      |      |
| mean     | 0.0973 | 0.2040 | 0.3001 | 0.3986 | 0.5014 |
| Bias     | -0.0027 | -0.0040 | -0.0001 | -0.0014 | 0.0014 |
| SE       | 0.0904 | 0.0920 | 0.0880 | 0.0873 | 0.0850 | 0.0834 |
| SE_j     | 0.0902 | 0.0894 | 0.0885 | 0.0879 | 0.0882 |
| MSE      | 0.0081 | 0.0080 | 0.0078 | 0.0078 | 0.0079 |
| C.P.     | 0.936  | 0.950  | 0.951  | 0.951  | 0.950  |
| Cen_Y    | 40.96  | 43.06  | 46.86  | 51.71  | 57.71  |

Notes: ‘mean’: the average of $\hat{\tau}_c$; ‘Bias’: the average bias; ‘SE’: the empirical standard deviation; ‘SE_j’: the mean of standard deviation estimator by the jackknife method; ‘MSE’: the mean square error; ‘C.P.’: the coverage probability of nominal 95% confidence interval; ‘Cen_X’: the censoring percentage of $X$; ‘Cen_Y’: the censoring percentage of $Y$. Sample size $n = 100$, replications: 1000. For the settings of $(a, b)$ in Beta$(a, b)$ are corresponding to the setting of $\tau_c$ with $b = 1$.

where $f_r(r)$ is a density function of frailty variable $r$. Here, we consider $r$ follows Beta$(a, b)$. Let $X$ be an exponential distribution with mean 0.6, $Y$ be an exponential distribution with mean 1. For Clayton and Frank copulas, let $C$ be a uniform distribution on $(0, 6)$, and for beta frailty model, let $C$ be a uniform distribution on $(0, 15)$. The variable $C$ for beta frailty model is different from others since we would control the censoring percentage. Furthermore, The sample size is 100 and the number of replications is 1000.

For estimations, we set $\tau_c$ from 0.1 to 0.8 for Clayton copula, from $-0.8$ to 0.8 for Frank copula, and from 0.1 to 0.5 for the beta frailty model. For the setting of $\tau_c$, we set it according to the relation, $\tau = \tau_c = \alpha/(\alpha + 2)$, for Clayton copula, and, for Frank copula and beta frailty, we compute the approximate true $\tau_c$ based on a large number of $(X_i, Y_i)$. The results are given in Tables 1 and 2. For testing procedure, let $\tau_0 = 0.5$ for Clayton copula; $\tau_0 = 0.5$ and $-0.5$ for Frank copula; and $\tau_0 = 0.3$ for the beta frailty model. The results are presented in Figure 1. From the results, it shows the biases and mean squared errors are very small and the coverage probability of the nominal 95% confidence interval is near 0.95. Furthermore, the average standard deviation by jackknife is close to the empirical standard deviation. Moreover, the performance of the suggested test statistic is well.

4.2 Truncated data

For truncated data, we also consider three dependence structures. For Clayton copula,

$$P(X \leq x, Y > y) = (F_X(x)^{-\alpha} + S_Y(y)^{-\alpha} - 1)^{-1/\alpha},$$

where $F_X(x)$ is the cumulative distribution function of $X$. For Frank copula,

$$P(X \leq x, Y > y) = \log_\alpha \left(1 + \frac{(\alpha F_X(x) - 1)(\alpha S_Y(y) - 1)}{\alpha - 1}\right).$$
Figure 1. Power of $T^2_s$ under semi-competing risks data. The first plot: Clayton copula with $H_0 : \tau_c = 0.5$; the second plot: Frank copula with $H_0 : \tau_c = 0.5$; the third plot: Frank copula with $H_0 : \tau_c = -0.5$; the fourth plot: beta frailty model with $H_0 : \tau_c = 0.3$. 
Table 2. Estimation of $\tau_c$ under semi-competing data for Frank copula.

| $\tau_c$ | 0.1  | 0.2  | 0.3  | 0.4  | 0.5  | 0.6  | 0.7  | 0.8  |
|----------|------|------|------|------|------|------|------|------|
| mean     | 0.0997 | 0.1995 | 0.2972 | 0.4012 | 0.5004 | 0.6005 | 0.6995 | 0.8003 |
| Bias     | -0.0003 | -0.0005 | -0.0028 | 0.0012 | 0.0004 | 0.0005 | -0.0005 | 0.0003 |
| SE       | 0.0891 | 0.0799 | 0.0784 | 0.0656 | 0.0581 | 0.0440 | 0.0357 | 0.0241 |
| SEj      | 0.0888 | 0.0842 | 0.0776 | 0.0688 | 0.0590 | 0.0484 | 0.0369 | 0.0256 |
| MSE      | 0.0079 | 0.0071 | 0.0060 | 0.0047 | 0.0035 | 0.0023 | 0.0014 | 0.0007 |
| C.P.     | 0.952 | 0.958 | 0.936 | 0.950 | 0.950 | 0.950 | 0.950 | 0.950 |
| CenX     | 40.38 | 39.25 | 38.18 | 36.05 | 33.15 | 29.82 | 25.1  | 19.05 |
| CenY     | 16.70 | 16.65 | 16.71 | 16.49 | 16.70 | 16.65 | 16.61 | 16.44 |

Notes: ‘mean’: the average of $\hat{\tau}_c$; ‘Bias’: the average bias; ‘SE’: the empirical standard deviation; ‘SEj’: the mean of standard deviation estimator by the jackknife method; ‘MSE’: the mean square error; ‘C.P.’: the coverage probability of nominal 95% confidence interval; ‘CenX’: the censoring percentage of X; ‘CenY’: the censoring percentage of Y. Sample size $n = 100$, replications: 1000.

For frailty model,

$$P(X \leq x, Y > y) = \int (F_X(x))'(S_Y(y))'f_r(r)\, dr,$$

where $r \sim$ Beta($a, b$). Let X be an exponential distribution with mean 0.6, Y be an exponential distribution with mean 1. For Clayton and Frank copulas, let D be a uniform distribution on (0, 4), and let D be a uniform distribution on (0, 6) for the beta frailty model. Furthermore, let sample size be 100 and the number of replications be 1000.

For estimations, set $\tau_c$ from $-0.8$ to $-0.1$ for Clayton copula, from $-0.8$ to 0.8 for Frank copula, and from $-0.5$ to $-0.1$ for the beta frailty model. The results are presented in Tables 3 and 4.

For testing, let $\tau_0 = -0.5$ for Clayton copula; $\tau_0 = 0.5$ and $-0.5$ for Frank copula; and $\tau_0 = -0.3$ for the beta frailty model. The results are shown in Figure 2. Similarly, the biases and mean squared errors are very small and the mean of standard deviation by jackknife is close to the empirical standard deviation. Furthermore, the coverage probability of the 95% C.I. is near 95%, and the test statistic performs well.

4.3 Testing for quasi-independence under truncated data

When $X$ and $Y$ are quasi-independent, it implies $\tau_c = 0$. Thus, we test $H_0 : \tau_c = 0$ instead of $H'_0 : X \perp_Q Y$. In this subsection, we compare our test statistic, Equation (7), Tsai’s test statistic, and Martin and Betensky’s test statistic for testing quasi-independence.

Based on the same setting in Section 4.2, the simulation results are shown in Figure 3. From the Figure 3, it reveal that, for Clayton copula, the power of our test statistic is close to Tsai’s, and slightly more powerful than Martin’s. For Frank copula, Tsai’s power is slightly higher when $\tau_c$ is negative. For the beta frailty model, our test statistic is slightly more powerful than Tsai’s and Martin’s. Furthermore, we present the empirical sizes of the three tests under the null hypothesis.
Figure 2. Power of $T^2$ under truncated data. The first plot: Clayton copula with $H_0 : \tau_c = -0.5$; the second plot: Frank copula with $H_0 : \tau_c = 0.5$; the third plot: Frank copula with $H_0 : \tau_c = -0.5$; the fourth plot: beta frailty model with $H_0 : \tau_c = -0.3$. 
Table 3. Estimation of $\tau_c$ under truncated data for Clayton copula and the beta frailty model.

| $\tau_c$ | 0.1   | 0.2   | 0.3   | 0.4   | 0.5   | 0.6   | 0.7   | 0.8   |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|
| Clayton  |       |       |       |       |       |       |       |       |
| mean     | -0.1012 | -0.1962 | -0.3031 | -0.3985 | -0.4994 | -0.5989 | -0.7021 | -0.8006 |
| Bias     | -0.0012 | 0.0038 | -0.0031 | 0.0015 | 0.0006 | 0.0011 | -0.0021 | -0.0006 |
| SE       | 0.0742 | 0.0761 | 0.0752 | 0.0703 | 0.0623 | 0.0556 | 0.0440 | 0.0329  |
| SE$_j$   | 0.0770 | 0.0778 | 0.0762 | 0.0710 | 0.0651 | 0.0564 | 0.0457 | 0.0332  |
| MSE      | 0.0059 | 0.0061 | 0.0058 | 0.0050 | 0.0042 | 0.0032 | 0.0021 | 0.0011  |
| C.P.     | 0.954  | 0.948  | 0.954  | 0.949  | 0.954  | 0.949  | 0.947  | 0.943   |
| Cen$_Y$  | 25.77  | 27.25  | 29.04  | 31.02  | 31.30  | 32.07  | 31.65  |         |
| Frailty  |       |       |       |       |       |       |       |       |
| mean     | -0.1010 | -0.2022 | -0.3008 | -0.3961 | -0.4990 |         |         |         |
| Bias     | -0.0010 | -0.0022 | -0.0008 | 0.0009 | 0.0010 |         |         |         |
| SE       | 0.0768 | 0.0802 | 0.0809 | 0.0841 | 0.0842 |         |         |         |
| SE$_j$   | 0.0790 | 0.0834 | 0.0860 | 0.0868 | 0.0883 |         |         |         |
| MSE      | 0.0062 | 0.0070 | 0.0074 | 0.0078 | 0.0085 |         |         |         |
| C.P.     | 0.956  | 0.955  | 0.954  | 0.951  | 0.950  |         |         |         |
| Cen$_Y$  | 28.97  | 37.86  | 45.91  | 53.41  | 60.95  |         |         |         |

Notes: ‘mean’: the average of $\hat{\tau}_c$; ‘Bias’: the average bias; ‘SE’: the empirical standard deviation; ‘SE$_j$’: the mean of standard deviation estimator by the jackknife method; ‘MSE’: the mean square error; ‘C.P.’: the coverage probability of nominal 95% confidence interval; ‘Cen$_Y$’: the censoring percentage of $Y$. Sample size $n = 100$, replications: 1000. For the settings of $(a, b)$ in Beta$(a, b)$ are corresponding to the setting of $\tau_c$ with $b = 1$.

Table 4. Estimation of $\tau_c$ under truncated data for Frank copula.

| $\tau_c$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|
| mean     | 0.1016 | 0.2030 | 0.2988 | 0.3939 | 0.5042 | 0.6031 | 0.6999 | 0.8010 |
| Bias     | 0.0016 | 0.0030 | -0.0012 | -0.0061 | 0.0042 | 0.0031 | -0.0001 | 0.0010 |
| SE       | 0.0692 | 0.0633 | 0.0601 | 0.0588 | 0.0517 | 0.0459 | 0.0382 | 0.0288 |
| SE$_j$   | 0.0711 | 0.0673 | 0.0647 | 0.0621 | 0.0568 | 0.0508 | 0.0435 | 0.0337 |
| MSE      | 0.0051 | 0.0045 | 0.0042 | 0.0039 | 0.0032 | 0.0026 | 0.0019 | 0.0011 |
| C.P.     | 0.945 | 0.954 | 0.961 | 0.965 | 0.954 | 0.959 | 0.965 | 0.970 |
| Cen$_Y$  | 20.40 | 15.52 | 12.71 | 11.59 | 11.07 | 10.72 | 10.27 | 10.15 |

Notes: ‘mean’: the average of $\hat{\tau}_c$; ‘Bias’: the average bias; ‘SE’: the empirical standard deviation; ‘SE$_j$’: the mean of standard deviation estimator by the jackknife method; ‘MSE’: the mean square error; ‘C.P.’: the coverage probability of nominal 95% confidence interval; ‘Cen$_Y$’: the censoring percentage of $Y$. Sample size $n = 100$, replications: 1000.

with $\alpha = 0.05$ in Table 5. We run the simulation three times and all the empirical sizes are close to 0.05.

5. Data analysis

5.1 Bone marrow transplants data

In this subsection, we apply our method to the bone marrow transplants data that were given in [5, p. 464]. There were 137 leukemia patients receiving bone marrow transplants. Let $X$ be the
Figure 3. Power of our test, Tsai’s test and Martin’s test for $H_0: \tau_c = 0$ under truncated data. The first plot: Clayton copula; the second plot: Frank copula; the third plot: beta frailty model.
Table 5. The empirical sizes of the three tests under the null hypothesis with $\alpha = 0.05$.

|                  | Our test | Tsai’s test | Martin’s test |
|------------------|----------|-------------|---------------|
| Empirical sizes  | 0.055    | 0.062       | 0.058         |
| Empirical sizes  | 0.056    | 0.048       | 0.050         |
| Empirical sizes  | 0.053    | 0.060       | 0.055         |

Notes: Sample size $n = 100$, replications: 1000.

Table 6. Estimation of $\tau_c$ and $p$-value.

|                  | G1       | G2       | G3       | G4       |
|------------------|----------|----------|----------|----------|
| BMT data         | $\hat{\tau}_c$ | 0.771    | 0.742    | 0.752    | 0.799    |
|                  | $SD_j$   | 0.092    | 0.122    | 0.092    | 0.044    |
|                  | $T^2_s$  | 70.824   | 37.117   | 67.340   | 328.443  |
|                  | $p$-value| < 0.001  | < 0.001  | < 0.001  | < 0.001  |
|                  |          | all      | male     | female   |          |
| CHR data         | $\hat{\tau}_t$ | 0.075    | 0.185    | 0.042    |
|                  | $SD_j$   | 0.047    | 0.091    | 0.056    |
|                  | $T^2_t$  | 2.491    | 4.125    | 0.569    |
|                  | $p$-value| 0.114    | 0.042    | 0.450    |
|                  |          | all      | male     | female   |
|                  | $p$-value (by Tsai) | 0.080    | 0.043    | 0.357    |
|                  |          |          |          |          |
|                  | $p$-value (by Martin) | 0.069    | 0.040    | 0.331    |

Notes: ‘BMT data’: Bone Marrow Transplants Data; ‘CHR data’: Channing House Retirement Community Data.

time of leukemia relapse, which is a non-terminal event time, $Y$ be the time of death, which is a terminal event time and $C$ be the time to end of study. Note that $X$ may be censored by $Y$, but not vice versa. Thus, these data are a semi-competing risks data and the observable variables are $S = \min(X, Y, C)$, $T = \min(Y, C)$, $\delta_X = I(X \leq Y \wedge C)$, and $\delta_Y = I(Y \leq C)$. These patients can be divided into three groups. The first group is the acute lymphoblastic leukemia (ALL) group with 38 patients. The second group is the acute myelogenous leukemia (AML) low-risk group with 54 patients. The third group is the AML high-risk group with 45 patients. We estimate $\tau_c$ and test $H_0 : \tau_c = 0$ for the three groups and all patients by our suggested approach. Let ALL group be G1, AML low-risk group be G2, AML high risk group be G3 and all patients be G4. The results are shown in the upper half of Table 6. From the results, all the $p$-values are close to zero. Therefore, $X$ and $Y$ are not quasi-independent for G1–G4. Furthermore, all of $\hat{\tau}_c$ are greater than 0.7. It means the associations between $X$ and $Y$ for G1–G4 are positive and higher under comparable set. Thus, we can conjecture that the association of $(X, Y)$ would be positive and high.

5.2 Channing house retirement community data

We also apply our method to Channing house retirement community data that were given in [5, p. 480]. There were 97 males and 365 females who were in residence during the period January 1964 to July 1975. Let $X$ be the entry age in months, $Y$ be the death age in months and $C$ be the age at study end in months. Note that we can observe a pair $(X, Y)$ only if $X < Y$. Thus, these data are a truncated data and the observable variables are $T = \min(Y, C)$ and $\delta = I(Y \leq C)$. By the proposed method, we estimate $\tau_c$ and test $H_0 : \tau_c = 0$ for these data. The results are shown in the lower half of Table 6. From the results, all of $\hat{\tau}_c$ are between 0 and 0.2, then we can say that the associations of $(X, Y)$ under comparable set for the three cases are positive, but lower.
For all people and females, the $p$-values are greater than 0.05, thus it is not significant from zero. Therefore, we cannot reject that $X$ and $Y$ are quasi-independent. For males, the $p$-value is smaller than 0.05, which is significant from 0 according to 0.05 significant level. Furthermore, the conclusion coincides with Tsai’s and Martin’s test statistic presented in the below of Table 6.

6. Concluding remarks

We follow Tsai’s definition of conditional Kendall’s tau, $\tau_c$, under truncated data. By the same way, we can define $\tau_c$ under semi-competing risks data. In this article, we consider the estimation and test of conditional Kendall’s tau, $\tau_c$, under semi-competing risks data and truncated data. We provide a consistent estimator, $\hat{\tau}_c^*(\tau_0)$, for semi-competing risks data and truncated data, which applies the IPCW technique to correct the bias. Furthermore, we use the proposed estimator of $\tau_c$ to construct a test statistic. Then, we can test $H_0: \tau_c = \tau_0$, where $\tau_0 \in (-1, 1)$. When $\tau_0 = 0$, the test is a proxy for quasi-independence. Tsai [10] and Martin and Betensky [8] provided test statistics for quasi-independence. We compare the three test statistics for three different dependence structures, such as Clayton copula, Frank copula, and the beta frailty model in simulation studies. From the results, we have the conclusion that there is no the best test statistic among the three test statistics.

We also provide the large sample properties for our proposed estimator, which also help us to construct the null distribution of the suggested test statistic. According to the results in simulation studies, it shows that the performance of our proposed estimator and test statistic works well. Finally, we apply our proposed method to analyze two data sets. For bone marrow transplants data, we discuss four cases and conclude that there is positive and high association between the time of leukemia relapse and the time of death under comparable set for each case. Moreover, it also concludes that the two event times are not quasi-independent. For Channing house retirement community data, the association between the entry age and the death age are positive and low under comparable set for three cases.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This paper was financially supported by the National Science Council of Taiwan [NSC99-2118-M-194-001].

Supplemental data

The appendix referenced in Section 3.2 is available under the paper information at doi:10.1080/02664763.2015.1004624.

References

[1] L.L. Chaieb, L.P. Rivest, and B. Abdous, Estimating survival under a dependent truncation, Biometrika 93(3) (2006), pp. 655–669.
[2] J.P. Fine, H. Jiang, and R. Chappell, On semi-competing risks data, Biometrika 88(4) (2001), pp. 907–919.
[3] D.G. Horvitz and D.J. Thompson, A generalization of sampling without replacement from a finite universe, J. Amer. Statist. Assoc. 47(260) (1952), pp. 663–685.
[4] M.G. Kendall, A new measure of rank correlation, Biometrika 30(1–2) (1938), pp. 81–93.
[5] J.P. Klein and M.L. Moeschberger, Survival Analysis: Techniques for Censored and Truncated, 2nd ed., Springer, New York, 2003.
[6] L. Lakhal, L.P. Rivest, and B. Abdous, Estimating survival and association in a semicompeting risks model, Biometrics 64(1) (2008), pp. 180–188.

[7] L. Lakhal, L.P. Rivest, and D. Beaudoin, IPCW estimator for Kendall’s tau under bivariate censoring, Int. J. Biostatistics 5 (2009), pp. 1–20.

[8] E.C. Martin and R.A. Betensky, Testing Quasi-independence of failure and truncation times via conditional Kendall’s tau, J. Amer. Statist. Assoc. 100(470) (2005), pp. 484–492.

[9] J. Shao and D. Tu, The Jackknife and Bootstrap, Springer, New York, 1995.

[10] W.-Y. Tsai, Testing the assumption of independence of truncation time and failure time, Biometrika 77(1) (1990), pp. 169–177.

[11] W. Wang, Estimating the association parameter for copula models under dependent censoring, J. R. Statist. Soc. Ser. B (Statist. Methodol.) 65(1) (2003), pp. 257–273.