

Information matrix equivalence in the presence of censoring: A goodness-of-fit test for semiparametric copula models with multivariate survival data

Qian M. Zhou

Department of Mathematics and Statistics, Mississippi State University,
MS, USA, qz70@msstate.edu

Abstract

Various goodness-of-fit tests are designed based on the so-called *information matrix equivalence*: if the assumed model is correctly specified, two information matrices that are derived from the likelihood function are equivalent. In the literature, this principle has been established for the likelihood function with fully observed data, but it has not been verified under the likelihood for censored data. In this manuscript, we prove the information matrix equivalence in the framework of semiparametric copula models for multivariate censored survival data. Based on this equivalence, we propose an information ratio (IR) test for the specification of the copula function. The IR statistic is constructed via comparing consistent estimates of the two information matrices. We derive the asymptotic distribution of the IR statistic and propose a parametric bootstrap procedure for the finite-sample P -value calculation. The performance of the IR test is investigated via a simulation study and a real data example.

Key words: blanket test, copula selection, in-and-out-of-sample pseudo likelihood ratio test, omnibus test, parametric bootstrap.

1 Introduction

As a graduate student, one learned an important derivation about the likelihood method: assume a random variable X has a distribution function $f(x; \theta)$ (probability density function or probability mass function) with a p -dimensional parameter θ . Under certain regularity conditions (White 1982) assumed on $f(x; \theta)$, we have the following equation:

$$-\int \frac{\partial^2 \log f(x; \theta)}{\partial \theta \partial \theta'} f(x; \theta) dx = \int \left[\frac{\partial \log f(x; \theta)}{\partial \theta} \right] \left[\frac{\partial \log f(x; \theta)}{\partial \theta} \right]' f(x; \theta) dx. \quad (1)$$

When $f(x; \theta)$ is the true data generating mechanism of X , the left-side of Equation (1) can be expressed as a $p \times p$ matrix

$$\mathbb{E}^0 \left[-\frac{\partial^2 \ell(\theta; X)}{\partial \theta \partial \theta'} \right] \triangleq \mathbf{S}^*(\theta),$$

where $\ell(\theta; x) = \log f(x; \theta)$ is the log-likelihood function, and $\mathbb{E}^0$ denotes the expectation with respect to (w.r.t.) the true distribution of X . This matrix is referred to as the Fisher information matrix, or the *sensitivity matrix*. The right-side of Equation (1) can be expressed as another $p \times p$ matrix

$$\mathbb{E}^0 \left\{ \left[\frac{\partial \ell(\theta; X)}{\partial \theta} \right] \left[\frac{\partial \ell(\theta; X)}{\partial \theta} \right]' \right\} \triangleq \mathbf{V}^*(\theta),$$

called the *variability matrix* (Varin et al. 2011). Equation (1) becomes $\mathbf{S}^*(\theta) = \mathbf{V}^*(\theta)$, which is referred to as the second Bartlett identity (Bartlett 1953a;b) or information matrix equivalence (White 1982).

Several goodness-of-fit (GoF) tests for detecting model misspecification were designed through comparing these two information matrices. White (1982) proposed an information matrix (IM) test based on the elements of $\mathbf{V}^*(\theta) - \mathbf{S}^*(\theta)$. Zhou et al. (2012) proposed an information ratio test by comparing $\mathbf{S}^*(\theta)^{-1} \mathbf{V}^*(\theta)$ with a p -dimensional identity matrix. Golden et al. (2013) and Golden et al. (2016) extended these two comparisons to a general framework, called generalized IM test, which covers a range of comparison forms. Later, these tests were applied to copula models for multivariate random variables (Huang and Prokhorov 2014, Zhang et al. 2016, Prokhorov et al. 2019).

Copulas have been a popular tool for modeling the dependence structure of multivariate data, such as multivariate time series (Chen and Fan 2006a;b) and multivariate survival times (Clayton 1978, Hougaard 1986, Oakes 1989, Shih and Louis 1995). In this manuscript, we are interested in a class of semiparametric survival copula models for multivariate survival times, denoted by $(T_1, T_2, \dots, T_d)$. The multivariate survival times can be times to different types of events collected on each subject, such as time to relapse or second cancer and time to cardiovascular disease among breast cancer survivors (Li et al. 2020). Or they are times to the same type of event from different individuals within a cluster, such as the survival times of acute lymphoblastic leukemia patients from 104 institutions (Othus and Li 2010).

A survival copula specifies the joint survival function $H(t_1, \dots, t_d) = \Pr(T_1 > t_1, \dots, T_d > t_d)$ as

$$H(t_1, \dots, t_d) = \mathbb{C}(H_1(t_1), \dots, H_d(t_d); \theta), \quad (2)$$

where $H_r(t) = \Pr(T_r > t)$, $r = 1, \dots, d$, are the marginal survival functions of individual survival times, and $\mathbb{C}(u_1, u_2, \dots, u_d; \theta) : [0, 1]^d \rightarrow [0, 1]$ is a copula function with a p -dimensional parameter θ . Copulas were originally proposed for modeling the joint cumulative distribution function (CDF) of multivariate random variables, and its properties have been extensively studied (Mikosch 2006, Nelsen 2007). The

way that the survival copula relates the joint survival function to marginal survival functions is completely analogous to the way that the original copula connects the joint CDF to marginal CDFs. Thus, the survival copula satisfies the properties of the original copula (Georges et al. 2001, Nelsen 2007).

Copulas enjoy the flexibility in coupling different marginal distributions with a wide variety of copula families that exhibit different dependence structures. A class of *semiparametric* copula models assumes a parametric form for the copula function but leaves the marginal distributions unspecified. Thus, a crucial element in such a model is the specification of the copula function. Archimedean copula families, such as Clayton, Frank, and Joe, are the most popular choices (Nelsen 2006). Li et al. (2008) and Othman and Li (2010) considered the Gaussian copula, which belongs to the elliptical families (including Gaussian and t copulas). Different copulas families display different features. For example, in terms of the tail dependence, Clayton has a lower-tail dependence; Joe has an upper-tail dependence; both Gaussian and Frank have no dependence for either lower-tail or upper-tail. Misspecification of the copula function can lead to incorrect estimation of the joint distribution as well as its derivatives, such as conditional distributions.

The above-mentioned IM-based GoF tests (Huang and Prokhorov 2014, Zhang et al. 2016, Prokhorov et al. 2019) were proposed for detecting misspecification of the copula function under a semiparametric copula model. They can be regarded as the *blanket* tests introduced in Genest et al. (2009): they can be applied to any copula families and do not require selection of smoothing parameters, weight functions, or kernel functions. However, they were designed based on the presumption that the data are fully observed, and may not be applicable for data with missing values. For example, survival times can be missing due to censoring, such as the termination of the follow-up or participants being lost to follow-up.

For censored survival times, several copula specification tests were proposed, but most are limited to Archimedean families by using their unique properties. For example, Shih (1998) and Emura et al. (2010) designed their test statistics using the cross-ratio function expressed as a function of the joint survival. The test statistics in Wang and Wells (2000), Wang (2010), and Lakhal-Chaieb (2010) used the Kendall distribution, expressed in terms of the generator function. By contrast, Yilmaz and Lawless (2011) and Lin and Wu (2020) proposed tests for *any* copula families while imposing assumptions on the form of copulas under the alternative hypothesis. For example, in Yilmaz and Lawless (2011), the null and alternative models are nested, i.e., the null is embedded in the alternative. Lin and Wu (2020) assumes a particular form for the alternative copulas. In addition, several tests, such as Shih (1998), Emura et al. (2010), and Andersen et al. (2005), require the choice of a weight function or bandwidth, or the partition of the data. According to Genest et al. (2009), they are not blanket tests.

Our goal is to propose a blanket test for multivariate censored survival times, and we adopt the information ratio (IR) test originally proposed in Zhou et al. (2012). First, the IR test can be applied to all parametric copula families. Second, it is likelihood-based and depends solely on the parametric form of the null copula. Thus, it does not impose any assumptions on the alternative copulas. Third, it does not require any smoothing parameters, weight functions, kernel functions, or partition of the

data. However, the first problem we encountered was whether the information matrix equivalence, the foundation of the IR test, still holds under the likelihood for censored data. No existing work has verified it. Thus, our first contribution is to prove this equivalence in the presence of censoring.

The IR test was first proposed under the quasi-likelihood for cross-sectional or longitudinal data (Zhou et al. 2012). Later, this test was extended to various models for univariate and multivariate time series (Zhang et al. 2012; 2016; 2019; 2021). The asymptotic properties of the IR statistic have been investigated for the above settings where data are fully observed, but not for censored data yet. Another contribution of our manuscript is to derive its asymptotic properties when the marginal distributions and the copula parameters are estimated in the presence of censoring.

In this paper, we will show that if the copula function is correctly specified, the IR statistic is asymptotically distributed as a normal random variable. However, the expression of its asymptotic variance is complicated, so it is difficult to use an analytic variance estimate to calculate P -values. Thus, we propose a bootstrap procedure to approximate the statistic's null distribution via generating replications of multivariate censored data from the null copula.

Zhang et al. (2016) established the asymptotic equivalence between the IR statistic and an in-and-out-of-sample pseudo (PIOS) likelihood ratio test statistic. The PIOS statistic is based on the comparison between two types of pseudo likelihood: the *in-sample* likelihood, which is the full likelihood, and the *out-of-sample* likelihood, which is a "leave-one-out" cross-validated likelihood. In this manuscript, we will prove the asymptotic equivalence between these two test statistics with censored data. We created an R package called `IRtests` that implements both tests for copula specification with bivariate censored data, and it is available at <https://github.com/michellezhou2009/IRtests>.

The remainder of the manuscript is organized as follows. In Section 2, we prove the information matrix equivalence under a semiparametric copula model for censored survival times. We define the IR statistic in Section 3 and discuss its asymptotic properties. In Section 4, we describe how to calculate P -values via bootstrap resampling and how to use the P -values to select the best copula family. Section 5 presents the simulation studies for investigating the finite-sample performance of the proposed IR test and comparing it with the other two forms of generalized IM tests. In Section 6, we apply the IR test to a data example. Concluding remarks are given in Section 7.

2 Information Matrix Equivalence in the Presence of Censoring

For ease of illustration, we present the proposed methods in the context of bivariate event times (T_1, T_2) . We denote the true marginal survival functions by $H_r^0(t)$ with the probability density function $f_r^0(t) = -dH_r^0(t)/dt$. We assume a copula model $\mathbb{C}(u_1, u_2; \theta)$ in Equation (2) for the joint survival function of (T_1, T_2) with $\theta \in \Theta \subset R^p$. To differentiate from this *assumed* copula, we denote the *true* copula by $\mathbb{C}^0(u_1, u_2)$. According to Sklar (1959), for a continuous random vector (T_1, T_2) , there exists a unique

copula function $\mathbb{C}^0(u_1, u_2)$ such that $Pr(T_1 > t_1, T_2 > t_2) = \mathbb{C}^0(H_1^0(t_1), H_2^0(t_2))$ for all (t_1, t_2) .

Definition 1 The assumed copula $\mathbb{C}(u_1, u_2; \theta)$ is said to be **correctly specified**, denoted as $\mathbb{C}^0(u_1, u_2) \in \mathcal{C}_\theta = \{\mathbb{C}(u_1, u_2; \theta), \theta \in \Theta\}$, if there exists $\theta_0 \in \Theta$ such that $\mathbb{C}(u_1, u_2; \theta_0) = \mathbb{C}^0(u_1, u_2)$ for all $(u_1, u_2) \in (0, 1)^2$. The value θ_0 is called the true value of the copula parameter. On the other hand, if for any $\theta \in \Theta$, there exists some $(u_1, u_2) \in (0, 1)^2$ such that $\mathbb{C}(u_1, u_2; \theta) \neq \mathbb{C}^0(u_1, u_2)$, we say that the assumed copula $\mathbb{C}(u_1, u_2; \theta)$ is **misspecified**.

In the remainder of the manuscript, we let g_θ and $g_{\theta\theta}$ denote the first-order and second-order partial derivatives of a function g w.r.t. θ .

2.1 Likelihood Function

As mentioned earlier, survival times (T_1, T_2) are often subject to censoring. Let (C_1, C_2) denote the bivariate censoring times. We assume independent censoring, i.e., (C_1, C_2) are independent of (T_1, T_2) . The observed variables include

$$X_r = \min\{T_r, C_r\}, \text{ and } \delta_r = I(T_r \leq C_r), r = 1, 2, \quad (3)$$

where $I(\cdot)$ is the identity function. Note that in some situations both event times are subject to a common censoring time, i.e., $C_1 = C_2$.

Under a semiparametric copula model, the parameters consist of the unspecified marginal survival functions and the copula parameter. Since our focus is the specification of the copula function, we regard θ as the parameter of interest and marginal survival functions as nuisance parameters. For now, let us assume that the true marginal survival functions H_r^0 , $r = 1, 2$, are known. Thus, given $(X_1, X_2, \delta_1, \delta_2)$, the log-likelihood under the assumed copula is a function of the copula parameter θ . It can be written as the sum of two components: $\ell(\theta) = \mathfrak{C}(\theta, U_1^0, U_2^0, \delta_1, \delta_2) + \mathfrak{F}(X_1, X_2, \delta_1, \delta_2)$, where $\mathfrak{C}$ is a function of the assumed copula on (U_1^0, U_2^0) with $U_r^0 = H_r^0(X_r)$, $r = 1, 2$. The second term $\mathfrak{F} = \delta_1 \log f_1^0(X_1) + \delta_2 \log f_2^0(X_2)$ is a function of marginal densities only, which can be regarded as a constant. Thus, the log-likelihood can be defined as

$$\begin{aligned} \ell(\theta; U_1^0, U_2^0, \delta_1, \delta_2) &= \mathfrak{C}(\theta, U_1^0, U_2^0, \delta_1, \delta_2) \\ &= \delta_1 \delta_2 \log \mathbb{C}(U_1^0, U_2^0; \theta) + \delta_1 (1 - \delta_2) \log \mathbb{C}_1(U_1^0, U_2^0; \theta) \\ &\quad + (1 - \delta_1) \delta_2 \log \mathbb{C}_2(U_1^0, U_2^0; \theta) + (1 - \delta_1)(1 - \delta_2) \log \mathbb{C}(U_1^0, U_2^0; \theta) \end{aligned} \quad (4)$$

with $\mathbb{C}_r(u_1, u_2; \theta) = \frac{\partial \mathbb{C}(u_1, u_2; \theta)}{\partial u_r}$ for $r = 1, 2$ and $\mathbb{C}(u_1, u_2; \theta) = \frac{\partial^2 \mathbb{C}(u_1, u_2; \theta)}{\partial u_1 \partial u_2}$.

Remark 1 The survival copulas $\mathbb{C}$ and $\mathbb{C}^0$ can be regarded as the assumed and true joint CDF for (Y_1, Y_2) with $Y_r = H_r^0(T_r)$, $r = 1, 2$, which are uniformly distributed on $(0, 1)$. The above log-likelihood function is also the log-likelihood function for data $(U_1^0, U_2^0, \delta_1, \delta_2)$, where U_r^0 and δ_r can be expressed as $U_r^0 = \max\{Y_r, H_r^0(C_r)\}$ and $\delta_r = I(Y_r \geq H_r^0(C_r))$.

Remark 2 If the bivariate event times (T_1, T_2) are fully observed, i.e., $\delta_1 = \delta_2 \equiv 1$,

$$\ell(\theta; U_1^0, U_2^0, \delta_1, \delta_2) = \log \mathbb{C}(U_1^0, U_2^0; \theta). \quad (5)$$

Definition 2 Given the log-likelihood function in Equation (4), we define

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta} \in \Theta} \mathbb{E}^0 [\ell(\boldsymbol{\theta}; U_1^0, U_2^0, \delta_1, \delta_2)], \quad (6)$$

as the pseudo-true value of the parameter $\boldsymbol{\theta}$, where $\mathbb{E}^0$ takes the expectation w.r.t. the true distributions of (T_1, T_2) and (C_1, C_2) .

If the assumed copula is correctly specified, the pseudo-true value $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ (Shih and Louis 1995, Chen et al. 2010); if the assumed copula is misspecified, $\boldsymbol{\theta}^*$ might not be equal to $\boldsymbol{\theta}_0$.

2.2 Information Matrix Equivalence

Given the above log-likelihood function in Equation (4), the sensitivity and variability information matrices, both $p \times p$ dimensional, are defined as

$$\mathbf{S}^*(\boldsymbol{\theta}) = \mathbb{E}^0 [-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}] \text{ and } \mathbf{V}^*(\boldsymbol{\theta}) = \mathbb{E}^0 [\ell_{\boldsymbol{\theta}} \ell_{\boldsymbol{\theta}}']. \quad (7)$$

In the supplementary material, we provide expressions of $\ell_{\boldsymbol{\theta}}$ and $\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}$ for four copula families: Clayton, Frank, Joe, and Gaussian.

The proof of the information matrix equivalence in Theorem 1 requires the following regularity conditions R1 - R6. First, we introduce all the required notation. Let $\|x\|$ denote the usual Euclidean metric of any p -dimensional vector $x = (x_1, \dots, x_p)$, i.e., $\|x\| = \sqrt{x_1^2 + \dots + x_p^2}$. For a $p \times p$ matrix A , define $\|A\| = \sqrt{\sum_{j,k=1}^p a_{jk}^2}$, where a_{jk} is the (j, k) -th element of A . For simplicity, in the remaining of the manuscript, we suppress δ_1 and δ_2 from the log-likelihood function $\ell(\boldsymbol{\theta}, u_1, u_2, \delta_1, \delta_2)$ as well as its partial derivatives defined as follows. For $j, k = 1, \dots, p$, let $\ell_{\theta_j} = \partial \ell / \partial \theta_j$ denote the j -th element of the $p \times 1$ vector $\ell_{\boldsymbol{\theta}}$, and let $\ell_{\theta_j \theta_k}(\boldsymbol{\theta}, u_1, u_2) = \partial^2 \ell / \partial \theta_j \partial \theta_k$ denote the (j, k) -th element of the $p \times p$ matrix $\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}, u_1, u_2)$. Define $\ell_{\theta_j, \boldsymbol{\theta}} = \partial \ell_{\theta_j} / \partial \boldsymbol{\theta}$ and $\ell_{\theta_j \theta_k, \boldsymbol{\theta}} = \partial \ell_{\theta_j \theta_k} / \partial \boldsymbol{\theta}$, both $p \times 1$ vectors. For $r = 1, 2$, let $\ell_{\theta_j, u_r} = \partial \ell_{\theta_j} / \partial u_r$ and $\ell_{\theta_j \theta_k, u_r} = \partial \ell_{\theta_j \theta_k} / \partial u_r$. Let $\ell_{\boldsymbol{\theta}, u_r}$ denote a $p \times 1$ vector with the j -th element ℓ_{θ_j, u_r} . Let $\ell_{\boldsymbol{\theta}\boldsymbol{\theta}, u_r}$ denote a $p \times p$ matrix with the (j, k) -th element $\ell_{\theta_j \theta_k, u_r}$.

Our regularity conditions are:

- R1** (i) $\{(T_{i1}, T_{i2}), i = 1, \dots, n\}$ is an independent and identically distributed (i.i.d.) sample from an unknown joint survival function $\mathbb{C}^0(H_1^0(t_1), H_2^0(t_2))$ with continuous marginal survival functions $H_r^0(\cdot), r = 1, 2$;
- (ii) $\{(C_{i1}, C_{i2}), i = 1, \dots, n\}$ is an i.i.d. sample with joint survival function $\mathcal{G}^0(t_1, t_2) = \Pr(C_{i1} > t_1, C_{i2} > t_2)$ and marginal survival functions $G_r^0(t) = \Pr(C_{ir} > t), r = 1, 2$;
- (iii) The censoring variables (C_{i1}, C_{i2}) are independent of (T_{i1}, T_{i2}) and there is no mass concentration at 0 in the sense that $G_r^0(\eta) \rightarrow 1$ as $\eta \rightarrow 0$.

R2 Let Θ be a compact space of $\mathcal{R}^p$. For every $\epsilon > 0$,

$$\liminf_{\boldsymbol{\theta} \in \Theta: \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \geq \epsilon} \mathbb{E}^0 [\ell(\boldsymbol{\theta}^*, U_1^0, U_2^0)] - \mathbb{E}^0 [\ell(\boldsymbol{\theta}, U_1^0, U_2^0)] > 0.$$

- R3** The true (unknown) copula function $\mathbb{C}^0(u_1, u_2)$ has continuous partial derivatives.
- R4** For any (u_1, u_2) , $\ell(\boldsymbol{\theta}, u_1, u_2)$ is a continuous function of $\boldsymbol{\theta} \in \Theta$.
- R5** Functions $\ell_{\theta_j}(\boldsymbol{\theta}, u_1, u_2)$, $\ell_{\theta_j, \theta_k}(\boldsymbol{\theta}, u_1, u_2)$, and $\ell_{\theta_j, u_r}(\boldsymbol{\theta}, u_1, u_2)$, $j, k = 1, \dots, p$, $r = 1, 2$, are well defined and continuous in $(u_1, u_2, \boldsymbol{\theta}) \in [0, 1]^2 \times \Theta$.
- R6** (i) $\|\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}, u_1, u_2)\| \leq q\{u_1(1-u_1)\}^{-a_1}\{u_2(1-u_2)\}^{-a_2}$ for some $q > 0$ and $a_1, a_2 \geq 0$ such that $\mathbb{E}^0[\{U_1^0(1-U_1^0)\}^{-2a_1}\{U_2^0(1-U_2^0)\}^{-2a_2}] < \infty$;
(ii) $\|\ell_{\theta, u_r}(\boldsymbol{\theta}, u_1, u_2)\| \leq q\{u_s(1-u_s)\}^{-a_s}\{u_r(1-u_r)\}^{-a_r}$ for some q, a_s, a_r , and $s \neq r$ such that $\mathbb{E}^0[\{U_s^0(1-U_s^0)\}^{\epsilon_s-a_s}\{U_r^0(1-U_r^0)\}^{-a_r}] < \infty$ for some $\epsilon_s \in (0, 1/2)$;
(iii) $\|\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}, u_1, u_2)\| \leq q\{u_1(1-u_1)\}^{-a_1}\{u_2(1-u_2)\}^{-a_2}$ for some $q > 0$ and $a_1, a_2 \geq 0$ such that $\mathbb{E}^0[\{U_1^0(1-U_1^0)\}^{-2a_1}\{U_2^0(1-U_2^0)\}^{-2a_2}] < \infty$.

These regularity conditions are similar to those listed in Shih and Louis (1995) and Chen et al. (2010), which also focused on semiparametric survival copula models for censored survival data. Shih and Louis (1995) referred them to as the standard regularity conditions for maximum likelihood estimation (White 1982) in the copula context. It is worth pointing out that our assumptions of homogenous censoring distribution (i.e., (C_{i1}, C_{i2}) follows the same joint distribution across subjects) was also imposed in Shih and Louis (1995). However, Chen et al. (2010) allowed different censoring distributions across subjects. We defer more discussions on this assumption to Section 7. In addition, these regularity conditions except for those related to censoring were used in Huang and Prokhorov (2014) to prove the information matrix equivalence under the log-likelihood function in Equation (5) for fully observed bivariate event times.

Theorem 1 (Information Matrix Equivalence) *Assume that conditions R1 - R6 hold. If the assumed copula is correctly specified, under the log-likelihood function in Equation (4), $\mathbf{S}^*(\boldsymbol{\theta}^*) = \mathbf{V}^*(\boldsymbol{\theta}^*)$, where $\mathbf{S}^*(\boldsymbol{\theta})$ and $\mathbf{V}^*(\boldsymbol{\theta})$ are the sensitivity and variability matrices defined in Equation (7), and $\boldsymbol{\theta}^*$ is the pseudo-true value of the parameter $\boldsymbol{\theta}$ defined in Equation (6).*

We prove this theorem in Appendix A, where we will show that $\mathbf{S}^*(\boldsymbol{\theta}) = \mathbf{V}^*(\boldsymbol{\theta}) - \mathbf{A}(\boldsymbol{\theta})$ for any assumed copula, where $\mathbf{A}(\boldsymbol{\theta}) = \mathbb{E}_{(C_1, C_2)}^0 \left\{ \mathbb{E}_{(T_1, T_2)}^0 [\Delta | C_1, C_2] \right\}$ with

$$\mathbb{E}_{(T_1, T_2)}^0 [\Delta | C_1, C_2] = \sum_{\delta_1, \delta_2=0,1} \int \int_{\Omega_{\delta_1, \delta_2}} \mathbb{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}(u_1, u_2; \boldsymbol{\theta}) w_{\delta_1, \delta_2}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2.$$

Here, for each censoring status (δ_1, δ_2) , $\Omega_{\delta_1, \delta_2}$ is the corresponding region for (Y_1, Y_2) (Remark 1) given (C_1, C_2) with $\bigcup_{\delta_1, \delta_2=0,1} \Omega_{\delta_1, \delta_2} = [0, 1]^2$. In addition, w_{δ_1, δ_2} is a ratio of the true copula function versus the assumed copula or a ratio of their partial derivatives.

If the assumed copula is correctly specified, $w_{\delta_1, \delta_2}(u_1, u_2; \boldsymbol{\theta}^*) \equiv 1$ for all (u_1, u_2) and (δ_1, δ_2) . It leads to

$$\mathbb{E}_{(T_1, T_2)}^0 [\Delta | C_1, C_2] = \int_0^1 \int_0^1 \mathbb{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}(u_1, u_2; \boldsymbol{\theta}^*) du_1 du_2 = 0, \quad (8)$$

because of the interchangeability between integrals and derivatives ensured by the regularity conditions. Equation (8) makes $\mathbf{A}(\boldsymbol{\theta}^*) = 0$, resulting in the information matrix equivalence stated in Theorem 1. On the other hand, if the assumed copula is misspecified, $w_{\delta_1, \delta_2}(u_1, u_2; \boldsymbol{\theta}^*) \neq 1$ for some (u_1, u_2) and (δ_1, δ_2) , and thus, Equation (8) will not hold, indicating $\mathbf{S}^*(\boldsymbol{\theta}^*) \neq \mathbf{V}^*(\boldsymbol{\theta}^*)$.

We define the following *information matrix ratio*: $\mathbf{R}^*(\boldsymbol{\theta}^*) = \mathbf{S}^*(\boldsymbol{\theta}^*)^{-1}\mathbf{V}^*(\boldsymbol{\theta}^*)$. If the assumed copula is correctly specified, $\mathbf{R}^*(\boldsymbol{\theta}^*) = I_p$, a p -dimensional identity matrix; otherwise, $\mathbf{R}^*(\boldsymbol{\theta}^*) \neq I_p$. The discrepancy between $\mathbf{R}^*(\boldsymbol{\theta}^*)$ and I_p implies copula misspecification, and it can be quantified by a scalar metric: $\text{tr}[\mathbf{R}^*(\boldsymbol{\theta}^*)] - p$, where $\text{tr}(\cdot)$ denotes the trace of a matrix. This is the motivation for the IR test we will propose for detecting copula misspecification.

3 Information Ratio Statistic under Semiparametric Copula Models

In this section, we propose an IR statistic, which is a consistent estimator of $\text{tr}[\mathbf{R}^*(\boldsymbol{\theta}^*)]$ under the semiparametric copula model.

3.1 IR Statistic

As described earlier, a semiparametric copula model assumes that the *true* marginal survival functions $H_1^0(t)$ and $H_2^0(t)$ are unspecified. Thus, obtaining a consistent estimator of $\text{tr}[\mathbf{R}^*(\boldsymbol{\theta}^*)]$ requires the consistent estimation for the marginal survival functions, the copula parameter, and two information matrices. Let $\{(X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}), i = 1, \dots, n\}$ be n independent realizations of $(X_1, X_2, \delta_1, \delta_2)$.

Consistent estimation of marginal survival functions and copula parameter. Shih and Louis (1995) proposed the following two-step procedure: at the first step, each marginal survival function is estimated by a nonparametric estimator $\hat{H}_r(t)$ with data $\{(X_{ir}, \delta_{ir}), i = 1, \dots, n\}$, $r = 1, 2$. Under the assumption that the censoring times are independent of the event times, we consider the Kaplan-Meier estimator, which is a consistent estimator for the marginal survival function (Kaplan and Meier 1958). Thus, $\hat{U}_{ir} = \hat{H}_r(X_{ir})$ is a consistent estimate of $U_{ir}^0 = H_r^0(X_{ir})$, $r = 1, 2, i = 1, \dots, n$.

At the second step, the copula parameter $\boldsymbol{\theta}$ is estimated by a pseudo maximum likelihood estimator (PMLE), which maximizes the *pseudo log-likelihood function* given as $\ell_n(\boldsymbol{\theta}) = \sum_{i=1}^n \ell(\boldsymbol{\theta}, \hat{U}_{i1}, \hat{U}_{i2})$. Specifically, the PMLE of $\boldsymbol{\theta}$ is given as

$$\hat{\boldsymbol{\theta}}_n = \arg \max_{\boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}). \quad (9)$$

Chen et al. (2010) established the asymptotic properties of $\hat{\boldsymbol{\theta}}_n$. One of them is that, under certain conditions, the PMLE $\hat{\boldsymbol{\theta}}_n$ converges in probability to the pseudo-true value $\boldsymbol{\theta}^*$ defined in Equation (6) as $n \rightarrow \infty$, *regardless of whether the assumed copula function is correctly specified or not*. Thus, $\hat{\boldsymbol{\theta}}_n$ is a consistent estimate of $\boldsymbol{\theta}^*$.

Consistent estimation of information matrices. By Equation (7), for a given value θ , $\mathbf{S}^*(\theta)$ and $\mathbf{V}^*(\theta)$ are the distributional means. If the true marginal survival functions are known, these two information matrices can be consistently estimated by the empirical means:

$$-n^{-1} \sum_{i=1}^n \ell_{\theta\theta}(\theta; U_{i1}^0, U_{i2}^0) \text{ and } n^{-1} \sum_{i=1}^n \ell_{\theta}(\theta; U_{i1}^0, U_{i2}^0) \ell_{\theta}(\theta; U_{i1}^0, U_{i2}^0)'.$$

Chen et al. (2010) provided the following consistent estimator for $\mathbf{S}^*(\theta^*)$:

$$\hat{\mathbf{S}}_n(\hat{\theta}_n) = -\frac{1}{n} \sum_{i=1}^n \ell_{\theta\theta}(\hat{\theta}_n; \hat{U}_{i1}, \hat{U}_{i2}),$$

where U_{ir}^0 is estimated by $\hat{U}_{ir}$, and θ^* is estimated by the PMLE $\hat{\theta}_n$. Following the same idea, we propose the following consistent estimator for $\mathbf{V}^*(\theta^*)$:

$$\hat{\mathbf{V}}_n(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^n \ell_{\theta}(\hat{\theta}_n; \hat{U}_{i1}, \hat{U}_{i2}) \ell_{\theta}(\hat{\theta}_n; \hat{U}_{i1}, \hat{U}_{i2})'.$$

With the above estimators, the IR statistic is defined as

$$R_n = \text{tr} \left[\hat{\mathbf{S}}_n(\hat{\theta}_n)^{-1} \hat{\mathbf{V}}_n(\hat{\theta}_n) \right].$$

Next, we will present two key asymptotic properties of this IR statistic. First, Theorem 2 establishes the consistency of R_n , i.e., it converges in probability to $\text{tr}[\mathbf{R}^*(\theta^*)]$ for *any* assumed copula. Second, Theorem 3 states the asymptotic normality of R_n if the assumed copula is correctly specified. This result will be used for designing the IR test and copula selection in Section 4.

3.2 Asymptotic Properties of IR statistic

The consistency and asymptotic normality of the IR statistic R_n requires the regularity conditions R1 - R6 listed in Section 2.2 and the following additional conditions:

C1 (i) Let $L = \sup_{\theta \in \Theta} |\ell(\theta, U_1^0, U_2^0)|$ and $L_{\theta} = \sup_{\theta \in \Theta} |\ell_{\theta}(\theta, U_1^0, U_2^0)|$. Then

$$\lim_{K \rightarrow \infty} \mathbb{E}^0 [L I(L \geq K) + L_{\theta} I(L_{\theta} \geq K)] = 0;$$

(ii) For any $\eta > 0$ and any $\epsilon > 0$, there is $K > 0$ such that $|\ell(\theta, u_1, u_2)| \leq K |\ell(\theta, u'_1, u'_2)|$ for all $\theta \in \Theta$ and all $u_r \in [\eta, 1)$ such that $1 - u_r \geq \epsilon(1 - u'_r)$, $r = 1, 2$.

C2 For $r = 1, 2$, if T_r is subject to non-trivial censoring (i.e., $C_r \neq \infty$), then the Kaplan-Meier estimator $\hat{H}_r$ is truncated at the tail in the sense that for some τ_r , $\hat{H}_r(t) = \hat{H}_r(\tau_r)$ for all $t \geq \tau_r$ and $G_r^0(\tau_r) H_r^0(\tau_r) > 0$.

C3 (i) Regularity condition R2 holds with $\theta^* \in \text{int}(\Theta^*)$, where Θ^* is a compact subset of Θ ;

- (ii) $\mathbb{E}^0 [-\ell_{\theta\theta}(\theta^*, U_1^0, U_2^0)]$ has all its eigenvalues bounded below and above by some finite positive constants;
 - (iii) $\text{Var}^0 [\ell_{\theta}(\theta^*, U_1^0, U_2^0) + W_1(\theta^*, X_1, \delta_1) + W_2(\theta^*, X_2, \delta_2)]$ has all its eigenvalues bounded below and above by some finite positive constants, where for $r = 1, 2$, $W_r(\theta^*, X_r, \delta_r)$ is defined in Equation (22) of Appendix C;
 - (iv) $\ell_{\theta}(\theta^*, U_1^0, U_2^0) + W_1(\theta^*, X_1, \delta_1) + W_2(\theta^*, X_2, \delta_2)$ satisfies Lindeberg condition.
- C4** (i) Let $L_{\theta, u_r} = \sup_{\theta \in \Theta} \|\ell_{\theta, u_r}(\theta, U_1^0, U_2^0)\|$ and $L_{\theta\theta} = \sup_{\theta \in \Theta} \|\ell_{\theta\theta}(\theta, U_1^0, U_2^0)\|$. Then

$$\lim_{K \rightarrow \infty} \mathbb{E}^0 [L_{\theta, u_r} I(L_{\theta, u_r} \geq K) + L_{\theta\theta} I(L_{\theta\theta} \geq K)] = 0;$$

- (ii) Let $Q(\theta, u_1, u_2) = \|\ell_{\theta}(\theta, u_1, u_2)\| + \|\ell_{\theta\theta}(\theta, u_1, u_2)\|$. For any $\eta > 0$ and any $\epsilon > 0$, there is $K > 0$ such that $Q(\theta, u_1, u_2) \leq KQ(\theta, u'_1, u'_2)$ for all $\theta \in \Theta$ and all $u_r \in [\eta, 1)$ such that $1 - u_r \geq \epsilon(1 - u'_r)$, $r = 1, 2$.
- C5** (i) For $j, k = 1, \dots, p$, $r = 1, 2$, let

$$LL_{\theta_j \theta_k, u_r} = \sup_{\theta \in \Theta} |\ell_{\theta_j, u_r}(\theta, U_1, U_2) \ell_{\theta_k}(\theta, U_1, U_2) + \ell_{\theta_k, u_r}(\theta, U_1, U_2) \ell_{\theta_j}(\theta, U_1, U_2)|,$$

$$LL_{\theta_j \theta_k, \theta} = \sup_{\theta \in \Theta} \|\ell_{\theta_j, \theta}(\theta, U_1, U_2) \ell_{\theta_k}(\theta, U_1, U_2) + \ell_{\theta_k, \theta}(\theta, U_1, U_2) \ell_{\theta_j}(\theta, U_1, U_2)\|.$$

Then,

$$\lim_{K \rightarrow \infty} \sup_{(j, k)} \mathbb{E}^0 [LL_{\theta_j \theta_k, u_r} I(LL_{\theta_j \theta_k, u_r} \geq K) + LL_{\theta_j \theta_k, \theta} I(LL_{\theta_j \theta_k, \theta} \geq K)] = 0;$$

- (ii) For $j, k = 1, \dots, p$, let

$$Q_{1, \theta_j \theta_k}(\theta, u_1, u_2) = |\ell_{\theta_j}(\theta, u_1, u_2) \ell_{\theta_k}(\theta, u_1, u_2)|$$

$$+ \|\ell_{\theta_j, \theta}(\theta, u_1, u_2) \ell_{\theta_k}(\theta, u_1, u_2) + \ell_{\theta_k, \theta}(\theta, u_1, u_2) \ell_{\theta_j}(\theta, u_1, u_2)\|.$$

For any $\eta > 0$ and any $\epsilon > 0$, there is $K > 0$, such that $Q_{1, \theta_j \theta_k}(\theta, u_1, u_2) \leq KQ_{1, \theta_j \theta_k}(\theta, u'_1, u'_2)$ for all $\theta \in \Theta$ and all $u_r \in [\eta, 1)$ such that $1 - u_r \geq \epsilon(1 - u'_r)$, $r = 1, 2$.

- C6** (i) Functions $\ell_{\theta_j \theta_k, \theta}(\theta, u_1, u_2)$ and $\ell_{\theta_j \theta_k, u_r}(\theta, u_1, u_2)$, $j, k = 1, \dots, p$, $r = 1, 2$, are well-defined and continuous in $(\theta, u_1, u_2) \in \Theta \times (0, 1)^2$;
- (ii) For $j, k = 1, \dots, p$, $r = 1, 2$, let $L_{\theta_j \theta_k, u_r} = \sup_{\theta \in \Theta} |\ell_{\theta_j \theta_k, u_r}(\theta, U_1^0, U_2^0)|$ and $L_{\theta_j \theta_k, \theta} = \sup_{\theta \in \Theta} \|\ell_{\theta_j \theta_k, \theta}(\theta, U_1^0, U_2^0)\|$. Then,

$$\lim_{K \rightarrow \infty} \sup_{(j, k)} \mathbb{E}^0 [L_{\theta_j \theta_k, u_r} I(L_{\theta_j \theta_k, u_r} \geq K) + L_{\theta_j \theta_k, \theta} I(L_{\theta_j \theta_k, \theta} \geq K)] = 0;$$

- (iii) For $j, k = 1, \dots, p$, let $Q_{2, \theta_j \theta_k}(\theta, u_1, u_2) = |\ell_{\theta_j \theta_k}(\theta, u_1, u_2)| + \|\ell_{\theta_j \theta_k, \theta}(\theta, u_1, u_2)\|$. For any $\eta > 0$ and any $\epsilon > 0$, there is $K > 0$, such that $Q_{2, \theta_j \theta_k}(\theta, u_1, u_2) \leq KQ_{2, \theta_j \theta_k}(\theta, u'_1, u'_2)$ for all $\theta \in \Theta$ and all $u_r \in [\eta, 1)$ such that $1 - u_r \geq \epsilon(1 - u'_r)$, $r = 1, 2$.

- C7** (i) $\|\ell_{\theta\theta, u_r}(\theta^*, u_1, u_2)\| \leq q\{u_s(1 - u_s)\}^{-a_s}\{u_r(1 - u_r)\}^{-a_r}$ for some q, a_s, a_r , and $s \neq r$ such that $\mathbb{E}^0[\{U_s^0(1 - U_s^0)\}^{\epsilon_s - a_s}\{U_r^0(1 - U_r^0)\}^{-a_r}] < \infty$ for some $\epsilon_s \in (0, 1/2)$;
- (ii) $\|\ell_{\theta}(\theta^*, u_1, u_2)\ell_{\theta}(\theta^*, u_1, u_2)'\| \leq q\{u_1(1 - u_1)\}^{-a_1}\{u_2(1 - u_2)\}^{-a_2}$ for some $q > 0$ and $a_1, a_2 \geq 0$ such that $\mathbb{E}^0[\{U_1^0(1 - U_1^0)\}^{-2a_1}\{U_2^0(1 - U_2^0)\}^{-2a_2}] < \infty$;
- (iii) $\|\ell_{\theta, u_r}(\theta^*, u_1, u_2)\ell_{\theta}(\theta^*, u_1, u_2)'\| \leq q\{u_s(1 - u_s)\}^{-a_s}\{u_r(1 - u_r)\}^{-a_r}$ for some q, a_s, a_r , and $s \neq r$ such that $\mathbb{E}^0[\{U_s^0(1 - U_s^0)\}^{\epsilon_s - a_s}\{U_r^0(1 - U_r^0)\}^{-a_r}] < \infty$ for some $\epsilon_s \in (0, 1/2)$.

Our regularity conditions R1 - R6 combined with the above conditions C1 - C4 are the conditions C1 - C5 and A1 - A4 of Chen et al. (2010) for the existence, consistency, and asymptotic normality of the PMLE $\hat{\theta}_n$.

Theorem 2 *Under conditions R1 - R6 and C1 - C5, we have $R_n \rightarrow \text{tr}[R^*(\theta^*)]$ in probability as $n \rightarrow \infty$.*

The proof of this theorem (Appendix B) requires the consistency of $\hat{S}_n(\hat{\theta}_n)$ and $\hat{V}_n(\hat{\theta}_n)$. Chen et al. (2010) has proved the consistency of $\hat{S}_n(\hat{\theta}_n)$, which requires their condition A4. We follow their arguments to prove the consistency of $\hat{V}_n(\hat{\theta}_n)$, where our conditions C5 is analogous to Chen et al.'s condition A4.

Theorem 3 *Assume conditions R1 - R6 and C1 - C7 hold. Define the null hypothesis $H_0: \mathbb{C}^0(u_1, u_2) \in \mathcal{C}_\theta = \{\mathbb{C}(u_1, u_2; \theta), \theta \in \Theta\}$, i.e., the assumed copula is correctly specified. If the null hypothesis H_0 is true, R_n converges to p in probability, and $\sqrt{n}(R_n - p)$ converges in distribution to a normal random variable with mean 0 and variance $\sigma_R^2 = \text{Var}[h_R(X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}, \theta)]$, where $h_R(X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}, \theta)$ is given by Equation (27) in Appendix C.*

The proof of this theorem (Appendix C) utilizes the Taylor expansion of $\hat{S}_n(\hat{\theta}_n)$ and $\hat{V}_n(\hat{\theta}_n)$. One step requires the consistency for the first-order derivative of $\hat{S}_n(\hat{\theta}_n)$ and of $\hat{V}_n(\hat{\theta}_n)$ w.r.t. θ . Again, we follow the arguments of Chen et al. (2010) for proving the consistency of $\hat{S}_n(\hat{\theta}_n)$, where our condition C6 (i) is analogous to Chen et al.'s Condition A2, and our conditions C6 (ii) & (iii) together are analogous to Chen et al.'s condition A4. Another component in the proof involves the expansion of the estimated pseudo-observations $\hat{U}_{ir} - U_{ir}^0$ using the asymptotic properties of the Kaplan-Meier estimator for the marginal survival functions. Our condition R6 (iii) is analogous to Chen et al.'s condition A3 (i), and so is our condition C7 (ii). Our condition C7 (i) is analogous to Chen et al.'s condition A3 (ii), and so is our condition C7 (iii). We want to point out that the expression of $h_R(X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}, \theta)$ is different from the expansion with fully observed data derived in Zhang et al. (2016; 2021). In their settings, the marginal distributions are estimated by the empirical distribution functions whose expansions are different from those of Kaplan-Meier estimators.

3.3 Asymptotic Equivalence to the In-and-Out-of-Sample Pseudo Likelihood Ratio Statistic

For semiparametric copula models with fully observed data, Zhang et al. (2016) showed that the IR statistic R_n is asymptotically equivalent to a class of in-and-out-

of-sample pseudo (PIOS) likelihood ratio test statistic. Theorem 4 below states this asymptotic equivalence still holds in the presence of censoring. The PIOS statistic is defined as a difference between two types of pseudo log-likelihood functions: in-sample and out-of-sample. Under our log-likelihood function in Equation (4), the in-sample pseudo log-likelihood is defined as $\ell_n^{in} = \sum_{i=1}^n \ell(\hat{\theta}_n, \hat{U}_{i1}, \hat{U}_{i2})$, where $\hat{\theta}_n$ is obtained from Equation (9) using all the observations. The out-of-sample pseudo log-likelihood employs the leave-one-out technique and is defined as $\ell_n^{out} = \sum_{i=1}^n \ell(\hat{\theta}_{(-i)}, \hat{U}_{i1}, \hat{U}_{i2})$, where $\hat{\theta}_{(-i)} = \arg \max_{\theta} \sum_{s=1, s \neq i}^n \ell(\theta, \hat{U}_{s1}, \hat{U}_{s2})$ is the PMLE using the data with the i -th observation deleted. The PIOS test statistic is defined as $T_n = \ell_n^{in} - \ell_n^{out}$. A large value of T_n suggests that the assumed copula model is a poor fit to the data since it is sensitive to the deletion of individual observations.

Theorem 4 *Under condition R1 - R6 and C1 - C4, $|R_n - T_n| = o_p(1)$.*

The proof is provided in Appendix D. Because of this asymptotic equivalence, if the null hypothesis H_0 is true, the PIOS statistic T_n also converges to p in probability, and $\sqrt{n}(T_n - p)$ also converges in distribution to a normal random variable with mean 0 and the same variance σ_R^2 .

4 Information Ratio Test and Copula Selection

In practice, it is challenging to calculate P -values using an analytical estimate of the asymptotic variance σ_R^2 because its expression is complicated. To address this issue, we suggest a parametric bootstrap resampling procedure for the P -value calculation. This approach is commonly employed in GoF tests, including those based on information matrix equivalence (Horowitz 1994, Dhaene and Hoorelbeke 2004, Golden et al. 2013, Huang and Prokhorov 2014, Golden et al. 2016, Prokhorov et al. 2019). Genest and Rémillard (2008) provided the validation of this procedure in the general setting of semi-parametric models.

4.1 P -value Calculation via Bootstrap Resampling

The key idea is to approximate null distribution of R_n by the test statistics values calculated from a large number of data replicates generated under the null copula (the copula family tested as the null hypothesis). These data replicates are referred to as the bootstrapped data, denoted by $\mathfrak{D}^{(b)}$; in contrast, we denote the original data by $\mathfrak{D}$. The bootstrapped data $\mathfrak{D}^{(b)}$ is obtained by generating bootstrapped resamples of the bivariate event times $(T_{i1}^{(b)}, T_{i2}^{(b)})$ and bivariate censoring times $(C_{i1}^{(b)}, C_{i2}^{(b)})$.

Generation of $(T_{i1}^{(b)}, T_{i2}^{(b)})$ under the null copula. For example, we test Clayton copula as the null hypothesis, i.e.,

$$H_0 : \mathbb{C}^0(u_1, u_2) = \mathbb{C}(u_1, u_2; \theta) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, \text{ for some } \theta > 0.$$

Let $\hat{\theta}_n$ be the PMLE of θ from the log-likelihood function using the original data $\mathfrak{D}$ and the above parametric form of Clayton copula function. First, we generate a bivariate variable $(U_{i1}^{(b)}, U_{i2}^{(b)})$ from the Clayton copula with parameter value $\hat{\theta}_n$. This step can be implemented using R function `rCopula` of `copula` package. Second, we obtain $T_{ir}^{(b)} = \hat{H}_r^{-1}(U_{ir}^{(b)})$, $r = 1, 2$, where $\hat{H}_r^{-1}$ is the inverse function of the Kaplan-Meier estimator of the marginal survival functions.

Generation of $(C_{i1}^{(b)}, C_{i2}^{(b)})$. The censoring times C_{i1} and C_{i2} might be correlated, but our method does not rely on their joint distribution. Thus, we can simulate them separately from their own marginal distributions. Under the assumption of independent censoring, the survival function $G_r(t)$ of C_{ir} can be consistently estimated by a Kaplan-Meier estimator $\hat{G}_r(t)$ using the data $\{(X_{ir}, 1 - \delta_{ir}), i = 1, \dots, n\}$. For each $r = 1, 2$, we first generate a random number $v_{ir}^{(b)}$ from a uniform distribution between 0 and 1, and then obtain $C_{ir}^{(b)} = \hat{G}_r^{-1}(v_{ir}^{(b)})$. In some cases, both event times are subject to the same censoring time, i.e., $C_{i1} = C_{i2} = C_i$, its sole survival function $G(t)$ can be estimated using the data $\{(\max\{X_{i1}, X_{i2}\}, 1 - \delta_{i1}\delta_{i2}), i = 1, \dots, n\}$.

Bootstrap resampling. The resampling procedure includes the following steps:

Step 1: Generate a bootstrapped resample of $\{(T_{i1}^{(b)}, T_{i2}^{(b)}, C_{i1}^{(b)}, C_{i2}^{(b)}), i = 1, \dots, n\}$ with the same sample size of the original data following the above description. This forms a bootstrapped data $\mathfrak{D}^{(b)} = \{(X_{i1}^{(b)}, X_{i2}^{(b)}, \delta_{i1}^{(b)}, \delta_{i2}^{(b)}), i = 1, \dots, n\}$, where $X_{ir}^{(b)} = \min\{T_{ir}^{(b)}, C_{ir}^{(b)}\}$ and $\delta_{ir}^{(b)} = I(T_{ir}^{(b)} \leq C_{ir}^{(b)})$, $r = 1, 2$.

Step 2: Based on the bootstrapped data $\mathfrak{D}^{(b)}$, calculate the test statistic, denoted as $R_n^{(b)}$, referred to as a bootstrap resample of R_n .

Step 3: Repeat Steps 1 and 2 B times, producing B bootstrap resamples $\{R_n^{(b)}, b = 1, \dots, B\}$.

The bootstrap resamples $\{\sqrt{n}(R_n^{(b)} - p), b = 1, \dots, B\}$ approximate the null distribution of $\sqrt{n}(R_n - p)$, and their sample variance approximates the asymptotic variance σ_R^2 . Thus, we calculate

$$\sigma^b = \sqrt{\frac{1}{B-1} \sum_{b=1}^B [R_n^{(b)} - \bar{R}_n^b]^2}$$

where $\bar{R}_n^b$ is the average of $\{R_n^{(b)}, b = 1, \dots, B\}$. The P -value of the IR test is

$$p\text{-value} = 2 \times \left[1 - \Phi \left(\frac{|R_n - p|}{\sigma^b} \right) \right],$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution.

If the calculated p -value is smaller than a significance level α , we reject the null hypothesis and conclude significant evidence suggesting copula misspecification. Alternatively, we can use critical values to make conclusions. Let $z_{\alpha/2}$ denote the upper

$100 * (\alpha/2)\%$ quantile of the standard normal distribution. If $\frac{|R_n - p|}{\sigma^b} > z_{\alpha/2}$, we reject the null hypothesis.

4.2 Selection of the Best Copula Family

For some data, a GoF test would fail to reject several copula families. It might be because the sample size is small or the censoring rate is high or both. As a result, the data do not contain sufficient information to reject the null hypothesis. In addition, if the level of dependence is not strong, several families appear similar, and consequently, it is more difficult for a test to tell them apart. In some situations, the underlying true dependence structure might be complicated, and any parametric copula family is merely an approximation. For these cases, we are more concerned with selecting the *best* copula family from several candidates in the sense that the data exhibit the weakest evidence against it, i.e., showing the highest agreement between the assumed copula and the data. Here, we propose using the P -value of the IR test as the selection criteria: the best copula family is the one with the largest P -value.

5 Simulation

In this section, we investigate the finite-sample performance of the proposed IR test through two simulation studies, where we consider different sample sizes, copula families with various dependence strength, and censoring rate (proportions of censored event times). The first study focuses on the null distribution of the IR statistic, i.e., the distribution when the null copula is the true copula. We compare it with the normal distribution and PIOS's null distribution. The second study examines the type I error rate and power of IR test as well as the performance of using IR's p -value for copula selection. As mentioned earlier, IR is a specific form of generalized IM tests. Thus, we compare our IR with two other forms: White test (difference between two IMs), and log IM test (difference between logarithms of two IMs).

5.1 Simulation Setting

We consider four copula families: Clayton, Frank, Joe, and Gaussian, each with a scalar copula parameter θ . The value of θ is determined by Kendall's τ coefficient, which reflects the dependence strength (Kendall 1938). The relationship between Kendall's τ and θ for each of the above copula families is described in the supplementary material.

Given a copula family $\mathbb{C}$ with a parameter value θ , we generate (T_{i1}, T_{i2}) whose marginal distributions are both exponential distribution with mean 1 and joint survival function follows the given copula family $\mathbb{C}$. For example, the copula is Clayton with parameter $\theta = 2$, corresponding to Kendall's $\tau = 0.5$. Following a similar procedure described in Section 4.1, we first generate (U_{i1}, U_{i2}) from Clayton copula with $\theta = 2$ using R function `rCopula`. Second, calculate $T_{ir} = -\log(U_{ir}), r = 1, 2$. Note that

$-\log(x)$ is the inverse function of the survival function for exponential distribution with mean 1.

In this simulation, both event times are subject to a common censoring time C_i , generated from an exponential distribution with mean 4, 3/2 or 3/7, that correspond to a censoring rate of 20%, 40%, or 70% for individual event times. In addition, we include a no-censoring setting, i.e., T_{i1} and T_{i2} are fully observed, to investigate the effect of censoring on the performance of the IR test. Thus, there are four censoring scenarios, denoted as "no-censoring", "20%-censored", "40%-censored", and "70%-censored". Figures 1 - 4 in the supplementary material plot the estimated pseudo-observations ($\hat{U}_{i1}, \hat{U}_{i2}$) obtained from one replication of the simulated bivariate censored survival times of sample size $n = 100$ or 600 generated from each of the four copula families with Kendall's $\tau = 0.3$ or 0.7.

5.2 Study I: Null Distributions of IR and PIOS statistics

In this study, we generate data from a copula family and test for the same copula family, i.e., the null copula is the true copula. Figure 5 - 8 in the supplementary material plot the normal quantile-quantile (QQ) plots of 500 replications of the IR and PIOS statistics under Clayton, Frank, Joe, or Gaussian with Kendall's $\tau = 0.5$ at sample size $n = 100, 300, 600$. These plots allow us to examine (1) whether IR's null distribution is close to normal, and (2) whether the null distributions of IR and PIOS statistics are similar to each other.

First, we focus on comparing the IR's null distribution with normality. For a given sample size, the distribution gets more skewed to the right as the censoring rate increases. For each censoring scenario, as the sample size increases, it is getting closer to the normal distribution, which confirms the asymptotic normality of the IR statistic (Theorem 3).

Second, we compare the distributions of IR and PIOS. The QQ plots clearly show that their distributions are close, and they get more similar as the sample size increases. It confirms the asymptotic equivalence between IR and PIOS (Theorem 4). However, their computational times are substantially different. The PIOS statistic requires repeated (n times) estimation of the copula parameter, $\hat{\theta}_{(-i)}$, when obtaining the out-of-sample pseudo log-likelihood. Thus, its computational burden is more intensive than IR. In addition, as the sample size increases, IR is more computationally efficient. Specifically, using a Dell desktop computer with 3.20 GHz Intel(R) Core(TM) i7-8700 CPU, the average computational time with sample size $n = 100$ is 0.0072 seconds for calculating the IR statistic and 0.11 seconds for PIOS (about 15 times of IR's time). When the sample size increases to 600, the average computational time is 0.02 seconds for IR and 1.1 seconds for PIOS (about 55 times of IR's time). Note that the computation of the PIOS statistic has been optimized via parallel computation using R packages "parallel", "foreach", and "doSNOW". If without the parallel computation, it would take even longer.

5.3 Study II: Test Size and Power of IR Test

In this study, we investigate the type I error rate and power of the proposed IR test. The bivariate event times (T_{i1}, T_{i2}) are generated from each of the four copula families; under a true copula, we test each of the four copula families as the null hypothesis. For example, in one scenario, (T_{i1}, T_{i2}) is generated from Clayton, i.e., the true copula is Clayton, and we test four different null copulas: Clayton, Frank, Joe, and Gaussian. We consider three different dependence levels: Kendall's $\tau = 0.3, 0.5, 0.7$ and three sample sizes $n = 100, 300, 600$. Figures 1 - 4 plots the proportion of rejecting the null hypothesis at the significance level 0.05 among 500 replications at sample size 600. For each simulation replication, the P -value is calculated from $B = 500$ bootstrap resamples. The rejection proportions for sample size $n = 100$ and 300 are plotted in Figures 9 - 16 of the supplementary material.

When the null copula is the same as the true copula, the rejection proportions are the empirical type I error rates, also extracted in Table 1. In most scenarios, the IR test can maintain the nominal test size, i.e., the empirical type I error rates are close to the significance level 0.05. When the null copula is different from the true copula, the rejection proportions are the empirical test power. The results indicate that Kendall's τ , sample size, and censoring rate all affect the power. First, Kendall's τ reflects the strength of the dependence between the bivariate event times. When τ is large, i.e., the event times are highly dependent with each other, the true copula's distinct features such as tail dependence are more pronounced, and thus, our IR test is more powerful to detect deviations from the null copula. However, when the dependency is weak, copula families appear similar to each other (See Figures 1 and 3 in the supplementary material). Thus, the IR test has a lower power for a smaller Kendall's τ . Similar patterns are observed in Genest et al. (2009) and Zhang et al. (2016). Second, as expected, when the sample size is larger or the censoring rate is lower or both, the data provide more information of the underlying true copula, and consequently, the IR test is more powerful.

We observe that when the censoring rate is 70%, the proportion of rejecting Clayton when the true copula is Clayton is much lower than the significance level 0.05. In other words, the IR test is over conservative against Clayton when the event times are heavily censored. A possible explanation is that when the censoring time follows an exponential distribution, it is more likely to censor smaller event times, leading to insufficient information on the lower-tail dependence, which is a distinct feature of Clayton. As a result, the data exhibit minimal evidence against Clayton. With the same reason, when the true copula is Frank, it is difficult to tell apart from Clayton because they appear alike under heavy censoring (Figures 1 - 4 of the supplementary material). Thus, the proportion of rejecting Clayton when the true copula is Frank is low. Similarly, the proportion of rejecting Frank when the true copula is Clayton is also low. By contrast, since Joe has the upper-tail dependence, the IR test has a much higher power of rejecting Joe when the true copula is Clayton or Frank, or rejecting Clayton or Frank when the true copula is Joe.

We also observe low proportions of rejecting Frank when the true copula is Gaussian for all sample sizes, Kendall's τ values, and censoring rates (Figure 2). It could be because both families have no dependence on either tails. However, when the true

copula is Frank and the null copula is Gaussian, the test performs better (Figure 4). It calls for more investigations.

As pointed out in the introduction, our IR test can be regarded as a specific form of comparing the two information matrices in the class of generalized IM tests (Prokhorov et al. 2019). The other forms include the White test, determinant White test, trace White test, determinant IR test, log trace IM test, log GAIC IM test, log eigenspectrum IM test, and eigenvalue test. However, for the case of scalar parameter, i.e., $p = 1$, some tests are equivalent. Specifically, the determinant IR and eigenvalue tests are the same as the IR test: $R_n = \hat{\mathbf{S}}_n(\hat{\boldsymbol{\theta}}_n)^{-1} \hat{\mathbf{V}}_n(\hat{\boldsymbol{\theta}}_n)$. The White, determinant White, and trace White are the same; they all take a difference: $T_n = \hat{\mathbf{V}}_n(\hat{\boldsymbol{\theta}}_n) - \hat{\mathbf{S}}_n(\hat{\boldsymbol{\theta}}_n)$. The log trace IM, log GAIC IM, and log eigenspectrum IM are equivalent, given as $Z_n = \log[\hat{\mathbf{S}}_n(\hat{\boldsymbol{\theta}}_n)] - \log[\hat{\mathbf{V}}_n(\hat{\boldsymbol{\theta}}_n)]$. Thus, in this study, we compare our IR test R_n with T_n , referred to as the White test, and Z_n , referred to as the log IM test. The P -values of these two tests are also obtained by the parametric bootstrap resampling procedure described in Section 4.1. The results show that these three tests perform similarly for most scenarios.

5.4 Copula Selection

We also examine how well using the P -value of the IR test as the criterion can correctly select the true copula as the best among the four families. With each simulated data, we obtain the P -value for testing each of Clayton, Frank, Joe, and Gaussian as the null hypothesis. Following Section 4.2, we select the copula family with the largest P -value as the best. Figures 17 - 28 in the supplementary material report the percentage of choosing each family as the best among the 500 replications. Consistent with our findings on the test power, when the sample size is larger or the dependence is stronger or the censoring rate is lower, the proportion of selecting the true copula as the best is higher. Copulas with similar properties are more difficult to tell apart. For example, when the true copula is Gaussian, Frank copula is a strong competitor, even when the sample size is 600, Kendall's $\tau = 0.7$, and the event times are fully observed. In addition, our IR test performs similarly to the other two generalized IM tests.

6 Data Example

The data example is 748 dizygotic female twin pairs from the Australian NHMRC Twin Registry (Duffy et al. 1990), and the bivariate event times (T_1, T_2) are the ages at appendicectomy measured for each twin pair. For this data, the event times are heavily censored with the censoring rate of about 74%. Among the 748 twin pairs, 82 (11%) pairs have both event times observed, 222 (30%) have one event time observed and the other censored, and 444 (59%) have both event times censored. Figure 5 plots the estimated pseudo-observations $\{(\hat{U}_{i1}, \hat{U}_{i2}), i = 1, \dots, n\}$.

Emura et al. (2010) analyzed this data and concluded that Gumbel provides the best fit over three other copula families: Clayton, Frank, and Log-copula. In this manuscript, we test for five copula families: Clayton, Frank, Gumbel, Joe, and Gaus-

sian using our proposed IR test as well as the White test and log IM test. Table 2 reports their test statistic values and P -values calculated using $B = 1000$ bootstrapped resamples.

Among the three tests, only the log IM test reaches the same conclusion as Emura et al. (2010): Gumbel is the best copula family with the P -value 0.307, and Clayton is the second best with the P -value 0.273. In contrast, for both our IR test and White test, Clayton is the best copula and Gumbel is the second. However, under the IR test, the difference of the P -values between Clayton and Gumbel is tiny: the P -values is 0.296 for Clayton and 0.291 for Gumbel. It indicates that Gumbel's goodness-of-fit is comparable with Clayton. On the other hand, under the White test, the lead of Clayton over Gumbel is more substantial (P -values 0.379 for Clayton and 0.258 for Gumbel).

7 Concluding Remarks

Information matrix equivalence plays an important role in model diagnosis, and a number of GoF tests have been established based on this principle. However, this equivalence has not been verified for censored data. Thus, one major contribution of this work is to prove the equivalence of the two information matrices under a class of semiparametric copula models for multivariate data in the presence of right censoring. The proof provides a framework which might be extended to other censoring schemes.

Based on this equivalence, we propose an IR test for the specification of the copula function via comparing consistent estimates of the two information matrices. This test is likelihood-based and depends on only the parametric form of the assumed copula function. Thus, it can be applied to all copula families, and do not rely on choices of weight functions, bandwidth, or smoothing parameters. In addition, the IR statistic is asymptotically equivalent to a class of PIOS test statistics, which provides a global measure of how the assumed model fits the data via the leave-one-out cross-validation. Furthermore, the IR test does not assume any parametric form of alternative copulas. It can be regarded as an omnibus test.

In this manuscript, we derive the asymptotic properties of the IR statistic following similar arguments in Chen et al. (2010). They considered a more general distributional assumption for censoring: the joint distribution of the bivariate censoring times could be different across subjects. Under this relaxed assumption, the pseudo-true value of the copula parameter is defined as

$$\theta_n^* = \arg \max_{\theta} n^{-1} \sum_{i=1}^n \mathbb{E}^0[\ell(\theta, U_{i1}^0, U_{i2}^0)].$$

This value depends on the sample size since the observed survival times might not be identically distributed due to non-identically distributed bivariate censoring times. Correspondingly, the definitions of the sensitivity and variability matrices can be modified as $\mathbf{S}^*(\theta) = n^{-1} \sum_{i=1}^n \mathbb{E}^0[-\ell_{\theta\theta}]$ and $\mathbf{V}^*(\theta) = n^{-1} \sum_{i=1}^n \mathbb{E}^0[\ell_{\theta} \ell'_{\theta}]$, which also depend on the sample size. It worths pointing out that the proof of Theorem 1 is still

valid, and thus, the information matrix equivalence still holds, and the IR test is still valid. However, for generating bootstrap resamples of censoring times, the Kaplan-Meier estimator of the censoring survival function is not appropriate when assuming heterogeneous censoring distribution. Under this assumption, it would require some subject-specific covariates Z_i to estimate the subject-specific censoring survival function $G_{ir}(t) = \Pr(C_{ir} > t \mid Z_i)$ for $r = 1, 2$.

In general, if testing within Archimedean families, the GoF tests that target these families are expected to be more powerful than our proposed IR test because they utilize their distinct properties such as cross-ratio functions or Kendall distribution. On the other hand, our proposed IR test can compare copula families beyond Archimedean. In Section 4.2, we demonstrate how to use the P -value of the IR test to select the best copula family among several candidates.

Equation (8) is the key step for proving the information matrix equivalence. It also implies that this equivalence holds for any censoring distribution when the assumed copula is correctly specified. However, when the assumed copula is misspecified, the difference between two information matrices depends on the censoring distribution. Our simulation study has shown that the censoring rate is one factor that affects the performance of the IR test. We hypothesize that besides the censoring rate, the shape of the censoring distribution might be another factor. For example, as discussed in our simulation, the shape of the exponential distribution for censoring leads to insufficient information on the lower-tail dependence. It causes the low power of differentiating between Clayton and Frank. It is our interest to conduct more studies to investigate other distributions for censoring, such as gamma, Weibull, or uniform distributions.

In the simulation study and data example, we compare our IR test with two other forms of generalized IM tests, and they perform similarly. For example, all three tests exhibit a lower power for rejecting Frank when the true copula is Gaussian because they both have neither upper-tail or lower-tail dependence. Our studies focus on the case of scalar copula parameter, i.e., $p = 1$, for which the class of generalized IM tests reduces to three forms of comparing IMs: ratio, difference, and difference of logarithm. However, if $p > 1$, the class would not be limited to only these three forms. In addition, different IM-based tests would perform more diversely. It warrants further investigation for cases with $p > 1$.

8 Supplementary Material

In the supplementary material, we present the expressions of the copula function, and the derivatives of the log-likelihood function for Clayton, Frank, Joe, and Gaussian copulas. We also show more results of the simulation study, including (i) scatter plots (Figures 1 - 4) of estimated pseudo-observations $(\hat{U}_{i1}, \hat{U}_{i2})$ from one simulated bivariate censored data, (ii) QQ plots (Figures 5 - 8) of the IR and PIOS statistics when the null copula is the true copula, (iii) bar plots (Figures 9 - 16) of proportions of rejecting the null hypothesis for sample sizes 100 and 300, and (iv) bar plots (Figures 17 - 28) of proportions of selecting different copula families as the best copula.

9 Acknowledgement

The work is supported by the grant DMS-2210481 from the National Science Foundation.

References

- Andersen, P. K., Ekstrøm, C. T., Klein, J. P., Shu, Y., and Zhang, M.-J. (2005). A class of goodness of fit tests for a copula based on bivariate right-censored data. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 47(6):815–824.
- Bartlett, M. S. (1953a). Approximate Confidence Intervals. *Biometrika*, 40(1/2):12–19.
- Bartlett, M. S. (1953b). Approximate Confidence Intervals.II. More than one Unknown Parameter. *Biometrika*, 40(3/4):306–317.
- Chen, X. and Fan, Y. (2006a). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of econometrics*, 135(1):125–154.
- Chen, X. and Fan, Y. (2006b). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2):307–335.
- Chen, X., Fan, Y., Pouzo, D., and Ying, Z. (2010). Estimation and model selection of semiparametric multivariate survival functions under general censorship. *Journal of Econometrics*, 157(1):129–142.
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.
- Dhaene, G. and Hoorelbeke, D. (2004). The information matrix test with bootstrap-based covariance matrix estimation. *Economics Letters*, 82(3):341–347.
- Duffy, D. L., Martin, N. G., and Mathews, J. D. (1990). Appendectomy in Australian twins. *American Journal of Human Genetics*, 47(3):590–592.
- Emura, T., Lin, C.-W., and Wang, W. (2010). A goodness-of-fit test for Archimedean copula models in the presence of right censoring. *Computational statistics & data analysis*, 54(12):3033–3043.
- Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. In *Annales de l’IHP Probabilités et statistiques*, volume 44, pages 1096–1127.
- Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, 44(2):199–213.

- Georges, P., Lamy, A.-G., Nicolas, E., Quibel, G., and Roncalli, T. (2001). Multivariate survival modelling: a unified approach with copulas. *Available at SSRN: <https://ssrn.com/abstract=1032559> or <http://dx.doi.org/10.2139/ssrn.1032559>.*
- Golden, R. M., Henley, S. S., White, H., and Kashner, T. M. (2013). New Directions in Information Matrix Testing: Eigenspectrum Tests. In Chen, X. and Swanson, N. R., editors, *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White Jr*, pages 145–177. Springer, New York, NY.
- Golden, R. M., Henley, S. S., White, H., and Kashner, T. M. (2016). Generalized Information Matrix Tests for Detecting Model Misspecification. *Econometrics*, 4(4):46.
- Horowitz, J. L. (1994). Bootstrap-based critical values for the information matrix test. *Journal of Econometrics*, 61(2):395–411.
- Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika*, 73(3):671–678.
- Huang, W. and Prokhorov, A. (2014). A goodness-of-fit test for copulas. *Econometric Reviews*, 33(7):751–771.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2):81–93.
- Lakhal-Chaieb, M. L. (2010). Copula inference under censoring. *Biometrika*, 97(2):505–512.
- Li, D., Hu, X. J., McBride, M. L., and Spinelli, J. J. (2020). Multiple event times in the presence of informative censoring: Modeling and analysis by copulas. *Lifetime Data Analysis*, 26(3):573–602.
- Li, Y., Prentice, R. L., and Lin, X. (2008). Semiparametric Maximum Likelihood Estimation in Normal Transformation Models for Bivariate Survival Data. *Biometrika*, 95(4):947–960.
- Lin, J. and Wu, X. (2020). A diagnostic test for specification of copulas under censorship. *Econometric Reviews*, pages 1–17.
- Mikosch, T. (2006). Copulas: Tales and facts—rejoinder. *Extremes*, 9(1):55–62.
- Nelsen, R. B. (2006). Archimedean Copulas. In *An Introduction to Copulas*, Springer Series in Statistics, pages 109–155. Springer, New York, NY.
- Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.
- Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84(406):487–493.

- Othus, M. and Li, Y. (2010). A Gaussian Copula Model for Multivariate Survival Data. *Statistics in Biosciences*, 2(2):154–179.
- Prokhorov, A., Schepsmeier, U., and Zhu, Y. (2019). Generalized information matrix tests for copulas. *Econometric Reviews*, 38(9):1024–1054.
- Shih, J. H. (1998). A goodness-of-fit test for association in a bivariate survival model. *Biometrika*, 85(1):189–200.
- Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51:1384–1399.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, 8:229–231.
- Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.
- Wang, A. (2010). Goodness-of-fit tests for archimedean copula models. *Statistica Sinica*, 20(1):441–453.
- Wang, W. and Wells, M. T. (2000). Model Selection and Semiparametric Inference for Bivariate Failure-Time Data. *Journal of the American Statistical Association*, 95(449):62–72.
- White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50(1):1–25.
- Yilmaz, Y. E. and Lawless, J. F. (2011). Likelihood ratio procedures and tests of fit in parametric and semiparametric copula models with censored data. *Lifetime Data Analysis*, 17(3):386–408.
- Zhang, S., Okhrin, O., Zhou, Q. M., and Song, P. X. K. (2016). Goodness-of-fit test for specification of semiparametric copula dependence models. *Journal of Econometrics*, 193(1):215–233.
- Zhang, S., Song, P. X.-K., Shi, D., and Zhou, Q. M. (2012). Information ratio test for model misspecification on parametric structures in stochastic diffusion models. *Computational Statistics and Data Analysis*, 56:3975–3987.
- Zhang, S., Zhou, Q. M., and Lin, H. (2021). Goodness-of-fit test of copula functions for semi-parametric univariate time series models. *Statistical Papers*, 62(4):1697–1721.
- Zhang, S., Zhou, Q. M., Zhu, D., and Song, P. X.-K. (2019). Goodness-of-Fit Test in Multivariate Jump Diffusion Models. *Journal of Business & Economic Statistics*, 37(2):275–287.
- Zhou, Q. M., Song, P. X.-K., and Thompson, M. E. (2012). Information Ratio Test for Model Misspecification in Quasi-Likelihood Inference. *Journal of the American Statistical Association*, 107(497):205–213.

Table 1: Simulation results: Empirical type I errors for the IR test R_n , White test T_n , and log IM test Z_n .

τ	n	No Cen.			20% Cen.			40% Cen.			70% Cen.		
		R_n	T_n	Z_n	R_n	T_n	Z_n	IR_n	T_n	Z_n	R_n	T_n	Z_n
Clayton Copula													
0.3	100	0.038	0.018	0.044	0.034	0.014	0.048	0.024	0.002	0.026	0.000	0.000	0.078
	300	0.036	0.026	0.032	0.030	0.018	0.034	0.022	0.010	0.020	0.012	0.002	0.016
	600	0.044	0.036	0.046	0.034	0.030	0.036	0.018	0.014	0.022	0.012	0.004	0.014
0.5	100	0.048	0.028	0.038	0.032	0.022	0.036	0.032	0.012	0.040	0.022	0.000	0.040
	300	0.036	0.030	0.042	0.028	0.018	0.024	0.024	0.018	0.028	0.008	0.004	0.010
	600	0.048	0.040	0.052	0.042	0.040	0.042	0.022	0.018	0.024	0.004	0.004	0.008
0.7	100	0.054	0.026	0.058	0.026	0.014	0.026	0.020	0.010	0.022	0.016	0.002	0.016
	300	0.050	0.034	0.050	0.036	0.024	0.040	0.014	0.012	0.018	0.016	0.006	0.018
	600	0.028	0.028	0.034	0.018	0.016	0.020	0.014	0.010	0.014	0.008	0.002	0.008
Frank Copula													
0.3	100	0.032	0.034	0.038	0.028	0.032	0.044	0.040	0.048	0.050	0.022	0.034	0.034
	300	0.050	0.052	0.052	0.048	0.046	0.052	0.046	0.044	0.042	0.058	0.060	0.050
	600	0.036	0.040	0.030	0.040	0.040	0.034	0.038	0.038	0.036	0.040	0.040	0.036
0.5	100	0.038	0.036	0.044	0.038	0.036	0.058	0.048	0.054	0.066	0.024	0.024	0.048
	300	0.054	0.054	0.048	0.038	0.036	0.040	0.040	0.040	0.048	0.030	0.034	0.048
	600	0.038	0.028	0.028	0.038	0.042	0.040	0.036	0.038	0.034	0.038	0.038	0.034
0.7	100	0.048	0.040	0.068	0.050	0.038	0.058	0.066	0.046	0.078	0.018	0.010	0.118
	300	0.066	0.060	0.058	0.054	0.046	0.048	0.054	0.040	0.058	0.030	0.030	0.052
	600	0.040	0.040	0.034	0.042	0.042	0.034	0.038	0.038	0.038	0.042	0.042	0.064
Joe Copula													
0.3	100	0.054	0.064	0.066	0.046	0.062	0.062	0.052	0.074	0.060	0.034	0.066	0.038
	300	0.066	0.078	0.078	0.058	0.066	0.076	0.050	0.060	0.070	0.046	0.050	0.054
	600	0.054	0.072	0.056	0.048	0.062	0.046	0.046	0.060	0.050	0.048	0.064	0.046
0.5	100	0.050	0.044	0.052	0.052	0.046	0.054	0.040	0.034	0.050	0.048	0.036	0.050
	300	0.068	0.062	0.082	0.058	0.060	0.078	0.050	0.048	0.060	0.056	0.054	0.062
	600	0.046	0.052	0.042	0.044	0.052	0.048	0.050	0.054	0.048	0.046	0.054	0.044
0.7	100	0.050	0.038	0.046	0.054	0.040	0.052	0.044	0.034	0.038	0.046	0.028	0.042
	300	0.056	0.042	0.050	0.050	0.044	0.050	0.052	0.042	0.050	0.064	0.046	0.052
	600	0.062	0.062	0.062	0.074	0.074	0.068	0.060	0.062	0.060	0.080	0.084	0.076
Gaussian Copula													
0.3	100	0.026	0.014	0.048	0.042	0.016	0.044	0.034	0.014	0.036	0.014	0.000	0.018
	300	0.040	0.028	0.046	0.040	0.034	0.044	0.052	0.042	0.052	0.024	0.012	0.034
	600	0.054	0.044	0.056	0.052	0.044	0.060	0.070	0.062	0.070	0.042	0.030	0.036
0.5	100	0.022	0.018	0.044	0.022	0.012	0.042	0.022	0.002	0.034	0.018	0.002	0.026
	300	0.046	0.042	0.042	0.040	0.034	0.048	0.054	0.036	0.054	0.024	0.010	0.032
	600	0.042	0.036	0.044	0.062	0.048	0.060	0.052	0.048	0.056	0.038	0.028	0.042
0.7	100	0.032	0.028	0.040	0.030	0.020	0.032	0.026	0.010	0.026	0.022	0.006	0.020
	300	0.048	0.042	0.044	0.044	0.040	0.040	0.046	0.046	0.042	0.012	0.008	0.016
	600	0.048	0.042	0.046	0.048	0.042	0.050	0.044	0.044	0.054	0.024	0.014	0.024

Table 2: Data example: The PMLE $\hat{\theta}_n$ of the copula parameter, the test statistic with the P -value (in the paratheses) of the IR, White, and log IM tests for Clayton, Frank, Gumbel, Joe, and Gaussian.

Copula	$\hat{\theta}_n$	IR	White	log IM
Clayton	0.750	1.085 (0.296)	0.003 (0.379)	0.081 (0.273)
Frank	1.795	1.075 (0.039)	0.001 (0.038)	0.072 (0.046)
Gumbel	1.162	1.060 (0.291)	0.051 (0.258)	0.058 (0.307)
Joe	1.204	1.085 (0.215)	0.045 (0.191)	0.081 (0.238)
Gaussian	0.304	1.083 (0.166)	0.035 (0.188)	0.079 (0.177)

Figure 1: Simulation results: Proportions of rejecting **Clayton** when the true copula is Clayton, Frank, Joe, or Gaussian and the sample size is 600. The dashed lines represent the significance level 0.05.

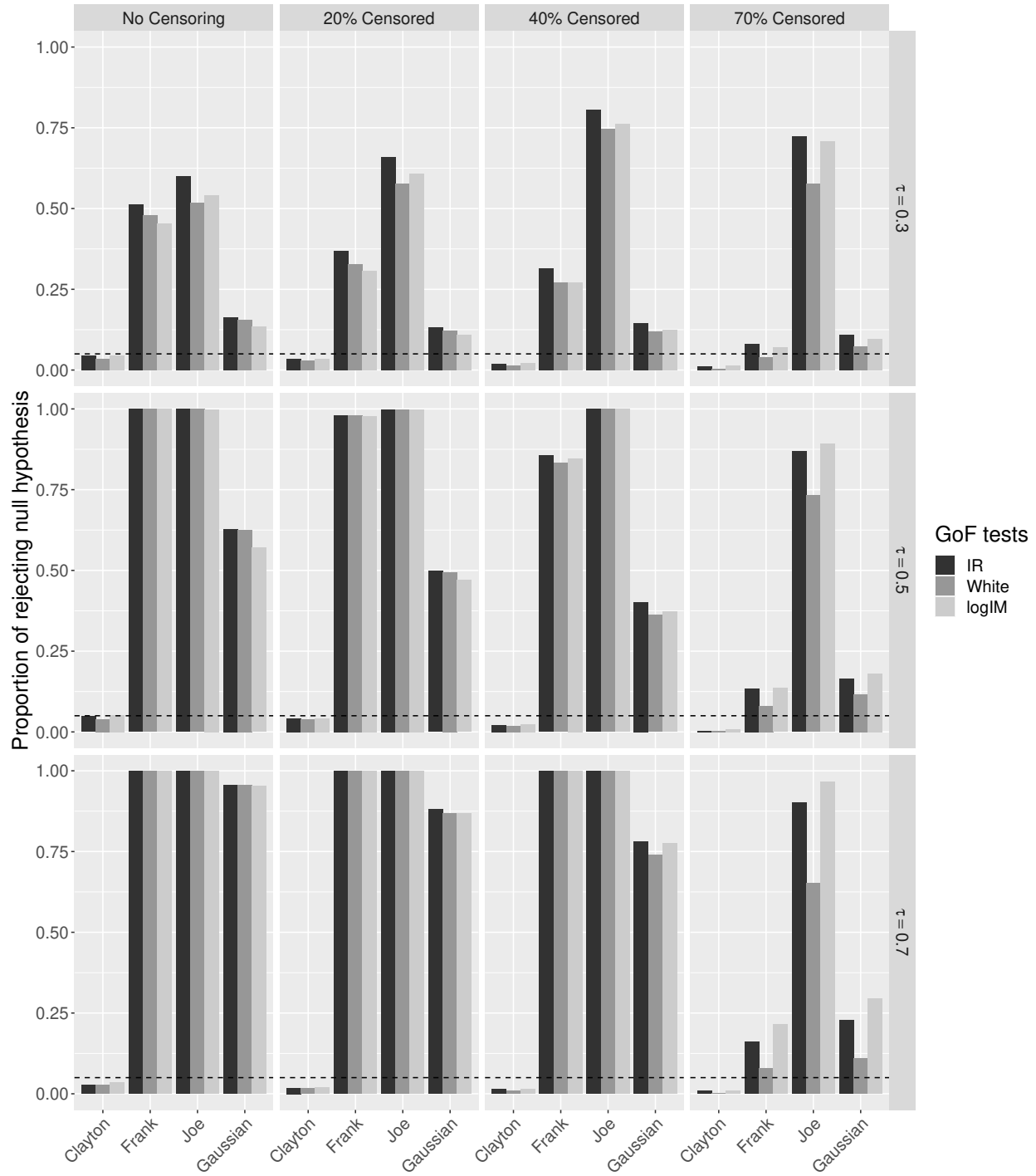


Figure 2: Simulation results: Proportions of rejecting **Frank** when the true copula is Clayton, Frank, Joe, or Gaussian and the sample size is 600. The dashed lines represent the significance level 0.05.

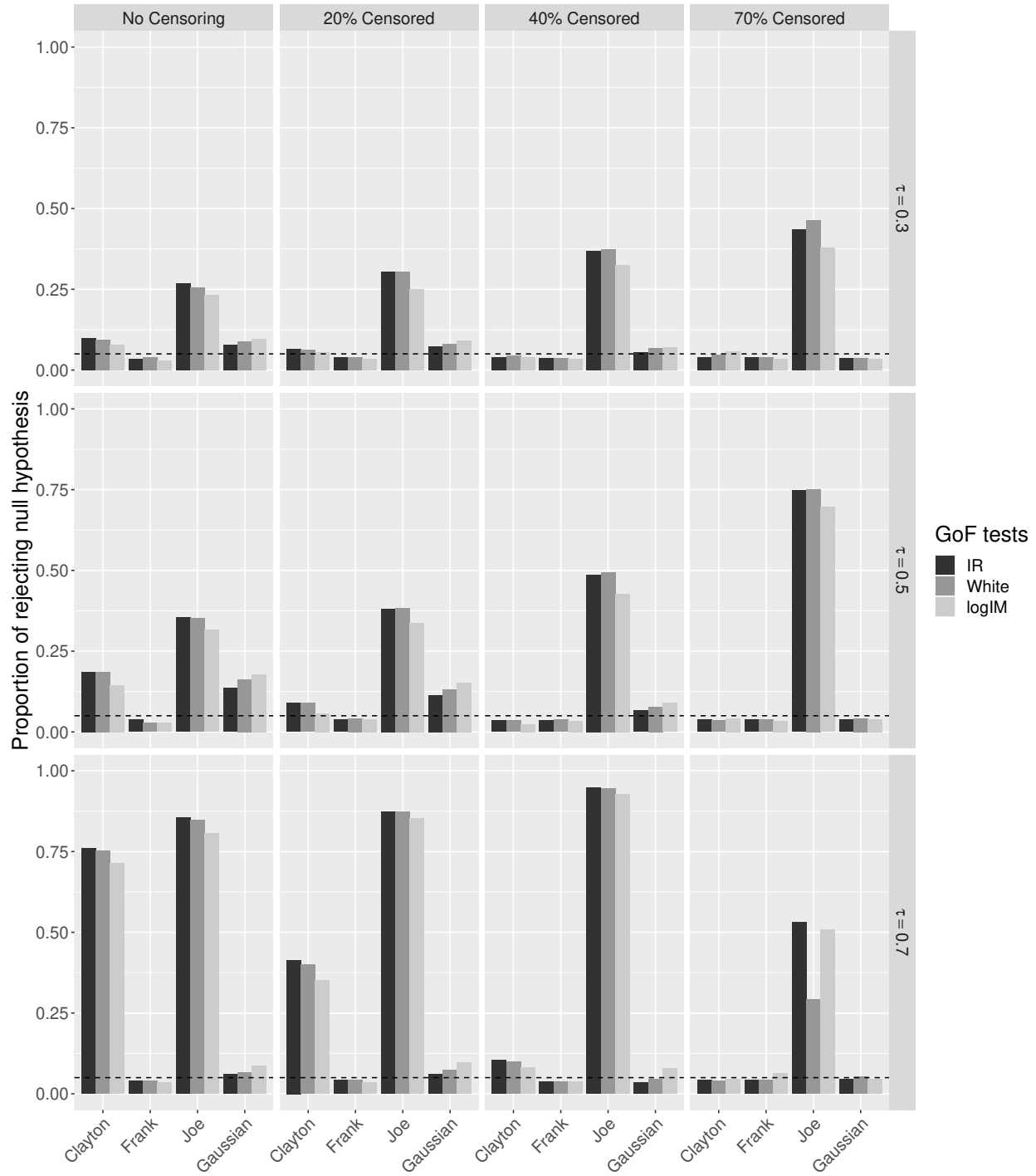


Figure 3: Simulation results: Proportions of rejecting **Joe** when the true copula is Clayton, Frank, Joe, or Gaussian and the sample size is 600. The dashed lines represent the significance level 0.05.

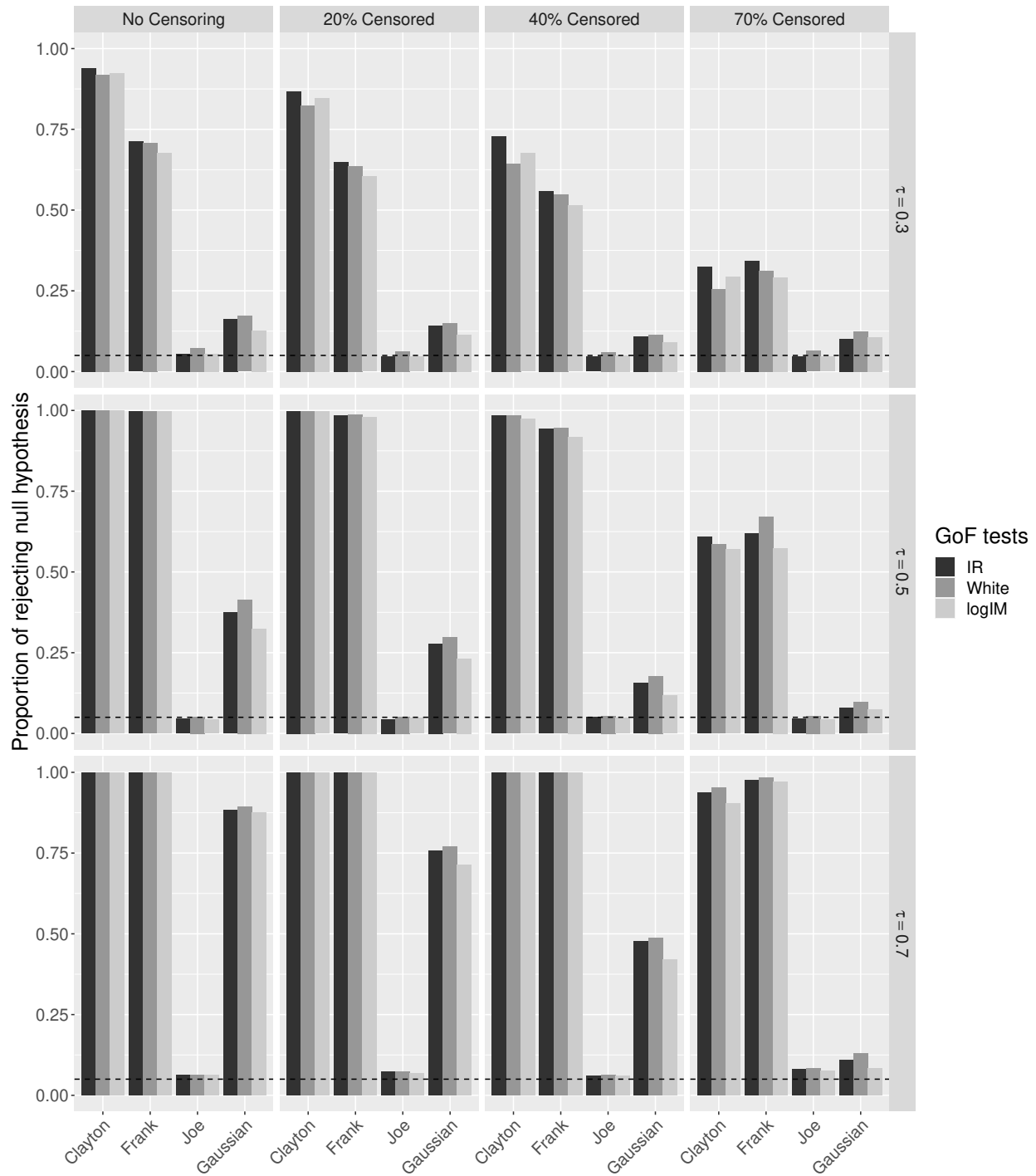


Figure 4: Simulation results: Proportions of rejecting **Gaussian** when the true copula is Clayton, Frank, Joe, or Gaussian and the sample size is 600. The dashed lines represent the significance level 0.05.

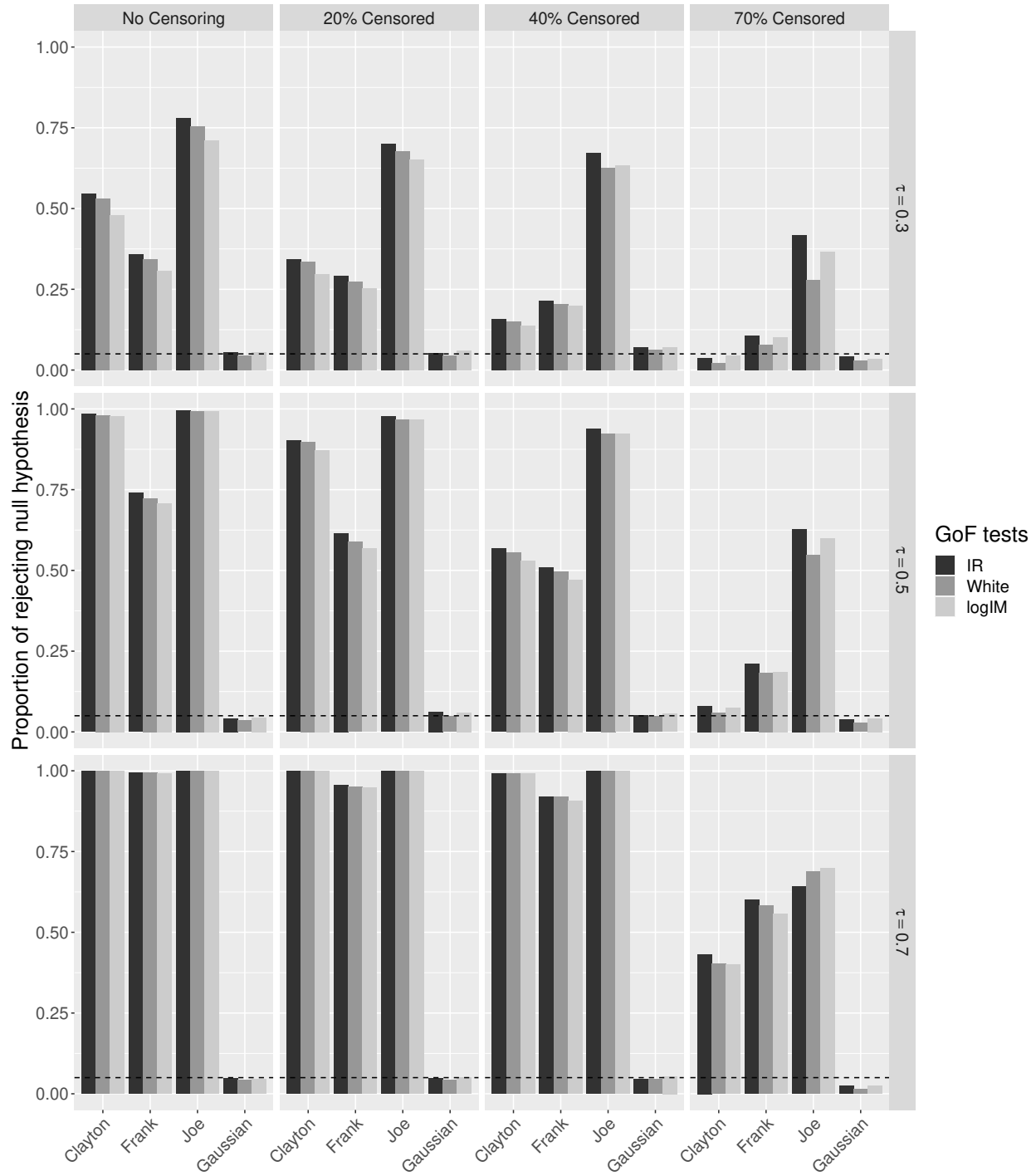
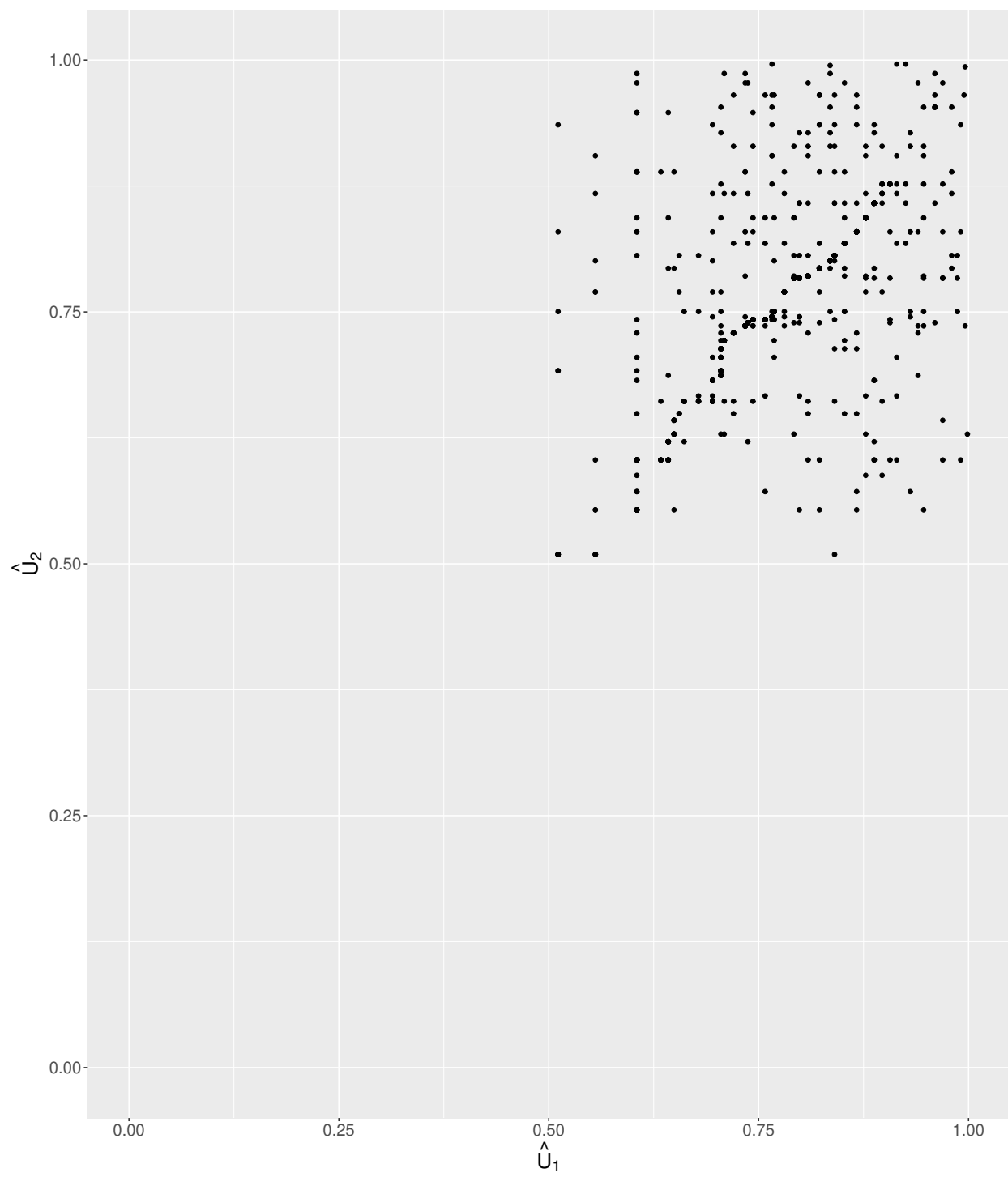


Figure 5: Data example: Scatter plot of estimated pseudo-observations $\hat{U}_{i1}$ and $\hat{U}_{i2}$.



Appendix

A Proof of Theorem 1

The outline of the prove is as follows. First, we will show that for an assumed copula with a given θ ,

$$\mathbf{S}^*(\theta) = \mathbf{V}^*(\theta) - \mathbf{A}(\theta), \quad (10)$$

where $\mathbf{A}(\theta)$ is a $p \times p$ matrix. Second, we prove that if the assumed copula is correctly specified, $\mathbf{A}(\theta^*) = \mathbf{0}_{p \times p}$.

Derive $\mathbf{S}^*(\theta) = \mathbf{V}^*(\theta) - \mathbf{A}(\theta)$. The sensitivity and variability matrices in Equation (7) are functions of $\ell_\theta \ell'_\theta$ and $\ell_{\theta\theta}$. We derive the expressions of these two quantities. By Equation (4), we have

$$\ell_\theta = \delta_1 \delta_2 \frac{\mathbb{C}_\theta}{\mathbb{C}} + \delta_1 (1 - \delta_2) \frac{\mathbb{C}_{1,\theta}}{\mathbb{C}_1} + (1 - \delta_1) \delta_2 \frac{\mathbb{C}_{2,\theta}}{\mathbb{C}_2} + (1 - \delta_1)(1 - \delta_2) \frac{\mathbb{C}_\theta}{\mathbb{C}}. \quad (11)$$

Thus,

$$\ell_\theta \ell'_\theta = \delta_1 \delta_2 \frac{\mathbb{C}_\theta \mathbb{C}'_\theta}{\mathbb{C}^2} + \delta_1 (1 - \delta_2) \frac{\mathbb{C}_{1,\theta} \mathbb{C}'_{1,\theta}}{\mathbb{C}_1^2} + (1 - \delta_1) \delta_2 \frac{\mathbb{C}_{2,\theta} \mathbb{C}'_{2,\theta}}{\mathbb{C}_2^2} + (1 - \delta_1)(1 - \delta_2) \frac{\mathbb{C}_\theta \mathbb{C}'_\theta}{\mathbb{C}^2}, \quad (12)$$

and

$$\begin{aligned} \ell_{\theta\theta} = & \delta_1 \delta_2 \left[\frac{\mathbb{C}_{\theta\theta}}{\mathbb{C}} - \frac{\mathbb{C}_\theta \mathbb{C}'_\theta}{\mathbb{C}^2} \right] + \delta_1 (1 - \delta_2) \left[\frac{\mathbb{C}_{1,\theta\theta}}{\mathbb{C}_1} - \frac{\mathbb{C}_{1,\theta} \mathbb{C}'_{1,\theta}}{\mathbb{C}_1^2} \right] \\ & + (1 - \delta_1) \delta_2 \left[\frac{\mathbb{C}_{2,\theta\theta}}{\mathbb{C}_2} - \frac{\mathbb{C}_{2,\theta} \mathbb{C}'_{2,\theta}}{\mathbb{C}_2^2} \right] + (1 - \delta_1)(1 - \delta_2) \left[\frac{\mathbb{C}_{\theta\theta}}{\mathbb{C}} - \frac{\mathbb{C}_\theta \mathbb{C}'_\theta}{\mathbb{C}^2} \right]. \end{aligned} \quad (13)$$

Consequently, $-\ell_{\theta\theta} = \ell_\theta \ell'_\theta - \Delta$, where Δ is a $p \times p$ matrix with the (j, k) -th element

$$\begin{aligned} \Delta_{jk}(\theta, u_1, u_2) = & \delta_1 \delta_2 \frac{\mathbb{C}_{\theta, \theta_k}(u_1, u_2; \theta)}{\mathbb{C}(u_1, u_2; \theta)} + \delta_1 (1 - \delta_2) \frac{\mathbb{C}_{1, \theta, \theta_k}(u_1, u_2; \theta)}{\mathbb{C}_1(u_1, u_2; \theta)} \\ & + (1 - \delta_1) \delta_2 \frac{\mathbb{C}_{2, \theta, \theta_k}(u_1, u_2; \theta)}{\mathbb{C}_2(u_1, u_2; \theta)} + (1 - \delta_1)(1 - \delta_2) \frac{\mathbb{C}_{\theta, \theta_k}(u_1, u_2; \theta)}{\mathbb{C}(u_1, u_2; \theta)}, \end{aligned}$$

Note that, for simplicity, we suppress (δ_1, δ_2) from the Δ function. Thus, $\mathbf{A}(\theta)$ is a $p \times p$ matrix with the (j, k) -th element $A_{jk}(\theta) = \mathbb{E}^0[\Delta_{jk}(\theta, U_1^0, U_2^0)]$.

To derive the expression of $A_{jk}(\theta)$, we invoke the double expectation theorem by conditioning on (C_1, C_2) , i.e., $A_{jk}(\theta) = \mathbb{E}_{(C_1, C_2)}^0 \left\{ \mathbb{E}_{(T_1, T_2)}^0 [\Delta_{jk} | C_1, C_2] \right\}$, where $\mathbb{E}_{(T_1, T_2)}^0$ and $\mathbb{E}_{(C_1, C_2)}^0$ denote the expectations w.r.t. the true distributions of (T_1, T_2) and (C_1, C_2) , respectively. In Remark 1, we stated that given that the true marginal survival functions $H_r^0(\cdot)$ are known, the copula $\mathbb{C}$ can be regarded as the joint CDF of (Y_1, Y_2) with

$Y_r = H_r^0(T_r)$, $r = 1, 2$, which are uniformly distributed on $(0, 1)$. Thus, the expectation w.r.t. (T_1, T_2) is equivalent to the expectation w.r.t. (Y_1, Y_2) , which gives

$$A_{jk}(\boldsymbol{\theta}) = \mathbb{E}_{(C_1, C_2)}^0 \left\{ \mathbb{E}_{(Y_1, Y_2)}^0 [\Delta_{jk}(\boldsymbol{\theta}; U_1^0, U_2^0) | C_1, C_2] \right\}. \quad (14)$$

In addition, U_r^0 can be expressed as $U_r^0 = \max\{Y_r, H_r^0(C_r)\}$ and $\delta_r = I(Y_r \geq H_r^0(C_r))$.

Proof of $A_{jk}(\boldsymbol{\theta}^*) = 0$ under correct specification. If the assumed copula $\mathbb{C}$ is correctly specified, $\mathbb{E}_{(Y_1, Y_2)}^0$ is taken w.r.t. $\mathbb{C}$ as follows:

$$\mathbb{E}_{(Y_1, Y_2)}^0 [\Delta_{jk} | C_1, C_2] = \iint \Delta_{jk}(\boldsymbol{\theta}; U_1^0, U_2^0) \mathbb{C}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2.$$

Given the regularity conditions R1 - R6, by the law of total probability, we can show that this conditional expectation $\mathbb{E}_{(Y_1, Y_2)}^0 [\Delta_{jk} | C_1, C_2] = \mathcal{A}_{(1,1)} + \mathcal{A}_{(1,0)} + \mathcal{A}_{(0,1)} + \mathcal{A}_{(0,0)}$ where $\mathcal{A}_{(1,1)}$, $\mathcal{A}_{(1,0)}$, $\mathcal{A}_{(0,1)}$, and $\mathcal{A}_{(0,0)}$ correspond to each censoring scenario (δ_1, δ_2) , given as

$$\mathcal{A}_{(1,1)}(\boldsymbol{\theta}, C_1, C_2) = \iint_{[H_1^0(C_1), 1] \times [H_2^0(C_2), 1]} \mathbb{C}_{\theta_j \theta_k}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2, \quad (15)$$

$$\mathcal{A}_{(1,0)}(\boldsymbol{\theta}, C_1, C_2) = \iint_{[H_1^0(C_1), 1] \times [0, H_2^0(C_2)]} \mathbb{C}_{\theta_j \theta_k}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2, \quad (16)$$

$$\mathcal{A}_{(0,1)}(\boldsymbol{\theta}, C_1, C_2) = \iint_{[0, H_1^0(C_1)] \times [H_2^0(C_2), 1]} \mathbb{C}_{\theta_j \theta_k}(u_1, u_2; \boldsymbol{\theta}) du_2 du_2, \quad (17)$$

$$\mathcal{A}_{(0,0)}(\boldsymbol{\theta}, C_1, C_2) = \iint_{[0, H_1^0(C_1)] \times [0, H_2^0(C_2)]} \mathbb{C}_{\theta_j \theta_k}(u_1, u_2; \boldsymbol{\theta}) du_2 du_2. \quad (18)$$

- Under the scenario with $\delta_1 = 1$ and $\delta_2 = 1$, $\Delta_{jk} = \mathbb{C}_{\theta_j \theta_k} / \mathbb{C}$ with both $U_1^0 = Y_1$ and $U_2^0 = Y_2$ being random variables. In addition, the integral under this scenario is taken over the region of (Y_1, Y_2) : $\Omega_{11} = [H_1^0(C_1), 1] \times [H_2^0(C_2), 1]$. Thus, the conditional expectation is

$$\iint_{\Omega_{11}} \frac{\mathbb{C}_{\theta_j \theta_k}(u_1, u_2; \boldsymbol{\theta})}{\mathbb{C}(u_1, u_2; \boldsymbol{\theta})} \mathbb{C}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2 = \iint_{\Omega_{11}} \mathbb{C}_{\theta_j \theta_k}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2,$$

which results in A_{11} in Equation (15).

- Under the scenario with $\delta_1 = 1$ and $\delta_2 = 0$, $\Delta = \mathbb{C}_{1, \theta_j \theta_k} / \mathbb{C}_1$ with $U_2^0 = H_2^0(C_2)$ as a fixed number and $U_1^0 = Y_1$ as the only random variable. In addition, the integral under this scenario is taken over the region of (Y_1, Y_2) : $\Omega_{10} = [H_1^0(C_1), 1] \times [0, H_2^0(C_2)]$. Thus, the conditional expectation is

$$\begin{aligned} & \int_{[H_1^0(C_1), 1]} \frac{\mathbb{C}_{1, \theta_j \theta_k}(u_1, H_2^0(C_2); \boldsymbol{\theta})}{\mathbb{C}_1(u_1, H_2^0(C_2); \boldsymbol{\theta})} \left\{ \int_{[0, H_2^0(C_2)]} \mathbb{C}(u_1, u_2; \boldsymbol{\theta}) du_2 \right\} du_1 \\ &= \int_{[H_1^0(C_1), 1]} \mathbb{C}_{1, \theta_j \theta_k}(u_1, H_2^0(C_2); \boldsymbol{\theta}) du_1, \end{aligned}$$

since $\int_{[0, H_2^0(C_2)]} \mathbb{C}(u_1, u_2; \boldsymbol{\theta}) du_2 = \mathbb{C}_1(u_1, H_2^0(C_2); \boldsymbol{\theta})$. The regularity condition R6 ensures that functions $\mathbb{C}_{\theta, \theta_k}, j, k = 1, \dots, p$, are dominated by integrable function w.r.t. (u_1, u_2) for all $\boldsymbol{\theta}$. It allows the following interchangeability between the integral and derivation in $\mathbb{C}_{1, \theta, \theta_k}$:

$$\frac{\partial^2}{\partial \theta_j \partial \theta_k} \int_{[0, H_2^0(C_2)]} \mathbb{C}(u_1, u_2; \boldsymbol{\theta}) du_2 = \int_{[0, H_2^0(C_2)]} \mathbb{C}_{\theta, \theta_k}(u_1, u_2; \boldsymbol{\theta}) du_2.$$

Thus, the conditional expectation is $\iint_{\Omega_{10}} \mathbb{C}_{\theta, \theta_k}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2$, which is A_{10} in Equation (16).

- Under the scenario with $\delta_1 = 0$ and $\delta_2 = 1$, $\Delta = \mathbb{C}_{2, \theta, \theta_k} / \mathbb{C}_2$ with $U_1^0 = H_1^0(C_1)$ as a fixed number and $U_2^0 = Y_2$ as the only random variable. In addition, the integral under this scenario is taken over the region of (Y_1, Y_2) : $\Omega_{01} = [0, H_1^0(C_1)] \times [H_2^0(C_2), 1]$. Thus, the conditional expectation is

$$\begin{aligned} & \int_{[H_2^0(C_2), 1]} \frac{\mathbb{C}_{2, \theta, \theta_k}(H_1^0(C_1), u_2; \boldsymbol{\theta})}{\mathbb{C}_2(H_1^0(C_1), u_2; \boldsymbol{\theta})} \left\{ \int_{[0, H_1^0(C_1)]} \mathbb{C}(u_1, u_2; \boldsymbol{\theta}) du_1 \right\} du_2 \\ &= \int_{[H_2^0(C_2), 1]} \mathbb{C}_{2, \theta, \theta_k}(H_1^0(C_1), u_2; \boldsymbol{\theta}) du_2, \end{aligned}$$

since $\int_{[0, H_1^0(C_1)]} \mathbb{C}(u_1, u_2; \boldsymbol{\theta}) du_1 = \mathbb{C}_2(H_1^0(C_1), u_2; \boldsymbol{\theta})$. Similarly, in $\mathbb{C}_{2, \theta, \theta_k}$, the integral and derivation are interchangeable as follows:

$$\frac{\partial^2}{\partial \theta_j \partial \theta_k} \int_{[0, H_1^0(C_1)]} \mathbb{C}(u_1, u_2; \boldsymbol{\theta}) du_1 = \int_{[0, H_1^0(C_1)]} \mathbb{C}_{\theta, \theta_k}(u_1, u_2; \boldsymbol{\theta}) du_1.$$

Thus, the conditional expectation is $\iint_{\Omega_{01}} \mathbb{C}_{\theta, \theta_k}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2$, which is A_{01} in Equation (17).

- Under the scenario with $\delta_1 = 0$ and $\delta_2 = 0$, $\Delta = \mathbb{C}_{\theta, \theta_k} / \mathbb{C}$ with both $U_1^0 = H_1^0(C_1)$ and $U_2^0 = H_2^0(C_2)$ as fixed numbers. In addition, the integral under this scenario is taken over the region of (Y_1, Y_2) : $\Omega_{00} = [0, H_1^0(C_1)] \times [0, H_2^0(C_2)]$. Thus, the conditional expectation is

$$\begin{aligned} & \frac{\mathbb{C}_{\theta, \theta_k}(H_1^0(C_1), H_2^0(C_2); \boldsymbol{\theta})}{\mathbb{C}(H_1^0(C_1), H_2^0(C_2); \boldsymbol{\theta})} \left\{ \iint_{[0, H_1^0(C_1)] \times [0, H_2^0(C_2)]} \mathbb{C}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2 \right\} \\ &= \mathbb{C}_{\theta, \theta_k}(H_1^0(C_1), H_2^0(C_2); \boldsymbol{\theta}). \end{aligned}$$

Again, in $\mathbb{C}_{\theta, \theta_k}$, the integral and derivations can be interchangeable as follows:

$$\begin{aligned} & \frac{\partial^2}{\partial \theta_j \partial \theta_k} \iint_{[0, H_1^0(C_1)] \times [0, H_2^0(C_2)]} \mathbb{C}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2 \\ &= \iint_{[0, H_1^0(C_1)] \times [0, H_2^0(C_2)]} \mathbb{C}_{\theta, \theta_k}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2. \end{aligned}$$

Thus, the conditional expectation is A_{00} in Equation (18).

Combining all the four censoring statuses, we have

$$\mathbb{E}_{(T_1, T_2)}^0[\Delta_{jk}|C_1, C_2] = \iint_{\Omega_{11} \cup \Omega_{10} \cup \Omega_{01} \cup \Omega_{00}} \mathbb{C}_{\theta, \theta_k}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2.$$

Again, due to the interchangeability between the integral and derivation, we have

$$\mathbb{E}_{(T_1, T_2)}^0[\Delta_{jk}|C_1, C_2] = \frac{\partial^2}{\partial \theta_j \partial \theta_k} \iint_{\Omega_{11} \cup \Omega_{10} \cup \Omega_{01} \cup \Omega_{00}} \mathbb{C}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2.$$

Since $\Omega_{11} \cup \Omega_{10} \cup \Omega_{01} \cup \Omega_{00} = [0, 1]^2$ and $\iint_{[0, 1]^2} \mathbb{C}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2 = 1$ for any $\boldsymbol{\theta}$, we can show that for any $\boldsymbol{\theta}$, $\mathbb{E}_{(Y_1, Y_2)}^0[\Delta_{jk}|C_1, C_2] = 0$ for $j, k = 1, \dots, p$. It implies that $A_{jk}(\boldsymbol{\theta}) = \mathbb{E}_{(C_1, C_2)}^0 \left\{ \mathbb{E}_{(Y_1, Y_2)}^0[\Delta_{jk}|C_1, C_2] \right\} = 0$. Furthermore, evaluating at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, we have $A_{jk}(\boldsymbol{\theta}^*) = 0$, which proves Theorem 1.

Expression of $A_{jk}(\boldsymbol{\theta}^*)$ under copula misspecification. Again, we use the double expectation theorem in Equation (14). If the assumed copula is misspecified, this conditional expectation $\mathbb{E}_{(Y_1, Y_2)}^0[\Delta_{jk}|C_1, C_2]$ is taken w.r.t. $\mathbb{C}^0$. Let $\mathbb{C}_r^0 = \frac{\partial \mathbb{C}^0(u_1, u_2)}{\partial u_r}$, $r = 1, 2$, and $\mathbb{C}^0 = \frac{\partial^2 \mathbb{C}^0(u_1, u_2)}{\partial u_1 \partial u_2}$. Following the above derivations under correct copula specification, we can show that

$$\begin{aligned} \mathbb{E}_{(Y_1, Y_2)}^0[\Delta_{jk}|C_1, C_2] &= \iint_{\Omega_{11}} \mathbb{C}_{\theta, \theta_k}(u_1, u_2; \boldsymbol{\theta}) w_{11}(u_1, u_2; \boldsymbol{\theta}) du_1 du_2 \\ &+ \iint_{\Omega_{10}} \mathbb{C}_{\theta, \theta_k}(u_1, H_2^0(c_2); \boldsymbol{\theta}) w_{10}(u_1, H_2^0(c_2); \boldsymbol{\theta}) du_1 \\ &+ \iint_{\Omega_{01}} \mathbb{C}_{\theta, \theta_k}(H_1^0(c_1), u_2; \boldsymbol{\theta}) w_{10}(H_1^0(c_1), u_2; \boldsymbol{\theta}) du_2 \\ &+ \iint_{\Omega_{00}} \mathbb{C}_{\theta, \theta_k}(H_1^0(c_1), H_2^0(c_2); \boldsymbol{\theta}) w_{00}(H_1^0(c_1), H_2^0(c_2); \boldsymbol{\theta}) du_1 du_2, \end{aligned}$$

where $w_{11}(u_1, u_2; \boldsymbol{\theta}) = \mathbb{C}^0(u_1, u_2) / \mathbb{C}(u_1, u_2; \boldsymbol{\theta})$, $w_{10}(u_1, u_2; \boldsymbol{\theta}) = \mathbb{C}_1^0(u_1, u_2) / \mathbb{C}_1(u_1, u_2; \boldsymbol{\theta})$, $w_{01}(u_1, u_2; \boldsymbol{\theta}) = \mathbb{C}_2^0(u_1, u_2) / \mathbb{C}_2(u_1, u_2; \boldsymbol{\theta})$, and $w_{00}(u_1, u_2; \boldsymbol{\theta}) = \mathbb{C}^0(u_1, u_2) / \mathbb{C}(u_1, u_2; \boldsymbol{\theta})$.

By Definition 1, when the assumed copula is misspecified, there exists some (u_1, u_2) such that $w_{d_1, d_2}(u_1, u_2; \boldsymbol{\theta}^*) \neq 1$ for some $d_1, d_2 = 0, 1$. Thus, for some $j, k = 1, \dots, p$,

$$\mathbb{E}_{(Y_1, Y_2)}^0[\Delta_{jk}(\boldsymbol{\theta}^*, U_1^0, U_2^0)|C_1, C_2] \neq \frac{\partial^2}{\partial \theta_j \partial \theta_k} \iint_{[0, 1]^2} \mathbb{C}(u_1, u_2; \boldsymbol{\theta}^*) du_1 du_2.$$

Consequently, $A_{jk}(\boldsymbol{\theta}^*) \neq 0$ for some $j, k = 1, \dots, p$.

B Proof of Theorem 2

To show $|R_n - \text{tr}[\mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \mathbf{V}^*(\boldsymbol{\theta}^*)]| = o_p(1)$, we need to first prove the consistency of $\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)$ and $\widehat{\mathbf{V}}_n(\widehat{\boldsymbol{\theta}}_n)$. Chen et al. (2010) has shown the consistency of $\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)$

which requires conditions A2 and A4 (i) & (ii) listed in their paper. To prove the consistency of $\widehat{\mathbf{V}}_n(\widehat{\boldsymbol{\theta}}_n)$, our conditions R5 and C5 are analogous to Chen et al.'s those two conditions, respectively. Thus, following the same arguments in their paper, we can show that $\sup_{\boldsymbol{\theta} \in \Theta} n^{-1} \sum_{i=1}^n \|\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \widehat{U}_{i1}, \widehat{U}_{i2}) \ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \widehat{U}_{i1}, \widehat{U}_{i2})' I(X_{ir} \leq \eta)\|$ is asymptotically ignorable as $\eta \rightarrow 0$. This together with the continuity of $\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}, u_1, u_2)$ (our condition R5), and the consistency of the Kaplan-Meier estimate $\widehat{H}_r$ and the PMLE $\widehat{\boldsymbol{\theta}}_n$, leads to $\|\widehat{\mathbf{V}}_n(\widehat{\boldsymbol{\theta}}_n) - \mathbf{V}^*(\boldsymbol{\theta}^*)\| = o_p(1)$.

Our condition C3 (ii) (which is equivalent to Condition A1 (ii) of Chen et al. (2010)) ensures that $\mathbf{S}^*(\boldsymbol{\theta}^*)$ is finite and non-singular. Thus, by Slutsky's Theorem, it implies $\text{tr} [\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)^{-1} \widehat{\mathbf{V}}_n(\widehat{\boldsymbol{\theta}}_n)]$ converges $\text{tr} [\mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \mathbf{V}^*(\boldsymbol{\theta}^*)]$ in probability as $n \rightarrow \infty$.

The proof of Theorem 2 ends.

C Proof of Theorem 3

To prove this theorem, we need to prove the following lemma:

Lemma 1 *Under Conditions R1 - R6 and C1 - C7,*

- (1) $\sqrt{n} \left\{ \widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n) - \mathbf{S}^*(\boldsymbol{\theta}^*) \right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{h}_S(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}) + o_p(1)$, where $\mathbf{h}_S(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2})$ is a $p \times p$ matrix with the (j, k) -th element $h_{S_{jk}}(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2})$ given in Equation (24) being independent random variables with mean 0.
- (2) $\sqrt{n} \left\{ \widehat{\mathbf{V}}_n(\widehat{\boldsymbol{\theta}}_n) - \mathbf{V}^*(\boldsymbol{\theta}^*) \right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{h}_V(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2})$, where $\mathbf{h}_V(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2})$ is a $p \times p$ matrix with the (j, k) -th element $h_{V_{jk}}(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2})$ given in Equation (25) being independent random variables with mean 0.

Proof of Lemma 1. Let $\mathbf{S}_{jk}^*(\boldsymbol{\theta}^*) = \mathbb{E}^0 \left[-\ell_{\theta_j \theta_k}(\boldsymbol{\theta}^*, U_1, U_2) \right]$ denote the (j, k) -th element of $\mathbf{S}^*(\boldsymbol{\theta}^*)$. Similarly, let $\widehat{\mathbf{S}}_{n,jk}(\widehat{\boldsymbol{\theta}}_n) = -n^{-1} \sum_{i=1}^n \ell_{\theta_j \theta_k}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{i1}, \widehat{U}_{i2})$ denote the (j, k) -th element of $\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)$, $j, k = 1, \dots, p$. By the mean-value theorem, we have

$$\widehat{\mathbf{S}}_{n,jk}(\widehat{\boldsymbol{\theta}}_n) = -n^{-1} \sum_{i=1}^n \ell_{\theta_j \theta_k}(\boldsymbol{\theta}^*, \widehat{U}_{i1}, \widehat{U}_{i2}) + \left[-n^{-1} \sum_{i=1}^n \ell_{\theta_j \theta_k, \boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}, \widehat{U}_{i1}, \widehat{U}_{i2}) \right]' (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*),$$

where $\tilde{\boldsymbol{\theta}}$ lies on the linear segment between $\boldsymbol{\theta}^*$ and $\widehat{\boldsymbol{\theta}}_n$.

Using the same arguments for the consistency of $\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)$, by condition C6 (i) (analogous to Chen et al.'s condition A2) and condition C6 (ii) & (iii) (analogous to Chen et al.'s Condition A4), we can show $\sup_{\boldsymbol{\theta} \in \Theta} n^{-1} \sum_{i=1}^n \|\ell_{\theta_j \theta_k, \boldsymbol{\theta}}(\boldsymbol{\theta}, \widehat{U}_{i1}, \widehat{U}_{i2}) I(X_{ir} \leq \eta)\|$ is asymptotically ignorable as $\eta \rightarrow 0$. This together with the continuity of $\ell_{\theta_j \theta_k, \boldsymbol{\theta}}$ (in our condition C6 (i)) and the consistency of the Kaplan-Meier estimate and PMLE $\widehat{\boldsymbol{\theta}}_n$, we can show that $\|n^{-1} \sum_{i=1}^n \ell_{\theta_j \theta_k, \boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}, \widehat{U}_{i1}, \widehat{U}_{i2}) - \mathbb{E}^0 [\ell_{\theta_j \theta_k, \boldsymbol{\theta}}(\boldsymbol{\theta}^*, U_1^0, U_2^0)]\| = o_p(1)$. Let $\mathbf{M}_{jk}(\boldsymbol{\theta}^*) = \mathbb{E}^0 [\ell_{\theta_j \theta_k, \boldsymbol{\theta}}(\boldsymbol{\theta}^*, U_1^0, U_2^0)]$ (a $p \times 1$ vector). Because $\widehat{\boldsymbol{\theta}}_n$ is $\sqrt{n}$ -consistent, we

have

$$\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)_{jk} = -n^{-1} \sum_{i=1}^n \ell_{\theta_j \theta_k}(\boldsymbol{\theta}^*, \widehat{U}_{i1}, \widehat{U}_{i2}) - \mathbf{M}_{jk}(\boldsymbol{\theta}^*)'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) + o_p(n^{-1/2}). \quad (19)$$

Again applying the mean-value theorem on Equation (19), we have

$$\begin{aligned} & \widehat{\mathbf{S}}_{n,jk}(\widehat{\boldsymbol{\theta}}_n) - \mathbf{S}_{jk}^*(\boldsymbol{\theta}^*) \\ &= n^{-1} \sum_{i=1}^n \left[-\ell_{\theta_j \theta_k}(\boldsymbol{\theta}^*, U_{i1}^0, U_{i2}^0) - \mathbf{S}_{jk}^*(\boldsymbol{\theta}^*) \right] - n^{-1} \sum_{r=1}^2 \sum_{i=1}^n \ell_{\theta_j \theta_k, u_r}(\boldsymbol{\theta}^*, \tilde{U}_{i1}, \tilde{U}_{i2})(\widehat{U}_{ir} - U_{ir}^0) \\ & \quad - \mathbf{M}_{jk}(\boldsymbol{\theta}^*)'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) + o_p(n^{-1/2}) \end{aligned} \quad (20)$$

where $(\tilde{U}_{i1}, \tilde{U}_{i2})$ lies on the linear segment between $(\widehat{U}_{i1}, \widehat{U}_{i2})$ and (U_{i1}^0, U_{i2}^0) .

Based on the expansion of $\widehat{\boldsymbol{\theta}}_n$ around $\boldsymbol{\theta}^*$ in Chen et al. (2010), we have

$$\begin{aligned} & \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^* \\ &= \mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \frac{1}{n} \sum_{i=1}^n [\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}^*, U_{i1}^0, U_{i2}^0) + W_1(\boldsymbol{\theta}^*, X_{i1}, \delta_{i1}) + W_2(\boldsymbol{\theta}^*, X_{i2}, \delta_{i2})] + o_p(n^{-1/2}) \end{aligned} \quad (21)$$

where for $r = 1, 2$,

$$W_r(\boldsymbol{\theta}^*, X_{ir}, \delta_{ir}) = \mathbb{E}^0 [\ell_{\boldsymbol{\theta}, u_r}(\boldsymbol{\theta}^*, U_1^0, U_2^0) I_{ir}(X_r) \mid X_{ir}, \delta_{ir}] \quad (22)$$

with

$$I_{ir}(X_r) = -H_r(X_r) \left[\int_{-\infty}^{X_r} \frac{dN_{ir}(u)}{P_{n,r}(u)} - \int_{-\infty}^{X_r} \frac{I(X_{ir} \geq u) d\Lambda_r^0(u)}{P_{n,r}(u)} \right]$$

with $\Lambda_r^0(u) = -\log H_r^0(u)$, the true cumulative hazard function of T_{ir} , $N_{ir}(u) = \delta_{ir} I(X_{ir} \leq u)$, $dN_{ir}(u) = N_{ir}(u) - N_{ir}(u-)$, and $P_{n,r}(u) = n^{-1} \sum_{k=1}^n Pr(X_{kr} \geq u)$. Using similar arguments for obtaining Equation (21), under our condition R6 (iii) and C7 (i) (analogous to Chen et al.'s condition A3 (i) & (ii)) and condition C6 (ii) (analogous to Chen et al.'s Condition A4 (i)), we can show that

$$n^{-1} \sum_{i=1}^n \ell_{\theta_j \theta_k, u_r}(\boldsymbol{\theta}^*, \tilde{U}_{i1}, \tilde{U}_{i2})(\widehat{U}_{ir} - U_{ir}^0) = n^{-1} \sum_{i=1}^n h_{S_{jk,r}}(\boldsymbol{\theta}^*, X_{ir}, \delta_{ir}) + o_p(n^{-1/2}), \quad (23)$$

where $h_{S_{jk,r}}(\boldsymbol{\theta}^*, X_{ir}, \delta_{ir}) = \mathbb{E}^0 [\ell_{\theta_j \theta_k, u_r}(\boldsymbol{\theta}^*, U_1^0, U_2^0) I_{ir}(X_r) \mid X_{ir}, \delta_{ir}]$.

From Equations (20), (21), and (23), we have

$$\sqrt{n} \left\{ \widehat{\mathbf{S}}_{n,jk}(\widehat{\boldsymbol{\theta}}_n) - \mathbf{S}_{jk}^*(\boldsymbol{\theta}^*) \right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^n h_{S_{jk}}(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}) + o_p(1),$$

where

$$\begin{aligned}
& h_{S_{jk}}(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}) \\
&= \left[-\ell_{\theta_j, \theta_k}(\boldsymbol{\theta}^*, U_{i1}^0, U_{i2}^0) - \mathbf{S}_{jk}^*(\boldsymbol{\theta}^*) \right] - h_{S_{jk,1}}(\boldsymbol{\theta}^*, X_{i1}, \delta_{i1}) - h_{S_{jk,2}}(\boldsymbol{\theta}^*, X_{i2}, \delta_{i2}) \\
&\quad - \mathbf{M}_{jk}(\boldsymbol{\theta}^*)' \mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} [\ell_{\theta}(\boldsymbol{\theta}^*, U_{i1}, U_{i2}) + W_1(\boldsymbol{\theta}^*, X_{i1}, \delta_{i1}) + W_2(\boldsymbol{\theta}^*, X_{i2}, \delta_{i2})] \quad (24)
\end{aligned}$$

Let $\mathbf{V}_{jk}^*(\boldsymbol{\theta}^*)$ and $\widehat{\mathbf{V}}_{n,jk}(\boldsymbol{\theta}^*)$ denote the (j, k) -th element of $\mathbf{V}^*(\boldsymbol{\theta}^*)$ and $\widehat{\mathbf{V}}_n(\boldsymbol{\theta}^*)$. We apply the same techniques above, with our condition C5 (i) (analogous to Chen et al.'s condition A4 (i)) and condition C7 (ii) & (iii) (analogous to Chen et al.'s condition A3 (i) & (ii)), we can show

$$\sqrt{n} \left\{ \widehat{\mathbf{V}}_{n,jk}(\widehat{\boldsymbol{\theta}}_n) - \mathbf{V}_{jk}^*(\boldsymbol{\theta}^*) \right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^n h_{V_{jk}}(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}) + o_p(1),$$

where

$$\begin{aligned}
& h_{V_{jk}}(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}) \\
&= \left[\ell_{\theta_j}(\boldsymbol{\theta}^*, U_{i1}^0, U_{i2}^0) \ell_{\theta_k}(\boldsymbol{\theta}^*, U_{i1}^0, U_{i2}^0)' - \mathbf{V}_{jk}^*(\boldsymbol{\theta}^*) \right] + h_{V_{jk,1}}(\boldsymbol{\theta}^*, X_{i1}, \delta_{i1}) + h_{V_{jk,2}}(\boldsymbol{\theta}^*, X_{i2}, \delta_{i2}) \\
&\quad + \mathbf{P}_{jk}(\boldsymbol{\theta}^*)' \mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} [\ell_{\theta}(\boldsymbol{\theta}^*, U_{i1}^0, U_{i2}^0) + W_1(\boldsymbol{\theta}^*, X_{i1}, \delta_{i1}) + W_2(\boldsymbol{\theta}^*, X_{i2}, \delta_{i2})] \quad (25)
\end{aligned}$$

with

$$\begin{aligned}
h_{V_{jk,r}}(\boldsymbol{\theta}^*, X_{ir}, \delta_{ir}) &= \mathbb{E}^0 \left\{ \left[\ell_{\theta_j, u_r}(\boldsymbol{\theta}^*, U_1^0, U_2^0) \ell_{\theta_k}(\boldsymbol{\theta}^*, U_1^0, U_2^0) \right. \right. \\
&\quad \left. \left. + \ell_{\theta_k, u_r}(\boldsymbol{\theta}^*, U_1^0, U_2^0) \ell_{\theta_j}(\boldsymbol{\theta}^*, U_1^0, U_2^0) \right] * I_{ir}(X_r) \mid X_{ir}, \delta_{ir} \right\}
\end{aligned}$$

and $\mathbf{P}_{jk}(\boldsymbol{\theta}^*) = \mathbb{E}^0 [\ell_{\theta_j, \theta}(\boldsymbol{\theta}^*, U_1^0, U_2^0) \ell_{\theta_k}(\boldsymbol{\theta}^*, U_1^0, U_2^0) + \ell_{\theta_k, \theta}(\boldsymbol{\theta}^*, U_1^0, U_2^0) \ell_{\theta_j}(\boldsymbol{\theta}^*, U_1^0, U_2^0)]$.

The proof of Lemma 1 ends.

Proof of Theorem 3: Under the null hypothesis that the assumed copula function is correctly specified, $\mathbf{R}^*(\boldsymbol{\theta}^*) = \mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \mathbf{V}^*(\boldsymbol{\theta}^*) = I_p$ due to Theorem 1, and consequently, by Theorem 2, $R_n \rightarrow p = \text{tr}[I_p]$ in probability as $n \rightarrow \infty$. In addition, $R_n - p$ can be expressed as $R_n - p = \text{tr}[\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)^{-1} \widehat{\mathbf{V}}_n(\widehat{\boldsymbol{\theta}}_n) - \mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \mathbf{V}^*(\boldsymbol{\theta}^*)]$. With algebraic derivations, we have

$$\begin{aligned}
& \sqrt{n}(R_n - p) = \sqrt{n} \text{tr} \left[\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)^{-1} \widehat{\mathbf{V}}_n(\widehat{\boldsymbol{\theta}}_n) - \mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \mathbf{V}^*(\boldsymbol{\theta}^*) \right] \\
&= \text{tr} \left[\mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \sqrt{n} \left\{ \widehat{\mathbf{V}}_n(\widehat{\boldsymbol{\theta}}_n) - \mathbf{V}^*(\boldsymbol{\theta}^*) \right\} \right] \\
&\quad + \text{tr} \left[\mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \mathbf{V}^*(\boldsymbol{\theta}^*) \mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \sqrt{n} \left\{ \mathbf{S}^*(\boldsymbol{\theta}^*) - \widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n) \right\} \right] \\
&\quad + \text{tr} \left[\left\{ \widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)^{-1} \widehat{\mathbf{V}}_n(\widehat{\boldsymbol{\theta}}_n) - \mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \mathbf{V}^*(\boldsymbol{\theta}^*) \right\} \mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \sqrt{n} \left\{ \mathbf{S}^*(\boldsymbol{\theta}^*) - \widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n) \right\} \right] \quad (26)
\end{aligned}$$

Under the null hypothesis, $\mathbf{S}^*(\boldsymbol{\theta}^*) = \mathbf{V}^*(\boldsymbol{\theta}^*)$, the second term in Equation (26) becomes $\text{tr}[\mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \sqrt{n} \left\{ \mathbf{S}^*(\boldsymbol{\theta}^*) - \widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n) \right\}]$. The third term in Equation (26) is $o_p(1)$

because $\|\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)^{-1}\widehat{\mathbf{V}}_n(\boldsymbol{\theta}^*) - \mathbf{S}^*(\boldsymbol{\theta}^*)^{-1}\mathbf{V}^*(\boldsymbol{\theta}^*)\| = o_p(1)$ shown in the proof of Theorem 2 (Appendix B) and $\|\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n) - \mathbf{S}^*(\boldsymbol{\theta}^*)\| = O_p(n^{-1/2})$ by Lemma 1. Thus, we can write

$$\sqrt{n}(R_n - p) = \frac{1}{\sqrt{n}} \sum_{i=1}^n h_R(X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}, \boldsymbol{\theta}) + o_p(1),$$

where

$$h_R(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}) = \text{tr} \left[\mathbf{S}^*(\boldsymbol{\theta}^*)^{-1} \{h_V(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}) - h_S(\boldsymbol{\theta}^*, X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2})\} \right]. \quad (27)$$

By Central Limit Theorem for independent random variables, we can show that $\sqrt{n}(R_n - p)$ converges in distribution to a normal random variable with mean 0 and variance $\sigma^2 = \text{Var}[h_R(X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2}, \boldsymbol{\theta})]$.

The proof of Theorem 3 ends.

D Proof of Theorem 4

To prove this theorem, we need to first prove the following lemma:

Lemma 2 Under Condition R1 - R6 and C1 - C4, $\sup_{1 \leq i \leq n} \|\widehat{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_{(-i)}\| = O_p(n^{-1})$.

Proof of Lemma 2 The "out-of-sample" PMLE $\widehat{\boldsymbol{\theta}}_{(-i)}$ is obtained by maximizing $\sum_{s=1, s \neq i}^n \ell(\boldsymbol{\theta}, \widehat{U}_{s1}, \widehat{U}_{s2})$, i.e., $\sum_{s=1, s \neq i}^n \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_{(-i)}, \widehat{U}_{s1}, \widehat{U}_{s2}) = 0$. Apply the mean-value theorem, we have

$$\begin{aligned} 0 &= \sum_{s=1, s \neq i}^n \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_{(-i)}, \widehat{U}_{s1}, \widehat{U}_{s2}) = \sum_{s=1, s \neq i}^n \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{s1}, \widehat{U}_{s2}) + \sum_{s=1, s \neq i}^n \ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\widetilde{\boldsymbol{\theta}}, \widehat{U}_{s1}, \widehat{U}_{s2})(\widehat{\boldsymbol{\theta}}_{(-i)} - \widehat{\boldsymbol{\theta}}_n) \\ &= \sum_{s=1}^n \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{s1}, \widehat{U}_{s2}) - \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{i1}, \widehat{U}_{i2}) + \sum_{s=1, s \neq i}^n \ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\widetilde{\boldsymbol{\theta}}, \widehat{U}_{s1}, \widehat{U}_{s2})(\widehat{\boldsymbol{\theta}}_{(-i)} - \widehat{\boldsymbol{\theta}}_n) \end{aligned}$$

where $\widetilde{\boldsymbol{\theta}}$ lies in the linear segment between $\widehat{\boldsymbol{\theta}}_{(-i)}$ and $\widehat{\boldsymbol{\theta}}_n$. Since $\sum_{s=1}^n \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{s1}, \widehat{U}_{s2}) = 0$ ($\widehat{\boldsymbol{\theta}}_n$ is the PMLE using all the observations), we have

$$\widehat{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_{(-i)} = \widehat{\mathbf{S}}_{(-i)}(\widetilde{\boldsymbol{\theta}})^{-1} n^{-1} \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{i1}, \widehat{U}_{i2}) \quad (28)$$

where $\widehat{\mathbf{S}}_{(-i)}(\widetilde{\boldsymbol{\theta}}) = -n^{-1} \sum_{s=1, s \neq i}^n \ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\widetilde{\boldsymbol{\theta}}, \widehat{U}_{s1}, \widehat{U}_{s2})$. Thus,

$$\sup_{1 \leq i \leq n} \|\widehat{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_{(-i)}\| \leq n^{-1} \sup_{1 \leq i \leq n} \left\| \widehat{\mathbf{S}}_{(-i)}(\widetilde{\boldsymbol{\theta}})^{-1} \right\| \times \sup_{1 \leq i \leq n} \left\| \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{i1}, \widehat{U}_{i2}) \right\|.$$

Using the same arguments for proving the consistency of $\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}})$, we can prove that as $n \rightarrow \infty$, $\widehat{\mathbf{S}}_{(-i)}(\widetilde{\boldsymbol{\theta}}) \rightarrow \mathbf{S}^*(\boldsymbol{\theta}^*)$ in probability. Our condition C3 (ii) (equivalent to Chen et al.'s Condition A1 (ii)) assumes the boundedness for the eigenvalues of $\mathbf{S}^*(\boldsymbol{\theta}^*)$,

which ensures that $\sup_{1 \leq i \leq n} \|\widehat{\mathbf{S}}_{(-i)}(\tilde{\boldsymbol{\theta}})^{-1}\| < \infty$. In addition, our condition R6 (i) & (ii) (equivalent to Chen et al.'s Condition A3) ensures that $\sup_{1 \leq i \leq n} \|\ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{i1}, \widehat{U}_{i2})\| = O_p(1)$, and thus, $\sup_{1 \leq i \leq n} \|n^{-1} \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{i1}, \widehat{U}_{i2})\| = O_p(1)$. It leads to $\sup_{1 \leq i \leq n} \|\widehat{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_{(-i)}\| = O_p(n^{-1})$.

The proof of Lemma 2 ends.

Proof of Theorem 4. Recall that the PIOS test statistic is defined as

$$T_n = \sum_{i=1}^n \ell(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{i1}, \widehat{U}_{i2}) - \sum_{i=1}^n \ell(\widehat{\boldsymbol{\theta}}_{(-i)}, \widehat{U}_{i1}, \widehat{U}_{i2}).$$

Applying the mean value theorem on $\ell(\widehat{\boldsymbol{\theta}}_{(-i)}, \widehat{U}_{i1}, \widehat{U}_{i2})$, we have

$$T_n = - \sum_{i=1}^n \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{i1}, \widehat{U}_{i2})' (\widehat{\boldsymbol{\theta}}_{(-i)} - \widehat{\boldsymbol{\theta}}_n) - \frac{1}{2} \sum_{i=1}^n \ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\check{\boldsymbol{\theta}}, \widehat{U}_{i1}, \widehat{U}_{i2}) (\widehat{\boldsymbol{\theta}}_{(-i)} - \widehat{\boldsymbol{\theta}}_n)^2,$$

where $\check{\boldsymbol{\theta}}$ lies on the linear segment between $\widehat{\boldsymbol{\theta}}_{(-i)}$ and $\widehat{\boldsymbol{\theta}}_n$. Plugging in Equation (28), we have

$$\begin{aligned} T_n &= n^{-1} \sum_{i=1}^n \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{i1}, \widehat{U}_{i2})' \left\{ \widehat{\mathbf{S}}_{(-i)}(\tilde{\boldsymbol{\theta}}) \right\}^{-1} \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{i1}, \widehat{U}_{i2}) \\ &\quad - \frac{1}{2} \sum_{i=1}^n \ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\check{\boldsymbol{\theta}}, \widehat{U}_{i1}, \widehat{U}_{i2}) (\widehat{\boldsymbol{\theta}}_{(-i)} - \widehat{\boldsymbol{\theta}}_n)^2 \\ &= \text{tr} \left[\widehat{\mathbf{S}}_{(-i)}(\tilde{\boldsymbol{\theta}})^{-1} \left\{ n^{-1} \sum_{i=1}^n \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{i1}, \widehat{U}_{i2}) \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_n, \widehat{U}_{i1}, \widehat{U}_{i2})' \right\} \right] \\ &\quad - \frac{1}{2} \left\{ n^{-1} \sum_{i=1}^n \ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\check{\boldsymbol{\theta}}, \widehat{U}_{i1}, \widehat{U}_{i2}) \right\} n (\widehat{\boldsymbol{\theta}}_{(-i)} - \widehat{\boldsymbol{\theta}}_n)^2. \end{aligned}$$

Thus,

$$\begin{aligned} T_n - R_n &= \text{tr} \left[\left\{ \widehat{\mathbf{S}}_{(-i)}(\tilde{\boldsymbol{\theta}})^{-1} - \widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)^{-1} \right\} \widehat{\mathbf{V}}_n(\widehat{\boldsymbol{\theta}}) \right] - \frac{1}{2} \left\{ n^{-1} \sum_{i=1}^n \ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\check{\boldsymbol{\theta}}, \widehat{U}_{i1}, \widehat{U}_{i2}) \right\} n (\widehat{\boldsymbol{\theta}}_{(-i)} - \widehat{\boldsymbol{\theta}}_n)^2. \end{aligned} \tag{29}$$

In the proof of Lemma 2, we have shown that $\|\widehat{\mathbf{S}}_{(-i)}(\tilde{\boldsymbol{\theta}}) - \mathbf{S}^*(\boldsymbol{\theta}^*)\| = o_p(1)$. In addition, because $\|\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n) - \mathbf{S}^*(\boldsymbol{\theta}^*)\| = o_p(1)$, we have $\|\widehat{\mathbf{S}}_{(-i)}(\tilde{\boldsymbol{\theta}}) - \widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)\| = o_p(1)$, and consequently, the first term in Equation (29) is $o_p(1)$. For the second term, following similar arguments, we can show $\|n^{-1} \sum_{i=1}^n \ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\check{\boldsymbol{\theta}}, \widehat{U}_{i1}, \widehat{U}_{i2}) - \mathbf{S}^*(\boldsymbol{\theta}^*)\| = o_p(1)$. Together with $\sup_{1 \leq i \leq n} \|\widehat{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_{(-i)}\| = O_p(n^{-1})$, the second term is $O_p(n^{-1})$. Combining the two terms, we have $|T_n - R_n| = o_p(1)$.

The proof of Theorem 4 ends.