Copula Specification Tests under General Censorship

Juan Lin∗  Ximing Wu†

October 15, 2015

Abstract

We propose a family of data-driven tests for the goodness-of-fit of copula-based multivariate survival models under general censorship. Appealing features of the tests include flexibility, ease of implementation, distribution free asymptotic distributions and informativeness regarding alternative copulas when a null distribution is rejected. Consistency and large sample properties of the tests, with parametrically or nonparametrically estimated marginal distributions, are established. Monte Carlo simulations demonstrate good finite sample performance of the proposed tests. The semiparametric tests are shown to rival the correctly specified parametric tests and at the same time are immune from risk of misspecification. Two empirical applications illustrate the usefulness of the proposed tests.

Keywords: copula specification; smooth test; censorship; survival analysis

JEL Classification: C12; C14; C34

∗Department of Finance, School of Economics & Wang Yanan Institute for Studies in Economics, Xiamen University, China. Email: linjuan.bnu@gmail.com
†Corresponding author. Department of Agricultural Economics, Texas A&M University, College Station, TX, 77843, U.S.A. Tel: (979) 458-1355, Fax: (979) 862-1563, Email: xwu@tamu.edu
1 Introduction

Copula is a popular tool to model the dependence structure of multivariate data. One major advantage of the copula approach is its flexibility in separately modeling the marginal distributions of individual variables and their dependence structure, cf. Cherubini et al. (2004, 2011), McNeil et al. (2005), Patton (2012) and references therein. It has been widely appreciated that the specification of copula models is of great importance since different copulas lead to multivariate models that may have very different dependence properties. For instance, the Gaussian copula has been widely used to model the collateralized debt obligation (CDO) due to its simplicity. However, it imposes the restriction that correlated events are asymptotically independent such that it cannot accommodate tail dependence during times of financial turmoil. Due to this drawback, it has been vilified as “the formula that killed Wall Street” (Salmon, 2009). A number of studies have attempted to address the issue of copula specification. For general overviews of copula specification tests for complete (i.e., uncensored) data, see among others Chen and Fan (2005, 2006), Genest et al. (2009), Berg (2009) and Fermanian (2013).

In practice, censored data are frequently encountered in many areas, such as actuarial sciences, biology, economics and finance; cf. Frees et al. (1996), Klugman and Parsa (1999), Denuit et al. (2006) among others. For example, in the field of credit risk, a firm’s lifetime is censored if either default does not occur by the end of the observation period, or if the firm drops out of the sample for reasons other than default such as merger, acquisition or liquidation. Another example of censored data is the insurance losses, which are usually bounded from above because the amount of claim cannot exceed the insured’s policy limit. In labor economics, high earnings are typically topcoded (right censored) in order to protect the confidentiality of respondents. Duration data are often censored because an unemployment spell might still be continuing at the end of the data collection period.

There is a small body of work that extends copula specification tests to censored data, with a focus on bivariate Archimedean copulas. Extending the probability integral transformation approach proposed by Genest and Rivest (1993), Wang and Wells (2000) proposed a goodness-of-fit test for the correct specification of an Archimedean copula by comparing nonparametric and model-based estimators. The unique structure of Archimedean copula allows the construction of test statistics that are based on univariate empirical processes and facilitate the tests considerably. Lakhal-Chaieb (2010) further modified Wang and Wells’s (2000) test. Emura et al. (2010) proposed a test by comparing the discrepancy between the unweighted and weighted concordance estimators of the associated Archimedean copula parameter. Chen and Fan (2007) proposed a model selection test for Archimedean copula
families. A few studies have also considered non-Archimedean copulas for censored data. Andersen et al. (2005) proposed a test that partitions the unit square and constructs a \( \chi^2 \)-square discrepancy in the partition frequencies according to a parametrically estimated copula and the empirical copula. Chen et al. (2010) proposed a likelihood-based criterion of copula selection that accommodates general copulas of two and higher dimensions.

In this paper we propose two flexible copula specification tests for censored data. Our tests are based on Neyman’s (1937) smooth test of distributions. This smooth test has been developed in many directions. For instance, Gray and Pierce (1985) extended it to censored data; Ledwina (1994) introduced a data driven version; Kallenberg and Ledwina (1999) generalized it to the test of bivariate independence; Lin and Wu (2015) applied it to copula specification for complete data. This study generalizes Lin and Wu (2015) in two directions. (i) We extend the smooth tests of copula specification to tackle censored data. We consider general censoring mechanisms, including the simple random censoring, fixed censoring and no censoring. Unlike most of the existing tests, the proposed tests accommodate both Archimedean and non-Archimedean copulas. (ii) Lin and Wu (2015) considered smooth tests with fixed designs. Instead we utilize data driven principles to construct the proposed tests and establish their consistency. The smooth test embeds the null distribution in a more general alternative parametrized by a set of ‘characterizing’ moment functions. It is diagnostic and conductive to improved copula specifications: if the null hypothesis is rejected under a selected set of moment functions, an alternative copula can be subsequently constructed by augmenting the null distribution with these additional moment functions in the spirit of Efron and Tibshirani (1996).

Copula estimation typically requires the estimation of marginal distributions. We consider both parametrically and non-parametrically estimated marginal distributions in the construction of the proposed tests. For both types of tests, we establish their consistency and large sample distributions, which are distribution free and can be tabulated via simple simulations. Our Monte Carlo simulations demonstrate good finite sample size and power performance of these tests. In particular, we find that the semiparametric tests are comparable to the parametric tests with correctly specified marginal distributions; at the same time, they are immune from the risk of misspecification of the marginal distributions, which may compromise the parametric tests. We apply the proposed tests to two real world examples with censored data.

The remainder of this paper is organized as follows. Section 2 briefly reviews Neyman’s smooth test of distribution and several extensions pertinent to this study. Section 3 proposes copula smooth tests for censored data under simple hypotheses without unknown parameters.
In Section 4, we present general copula smooth tests with estimated marginal distributions and copula parameters and derive their large sample properties. Finite sample performances of the proposed tests are investigated in section 5 through Monte Carlo simulations. Section 6 provides two empirical applications and the last section briefly concludes. All proofs are relegated to the appendix.

Throughout the paper, we use upper case letters to denote cumulative distribution functions and corresponding lower case letters to denote density functions. We use subscript \( t \) to index observations and subscript \( j \) to denote the coordinate of multivariate random vectors. We use \( I_k \) to denote an \( k \times k \) identity matrix and a prime symbol to indicate matrix transpose. For simplicity, we focus on bivariate copulas in this study. Extensions to higher dimensional copulas are straightforward.

2 Background

Neyman (1937) introduced the smooth tests of goodness of fit. The main thrust of smooth tests is to embed a null density into a larger exponential family of alternatives, transforming the hypothesis regarding a certain distribution into that of a general distribution against a nested special case. Gray and Pierce (1985) extended Neyman’s smooth test to accommodate censored data. Denote by \( T \) the survival time of interest from an unknown continuous distribution and by \( D \) a censoring variable. We assume that \( D \) is independent of \( T \). Suppose \( n \) independently (but not necessarily identically) distributed observations \( \{(X_t, \delta_t)\}_{t=1}^{n} \) are available, where \( X_t = T_t \wedge D_t \) and \( \delta_t = I(T_t \leq D_t) \), in which \( a \wedge b = \min(a, b) \) for real numbers \( a \) and \( b \) and \( I(\cdot) \) is the indicator function. The hypothesis of interest is whether a random sample \( T_1, ..., T_n \) are generated from a specific distribution \( F(t; \theta) \), where \( \theta \) is a finite-dimensional vector of unknown parameters. Let \( f(t; \theta) \) and \( S(t; \theta) \) be the corresponding density and survival function respectively. Under the null hypothesis, \( \{S(T_t; \theta)\}_{t=1}^{n} \) are distributed according to the standard uniform distribution. Gray and Pierce (1985) considered a family of alternative densities for \( T \) of the form

\[
f(t; \theta) \exp(\lambda' \psi(S(t; \theta)) - \lambda_0),
\]

where \( \psi = (\psi_1, ..., \psi_k)' \) is a vector of linearly independent, bounded real valued functions defined on \([0, 1]\), \( \lambda = (\lambda_1, ..., \lambda_k)' \in \mathbb{R}^k \), and \( \lambda_0 \) is a normalization constant.\(^1\) Neyman’s smooth test of distribution then amounts to the test of \( \lambda = 0 \).

\(^1\)In a similar spirit, Peña (1998) considered a smooth goodness-of-fit test for hazard functions, embedding a baseline hazard function in a larger family of hazard functions.
Define $u_t = S(X_t; \theta)$. The average log-likelihood of the data under alternative (1) is given by

$$
\frac{1}{n} \sum_{t=1}^{n} \left\{ \delta_t [\log f(X_t; \theta) + X' \psi(u_t)] + (1 - \delta_t) \log \int_{0}^{u_t} \exp(X' \psi(v))dv \right\} - \lambda_0.
$$

(2)

Let $\hat{\theta}$ be the maximum likelihood estimate for $\theta$ and $\hat{u}_t = S(X_t; \hat{\theta})$. The score of (2) with respect to $\lambda$, evaluated at $\lambda = 0$, is given by

$$\hat{\Psi} = \frac{1}{n} \sum_{t=1}^{n} \Psi(\hat{u}_t, \delta_t),$$

with

$$\Psi(u_t, \delta_t) = \delta_t \psi(u_t) + (1 - \delta_t) \int_{0}^{u_t} \psi(u)du / u_t - \int_{0}^{1} \psi(u)du$$

$$= \delta_t \psi(u_t) + (1 - \delta_t)E[\psi(U)|U < u_t] - E[\psi(U)],$$

where $U$ is a standard uniform random variable. It is seen that the uncensored observations contribute to the score function in the usual form of $\psi(S(X_t; \hat{\theta}))$, while contributions from the censored observations take the form $E[\psi(S(T; \hat{\theta}))|T > X_t]$. Thus, $\hat{\Psi}$ yields the conditional expectation of the score function for the complete data $\{T_t\}_{t=1}^{n}$ given the observed data $\{(X_t, \delta_t)\}_{t=1}^{n}$. This is similar to the idea of the EM algorithm for censored data proposed by Dempster et al. (1977).

It is straightforward to show that under the null hypothesis, $n^{-1} \sum_{t=1}^{n} E[\Psi(\hat{u}_t, \delta_t)] = 0$. Let $V = n^{-1} \sum_{t=1}^{n} \text{var}[\Psi(\hat{u}_t, \delta_t)]$. It follows that under the null hypothesis, $\hat{\Psi} \overset{p}{\to} 0$ and $\sqrt{n} V^{-1/2} \hat{\Psi} \overset{d}{\to} N(0, I_k)$ as $n \to \infty$ under suitable regularity conditions. Let $\hat{V}$ be a consistent estimate of $V$. Neyman’s smooth test of distribution for censored data is then given by

$$N_k = n \hat{\Psi}' \hat{V}^{-1} \hat{\Psi}.$$ 

Under the null hypothesis, $N_k$ converges in distribution to the $\chi^2$ distribution with $k$ degrees of freedom as $n \to \infty$.

Compared to the omnibus tests (e.g. the Komogorov-Smirnov test or Cramér-von Mises test), the smooth tests have attractive finite sample properties (see Rayner and Best (1990) for a comprehensive review of the smooth tests). In practice, caution should be exercised regarding the choice of the number of basis functions $k$, which influences the power of smooth tests. A customary practice is to use a small $k$, usually less than 4. Alternatively Led-
wina (1994) proposed a data driven smooth test. She suggested first applying the Bayesian Information Criterion (BIC) to choose a suitable distribution within family (1) and then constructing the test statistic based on the chosen distribution. In an extension to Ledwina (1994), Inglot and Ledwina (2006) considered a hybrid selection criterion, which combines the strength of the BIC and Akaike Information Criterion (AIC) and is shown to improve the power.

Kallenberg and Ledwina (1999) and Kallenberg (2009) proposed smooth tests for bivariate distributions. Let $\psi_{i_1, i_2}(v_1, v_2) = \psi_{i_1}(v_1)\psi_{i_2}(v_2)$ and $\Psi$ be a non-empty subset of $\{\psi_{i_1, i_2} : 1 \leq i_1, i_2 \leq M\}$. Denote the cardinality of $\Psi$ by $k = |\Psi|$ and write $\Psi = \{\Psi_1, ..., \Psi_k\}$. They considered a bivariate version of (1) for the joint distribution of the empirical distributions of two random variables,

$$
\exp\left(\sum_{i=1}^{k} \lambda_i \Psi_i(v_1, v_2) - \lambda_0\right), \quad (v_1, v_2) \in [0, 1]^2,
$$

where $\lambda_0$ is a normalization constant. The test of independence within the exponential family (3) is equivalent to a test of the hypothesis: $\lambda = (\lambda_1, \ldots, \lambda_k)' = 0$. A smooth test of independence can then be constructed analogously to the test of univariate uniformity.

In their study of bivariate distributions, Kallenberg and Ledwina (1999) suggested two specifications of $\psi_{i_1, i_2}$. The first method considers only ‘diagonal’ entries such as $\psi_{11}, \psi_{22}, \ldots$ and so on, to which the BIC truncation rule is applied. The second method is more flexible and allows ‘non-diagonal’ entries. In particular, let $\{Z_{1,t}, Z_{2,t}\}_{t=1}^{n}$ be an i.i.d. sample and $\{\hat{V}_{1,t}, \hat{V}_{2,t}\}_{t=1}^{n}$ their corresponding empirical distributions. Define $\hat{\psi}_{i_1, i_2} = n^{-1} \sum_{t=1}^{n} \psi_{i_1, i_2}(\hat{V}_{1,t}, \hat{V}_{2,t})$, $i_1, i_2 \in (1, \ldots, M)$. The set $\Psi$ is arranged such that $\Psi_1 = \psi_{11}$ and the remaining entries are arranged in the descending order according to the absolute value of $\hat{\psi}_{i_1, i_2}$. The BIC truncation rule is then applied to this ordered set.\(^2\) Denote by $K$ the number of moment functions selected according to the BIC. The test statistic is then given by $T_K = n \sum_{i=1}^{K} \hat{\Psi}_i^2$.

Kallenberg and Ledwina (1999) showed that $K \xrightarrow{p} 1$ and $T_K \xrightarrow{d} \chi^2$ under independence as $n \to \infty$.

\(^2\)In a similar spirit, Kallenberg (2009) used a threshold rule to select significant elements of $\Psi$ for copula specifications.
3 Smooth tests of copula specification for censored data: simple hypothesis

To fix idea, in this section we introduce the smooth copula specification tests for censored data under the simplifying assumption of no unknown parameters. The general tests under unknown marginal distributions and copula parameters are given in the next section.

A copula is a multivariate probability distribution with standard uniform marginal distributions. Copulas describe the dependence between random variables. Suppose \((T_1, T_2)\) is a paired survival times with joint survival function \(S(t_1, t_2)\) and density function \(f(t_1, t_2)\). Let \((S_1, S_2)\) and \((f_1, f_2)\) denote the corresponding continuous marginal survival functions and density functions for \((T_1, T_2)\). A straightforward application of Sklar's (1959) Theorem yields that there exists a unique copula function \(C(\cdot, \cdot)\) with density function \(c(\cdot, \cdot)\) on \([0, 1]^2\), such that

\[
S(t_1, t_2) = C(S_1(t_1), S_2(t_2)),
\]

\[
f(t_1, t_2) = c(S_1(t_1), S_2(t_2)) f_1(t_1) f_2(t_2),
\]

where \(C(u_1, u_2)\) is sometimes called the survival copula.

Let \(c_0(u_1, u_2; \alpha)\), where \((u_1, u_2) \in [0, 1]^2\), be a class of copula density functions parametrized by a finite \(p\)-dimensional coefficient \(\alpha \in A \subset \mathbb{R}^p\). In this section we consider the simplest case where the marginal distributions are known and the hypothesized copula distribution is completely specified with \(\alpha = \alpha_0\). The corresponding simple null hypothesis then takes the form

\[
H_0 : \Pr(c(u_1, u_2) = c_0(u_1, u_2; \alpha_0)) = 1,
\]

against the alternative hypothesis

\[
H_1 : \Pr(c(u_1, u_2) = c_0(u_1, u_2; \alpha_0)) < 1.
\]

We propose a smooth test of copula specification in the presence of censored data. Let \(\psi = (\psi_1, \ldots, \psi_k)'\) be a vector of linearly independent bounded real valued functions defined on \([0, 1]^2\) and \(\lambda = (\lambda_1, \ldots, \lambda_k)' \in \mathbb{R}^k\). Similar to Neyman's smooth test of distribution, our smooth test of copula specification can be motivated by a smooth alternative. Denote \(U_j^0 = S_j(T_j), j = 1, 2\). Consider a family of densities for \((U_1^0, U_2^0)\) of the form,

\[
c_\psi(u_1, u_2; \alpha_0, \lambda) = c_0(u_1, u_2; \alpha_0) \exp(\lambda \psi(u_1, u_2) - \lambda_0)
\]

where \(C(u_1, u_2)\) is sometimes called the survival copula.
with a normalization constant \( \lambda_0 = \log \int c_0(u_1, u_2; \alpha_0) \exp(\lambda \psi(u_1, u_2))du_1du_2 \). This construction has an appealing information theoretic interpretation: it can be derived as the density that minimizes the conditional Kullback-Leibler information criterion between the target density \( c_\psi \) in (5) and the null density subject to moment conditions associated with \( \psi \) (cf. Efron and Tibshirani (1996)).

Under the assumption that \((U_1^0, U_2^0)\) are distributed according to a distribution in family (5), hypothesis (4) is equivalent to that \( \lambda = 0 \). Define

\[
\mu_\psi = E[\psi(U_1^0, U_2^0)] = \int_{[0,1]^2} \psi(u_1, u_2)c_0(u_1, u_2; \alpha_0)du_1du_2,
\]

where the expectation is taken with respect to the hypothesized null copula distribution. For simplicity, we shall omit the subscript \([0,1]^2\) in the double integration whenever there is no ambiguity. One can then construct a score test on \( \lambda = 0 \) based on the discrepancy between \( \mu_\psi \), which is the population mean of \( \psi(U_1^0, U_2^0) \) under the null, and the sample analog of the conditional mean of the \( \psi(U_1^0, U_2^0) \) given the observations \( \{(X_{1t}, X_{2t}, \delta_{1t}, \delta_{2t})\}_{t=1}^n \).

In the presence of right censorship, suppose we have \( n \) independent (but not necessarily identically distributed) observations \( \{(X_{1t}, X_{2t}, \delta_{1t}, \delta_{2t})\}_{t=1}^n \), where \( T_{jt} \) and \( D_{jt} \) are survival and censoring variables, \( X_{jt} = T_{jt} \wedge D_{jt} \) and \( \delta_{jt} = I(T_{jt} \leq D_{jt}), j = 1, 2 \). Let \( U_{jt} = S_j(X_{jt}), j = 1, 2 \). Denote by \( C_\psi(u_1, u_2; \alpha_0) \) the distribution function of the density \( c_\psi(u_1, u_2; \alpha_0) \). The average log-likelihood under the parametric copula family (5) is given by

\[
n^{-1} \sum_{t=1}^n l_\psi(U_{1t}, U_{2t}; \alpha_0),
\]

where

\[
l_\psi(u_{1t}, u_{2t}; \alpha) = \delta_{1t}\delta_{2t} \log c_\psi(u_{1t}, u_{2t}; \alpha) + \delta_{1t}(1 - \delta_{2t}) \log \frac{\partial C_\psi(u_{1t}, u_{2t}; \alpha)}{\partial u_1}
+ (1 - \delta_{1t})\delta_{2t} \log \frac{\partial C_\psi(u_{1t}, u_{2t}; \alpha)}{\partial u_2} + (1 - \delta_{1t})(1 - \delta_{2t}) \log C_\psi(u_{1t}, u_{2t}; \alpha).
\]

Note that we suppress the dependence of \( l_\psi(u_{1t}, u_{2t}; \alpha) \) on \( \delta_{1t}, \delta_{2t} \) for notational simplicity. Such simplifications will also be applied to other likelihood functions. Substituting (5) into
(6) yields the following log-likelihood

\[
\frac{1}{n} \sum_{t=1}^{n} \left\{ \delta_{1t} \delta_{2t} \left[ X' \psi(U_{1t}, U_{2t}) + \log c_0(U_{1t}, U_{2t}; \alpha_0) \right] \\
+ \delta_{1t}(1 - \delta_{2t}) \log \int_{0}^{U_{2t}} \exp(X' \psi(U_{1t}, u_2)) c_0(U_{1t}, u_2; \alpha_0) \, du_2 \\
+ (1 - \delta_{1t}) \delta_{2t} \log \int_{0}^{U_{1t}} \exp(X' \psi(u_1, U_{2t})) c_0(u_1, U_{2t}; \alpha_0) \, du_1 \\
+ (1 - \delta_{1t})(1 - \delta_{2t}) \log \int_{0}^{U_{2t}} \int_{0}^{U_{1t}} \exp(X' \psi(u_1, u_2)) c_0(u_1, u_2; \alpha_0) \, du_1 \, du_2 - \lambda_0 \right\}. \tag{7}
\]

The derivative of (7) with respect to \( \lambda \), evaluated at \( \lambda = 0 \), is given by

\[
\hat{g}_\psi = \frac{1}{n} \sum_{t=1}^{n} g_\psi(U_{1t}, U_{2t}; \alpha_0), \tag{8}
\]

where

\[
g_\psi(u_{1t}, u_{2t}; \alpha) = \delta_{1t} \delta_{2t} \psi(u_{1t}, u_{2t}) + \delta_{1t}(1 - \delta_{2t}) \int_{0}^{u_{2t}} \frac{\psi(u_{1t}, u_2) c_0(u_{1t}, u_2; \alpha) \, du_2}{C_1(u_{1t}, u_{2t}; \alpha)} \\
+ (1 - \delta_{1t}) \delta_{2t} \int_{0}^{u_{1t}} \frac{\psi(u_1, u_{2t}) c_0(u_1, u_{2t}; \alpha) \, du_1}{C_2(u_{1t}, u_{2t}; \alpha)} \\
+ (1 - \delta_{1t})(1 - \delta_{2t}) \int_{0}^{u_{2t}} \int_{0}^{u_{1t}} \frac{\psi(u_1, u_2) c_0(u_1, u_2; \alpha) \, du_1 \, du_2}{C_0(u_{1t}, u_{2t}; \alpha)} \\
- \int \psi(u_1, u_2) c_0(u_1, u_2; \alpha) \, du_1 \, du_2 \tag{9}
\]

with \( C_1(u_{1t}, u_{2t}; \alpha) = \int_{0}^{u_{2t}} c_0(u_{1t}, u_2; \alpha) \, du_2 \) and \( C_2(u_{1t}, u_{2t}; \alpha) = \int_{0}^{u_{1t}} c_0(u_1, u_{2t}; \alpha) \, du_1 \). Again we suppress the dependence of \( g_\psi \) on \( \delta_{1t} \) and \( \delta_{2t} \) when there is no confusion.

According to (9), different types of observations contribute to the score function \( \hat{g}_\psi \) differently:

- the uncensored observations contribute to the score in the usual form
  \[
  \psi(U_{1t}, U_{2t}) = \psi(S_1(T_{1t}), S_2(T_{2t}));
  \]

- if an observation is censored in both margins, its contribution to the score takes the
form
\[
\int_0^{U_{2t}} \int_0^{U_{1t}} \psi(u_1, u_2) c_0(u_1, u_2; \alpha_0) du_1 du_2 \cr C_0(U_{1t}, U_{2t}; \alpha_0) = E[\psi(S_1(T_1), S_2(T_2))|T_1 > X_{1t}, T_2 > X_{2t}],
\]
which is the conditional mean of \( \psi \) given that \( T_1 > X_{1t} \) and \( T_2 > X_{2t} \);

- if an observation is censored, say, in the first but not the second margin, its contribution to the score function is given by
\[
\int_0^{U_{1t}} \psi(u_1, U_{2t}) c_0(u_1, U_{2t}; \alpha_0) du_1 \cr C_2(U_{1t}, U_{2t}; \alpha_0) = E[\psi(S_1(T_1), S_2(T_2))|T_1 > X_{1t}],
\]
which is the conditional mean of \( \psi \) given that \( T_1 > X_{1t} \) and \( T_2 = T_{2t} \).

Similar to \( \hat{\Psi} \) defined in Section 2, \( \hat{g}_{\psi} \) estimates the conditional expectation of the score function for the (unobserved) complete data given the observations \( \{(X_{1t}, X_{2t}, \delta_{1t}, \delta_{2t})\}_{t=1}^n \). Define
\[
I_n = \frac{1}{n} \sum_{t=1}^n \text{var}[g_{\psi}(U_{1t}, U_{2t}; \alpha_0)]. \tag{10}
\]
It follows that under the null hypothesis, \( \hat{g}_{\psi} \xrightarrow{p} 0 \) and \( \sqrt{n}I_n^{-1/2} \hat{g}_{\psi} \xrightarrow{d} N(0, I_k) \) as \( n \to \infty \) under suitable regularity conditions.

Next define \( \hat{I}_n = n^{-1} \sum_{t=1}^n g_{\psi}(U_{1t}, U_{2t}; \alpha_0) g_{\psi}(U_{1t}, U_{2t}; \alpha_0)' \). We construct a smooth test for copula specification as follows:
\[
Q_k = n \hat{g}_{\psi}' \hat{I}_n^{-1} \hat{g}_{\psi}. \tag{11}
\]
Under null hypothesis (4), \( Q_k \) can be shown to converge in distribution to the \( \chi^2 \) distribution with \( k \) degrees of freedom under suitable regularity conditions given in the next section.

4 Smooth tests of copula specification for censored data: composite hypothesis

In practice, the marginal distributions and copula coefficients are usually unknown and have to be estimated. In the presence of unknown parameters, we now face the following composite
hypthesis:

\[ H_0 : \Pr(c(u_1, u_2) = c_0(u_1, u_2; \alpha_0)) = 1 \text{ for some } \alpha_0 \in A, \quad (12) \]

against the alternative hypothesis

\[ H_1 : \Pr(c(u_1, u_2) = c_0(u_1, u_2; \alpha)) < 1 \text{ for all } \alpha \in A. \]

One can estimate the marginal distributions using parametric or nonparametric methods. In general, tests with parametrically estimated marginal distributions are more efficient if the parametric margins are correctly specified. However, under incorrect distributional assumptions, the tests can suffer persistent size distortions even if the sample size goes to infinity. In contrast, tests with non-parametrically estimated marginals are immune from misspecification of the marginal distributions at the expense of convergence rate. We investigate both testing strategies in this study.

4.1 Tests with parametrically estimated marginal distributions

We first consider the case where the marginal distributions are assumed to be known up to a finite number of unknown parameters. Denote by \( \beta_j \in B_j \subset \mathbb{R}^{q_j}, j = 1, 2 \), the coefficients for marginal distributions. Let \( \theta = (\beta'_1, \beta'_2, \alpha)' \), \( \theta \in \Theta \subset \mathbb{R}^{q_1 + q_2 + p} \). Under null hypothesis (12), the average log-likelihood of \( \{(X_{1t}, X_{2t}, \delta_{1t}, \delta_{2t})\}_{t=1}^n \) is given by

\[
L^*(\theta) = n^{-1} \sum_{t=1}^n l^*(X_{1t}, X_{2t}; \theta),
\]

where

\[
l^*(X_{1t}, X_{2t}; \theta) = \delta_{1t} \delta_{2t} \log f(X_{1t}, X_{2t}; \theta) + \delta_{1t} (1 - \delta_{2t}) \log \left[-\frac{\partial S(X_{1t}, X_{2t}; \theta)}{\partial X_{1t}}\right] + (1 - \delta_{1t}) \delta_{2t} \log \left[-\frac{\partial S(X_{1t}, X_{2t}; \theta)}{\partial X_{2t}}\right] + (1 - \delta_{1t})(1 - \delta_{2t}) \log S(X_{1t}, X_{2t}; \theta). \quad (13)
\]

Let \( U_{jt} = S_j(X_{jt}, \beta_j), j = 1, 2 \). (13) can be rewritten as

\[
l^*(X_{1t}, X_{2t}; \theta) = l(U_{1t}, U_{2t}; \alpha) + \delta_{1t} \log f_1(X_{1t}, \beta_1) + \delta_{2t} \log f_2(X_{2t}, \beta_2),
\]

\[\text{See Equation (11.2.14) of Lawless (2003) for the form of likelihood function for fully parametric models.}\]
where \( f_j \) is the marginal distribution of \( T_j \) and

\[
l(u_{1t}, u_{2t}; \alpha) = \delta_1 \delta_2 \log c_0(u_{1t}, u_{2t}; \alpha) + \delta_1(1 - \delta_2) \log \frac{\partial C_0(u_{1t}, u_{2t}; \alpha)}{\partial u_1} \\
+ \delta_2(1 - \delta_1) \log \frac{\partial C_0(u_{1t}, u_{2t}; \alpha)}{\partial u_2} + (1 - \delta_1)(1 - \delta_2) \log C_0(u_{1t}, u_{2t}; \alpha).
\]

(14)

It’s easy to see that \( l(u_{1t}, u_{2t}; \alpha) \) is identical to (6) under null hypothesis (12).\(^4\)

Copula models with unknown parametric marginal distributions can be estimated using two strategies: (i) the two-staged estimation procedure estimates the marginal distributions first and then the copula parameter, taking the estimated marginal distributions as given; (ii) the simultaneous procedure that estimates the marginal and copula parameters in one stage. We employ the second strategy in this study. Let \( L_{\hat{\beta}_1}(\theta), L_{\hat{\beta}_2}(\theta) \) and \( L_{\alpha}(\theta) \) be the derivatives of \( L^*(\theta) \) with respect to \( \beta_1, \beta_2 \) and \( \alpha \). We obtain the maximum likelihood estimate \( \hat{\theta}_n = (\hat{\beta}_1^n, \hat{\beta}_2^n, \hat{\alpha}_n)^\prime \) by solving the system of equations: \( L_{\hat{\beta}_1}(\theta) = 0, L_{\hat{\beta}_2}(\theta) = 0 \) and \( L_{\alpha}(\theta) = 0 \).

Let \( \Psi_P \) be a non-empty subset of basis functions \( \{\psi_{i_1i_2} : 1 \leq i_1, i_2 \leq M\} \). Denote the cardinality of \( \Psi_P \) by \( K_P = |\Psi_P| \). Similar to (5), we consider a smooth alternative distribution given by

\[
c_P(u_1, u_2; \alpha_0, \lambda_P) = c_0(u_1, u_2; \alpha_0) \exp \left( \sum_{i=1}^{K_P} \lambda_{P,i} \psi_{P,i}(u_1, u_2) - \lambda_{P,0} \right),
\]

(15)

where \( \lambda_P = (\lambda_{P,1}, ..., \lambda_{P,K_P})^\prime \) and \( \lambda_{P,0} \) is a normalization constant.

Let \( \hat{U}_{jt} = S_j(X_{jt}; \hat{\beta}_{jn}) \), \( j = 1, 2 \). One can envision that a smooth test on hypothesis (12) can be constructed based on the estimated score

\[
\hat{g}_P(\hat{\theta}_n) = \frac{1}{n} \sum_{t=1}^n g_P(X_{1t}, X_{2t}; \hat{\theta}_n),
\]

(16)

where \( g_P \) is defined analogously to (9) with \( \psi \) replaced by \( \Psi_P \), \( \hat{U}_{jt} \) replaced by \( \hat{U}_{jt} \), \( j = 1, 2 \) and \( \alpha_0 \) by \( \hat{\alpha}_n \).\(^5\) However, test (11) for simple hypotheses is not directly applicable here because of the presence of nuisance parameters in (16): the copula parameter \( \alpha \) and the parameters of the marginal distributions \( \beta_j \), \( j = 1, 2 \).

In order to construct a test based on (16), we need to account for the influence of nuisance parameters.

\(^4\) In the Appendix, we list some assumptions on \( l(U_{1t}, U_{2t}; \alpha) \) in order to allow the score function and its derivatives to blow up at the boundaries. These conditions are needed to accommodate many commonly used copulas. See Chen et al. (2010) for a detailed discussion on this.

\(^5\) Note here \( g_P \) is not only a function of \( \alpha \) but also that of \( \beta_j \)’s, as \( X_{jt} \) enters the function via \( \hat{U}_{jt} \), which depends on \( \hat{\beta}_{jn} \).
parameters. We start with the asymptotic distribution of \( \hat{\theta}_n \), which is originally studied by Shih and Louis (1995). Let \( \theta_0 = (\beta_{10}', \beta_{20}', \alpha_0') \) denote the vector of true coefficients. For \( i = 1, 2, j = 1, 2 \), let

\[
\Sigma_{i,j,n}^* = -\frac{1}{n} \sum_{t=1}^{n} E\{l_{\beta_i,\beta_j}^*(X_{1t}, X_{2t}; \theta_0)\}, \quad \Sigma_{j,3,n}^* = -\frac{1}{n} \sum_{t=1}^{n} E\{l_{\beta_j,\alpha}^*(X_{1t}, X_{2t}; \theta_0)\},
\]

and also define

\[
\Sigma_{33,n}^* = -\frac{1}{n} \sum_{t=1}^{n} E\{l_{\alpha,\alpha}^*(X_{1t}, X_{2t}; \theta_0)\},
\]

where \( l_{\beta_i,\beta_j}^*(X_1, X_2; \theta) = \partial^2 l^*(X_1, X_2; \theta)/\partial \beta_i \partial \beta_j', \ l_{\beta_j,\alpha}^*(X_1, X_2; \theta) = \partial^2 l^*(X_1, X_2; \theta)/\partial \beta_j \partial \alpha' \) and \( l_{\alpha,\alpha}^*(X_1, X_2; \theta) = \partial^2 l^*(X_1, X_2; \theta)/\partial \alpha \partial \alpha' \).

Under null hypothesis (12), the distribution of \( \hat{\theta}_n \) is given by the following theorem.

**Theorem 1.** (a) Under conditions \( \text{C1-C4} \), \( \|\hat{\theta}_n - \theta_0\| = o_p(1) \). (b) If conditions \( \text{A1-A4} \) also hold, then \( \sqrt{n} \Sigma^*_n (\hat{\theta}_n - \theta_0) \to N(0, I_{q_1 + q_2 + p}) \), where \( \Sigma^*_n \) is partitioned into

\[
\Sigma^*_n = \begin{bmatrix}
\Sigma_{11,n}^* & \Sigma_{12,n}^* & \Sigma_{13,n}^* \\
\Sigma_{21,n}^* & \Sigma_{22,n}^* & \Sigma_{23,n}^* \\
\Sigma_{31,n}^* & \Sigma_{32,n}^* & \Sigma_{33,n}^*
\end{bmatrix}.
\]

**Remark 1.** This theorem extends Theorem 1 of Shih and Louis (1995) by allowing for more general censoring mechanisms than the simple random censoring considered in Shih and Louis (1995). Since the censoring variables \( \{D_{1t}, D_{2t}\}_{t=1}^n \) in this paper are not necessarily identically distributed, \( \Sigma^*_n \) is defined based on individual observations instead of sample averages.

The asymptotic variance of \( \hat{\theta}_n \) can be consistently estimated by

\[
\hat{\Sigma}^*_n = \begin{bmatrix}
\hat{\Sigma}_{11,n}^* & \hat{\Sigma}_{12,n}^* & \hat{\Sigma}_{13,n}^* \\
\hat{\Sigma}_{21,n}^* & \hat{\Sigma}_{22,n}^* & \hat{\Sigma}_{23,n}^* \\
\hat{\Sigma}_{31,n}^* & \hat{\Sigma}_{32,n}^* & \hat{\Sigma}_{33,n}^*
\end{bmatrix}.
\]
where for $i,j = 1,2$,

\[
\hat{\Sigma}_{ij,n} = \frac{1}{n} \sum_{t=1}^{n} l_{\beta_i}(X_{1t}, X_{2t}; \hat{\theta}_n)l_{\beta_j}(X_{1t}, X_{2t}; \hat{\theta}_n)',
\]

\[
\hat{\Sigma}_{j3,n} = \frac{1}{n} \sum_{t=1}^{n} l_{\beta_j}(X_{1t}, X_{2t}; \hat{\theta}_n)l_{\alpha}(X_{1t}, X_{2t}; \hat{\theta}_n)',
\]

\[
\hat{\Sigma}_{33,n} = \frac{1}{n} \sum_{t=1}^{n} l_{\alpha}(X_{1t}, X_{2t}; \hat{\theta}_n)l_{\alpha}(X_{1t}, X_{2t}; \hat{\theta}_n)'.
\]

Further define $G_P = (G_{P,\beta_1}', G_{P,\beta_2}', G_{P,\alpha})'$ with

\[
G_{P,\beta_j} = \frac{1}{n} \sum_{t=1}^{n} E\left[ g_{P,\beta_j}(X_{1t}, X_{2t}; \theta_0) \right], \quad G_{P,\alpha} = \frac{1}{n} \sum_{t=1}^{n} E\left[ g_{P,\alpha}(X_{1t}, X_{2t}; \theta_0) \right],
\]

where for $j = 1,2$, $g_{P,\beta_j}(X_{1t}, X_{2t}; \theta)$ and $g_{P,\alpha}(X_{1t}, X_{2t}; \theta)$ are the partial derivatives of $g_P(X_{1t}, X_{2t}; \theta)$ with respect to $\beta_j$ and $\alpha$. We obtain the following asymptotic distribution of $\hat{g}_P(\hat{\theta}_n)$.

**Theorem 2.** Suppose all conditions of Theorem 1 are satisfied and conditions $T1$-$T4$ given in the Appendix hold. Under null hypothesis (12), $\hat{g}_P(\hat{\theta}_n) \xrightarrow{p} 0$ and $\sqrt{n}\Omega^{-1/2}_P \hat{g}_P(\hat{\theta}_n) \xrightarrow{d} N(0, I_{K_P})$, where

\[
\Omega_P = \frac{1}{n} \sum_{t=1}^{n} \text{var}\left[ g_P(X_{1t}, X_{2t}; \theta_0) + G_{P,\beta_j}^* \hat{\Sigma}_n^{-1} l_{\theta}(X_{1t}, X_{2t}; \theta_0) \right]
\]

with $l_{\theta}'(X_{1t}, X_{2t}; \theta) = (l_{\beta_1}'(X_{1t}, X_{2t}; \theta)', l_{\beta_2}'(X_{1t}, X_{2t}; \theta)', l_{\alpha}'(X_{1t}, X_{2t}; \theta)')'$.  

**Remark 2.** The asymptotic variance $\Omega_P$ reflects the influence of nuisance parameters in $\hat{g}_P(\hat{\theta}_n)$. When both the marginal distributions and copula parameter are known, $\Omega_P$ is reduced to $n^{-1} \sum_{t=1}^{n} \text{var}\{g_P(X_{1t}, X_{2t}; \theta_0)\}$, which is exactly the variance (10) in the absence of nuisance parameters.

Next we present a consistent estimator for $\Omega_P$. Let

\[
\hat{\varphi}_t = g_P(X_{1t}, X_{2t}; \hat{\theta}_n) + \hat{G}_{P,\beta}^* \hat{\Sigma}_n^{-1} l_{\theta}(X_{1t}, X_{2t}; \hat{\theta}_n),
\]

where $\hat{G}_P = (\hat{G}_{P,\beta_1}', \hat{G}_{P,\beta_2}', \hat{G}_{P,\alpha})'$ and for $j = 1,2$, $\hat{G}_{P,\beta_j} = n^{-1} \sum_{t=1}^{n} g_{P,\beta_j}(X_{1t}, X_{2t}; \hat{\theta}_n)$ and
\[ \hat{g}_{P,\alpha} = n^{-1} \sum_{t=1}^{n} g_{P,\alpha}(X_{1t}, X_{2t}; \hat{\theta}_n). \]

It follows that \( \Omega_P \) can be estimated consistently by

\[ \hat{\Omega}_P = \frac{1}{n} \sum_{t=1}^{n} (\hat{\varphi}_t - \bar{\varphi})(\hat{\varphi}_t - \bar{\varphi})', \tag{19} \]

where \( \bar{\varphi} = n^{-1} \sum_{t=1}^{n} \hat{\varphi}_t \). We can now construct a parametric smooth test of copula specification as follows.

**Theorem 3.** Suppose the conditions of Theorem 1 and 2 are satisfied. The parametric smooth test of copula specification for censored data is given by

\[ \hat{Q}_P = n \hat{g}_P(\hat{\theta}_n)'\hat{\Omega}_P^{-1} \hat{g}_P(\hat{\theta}_n). \tag{20} \]

Under null hypothesis (12), \( \hat{Q}_P \xrightarrow{d} \chi^2_{K_P} \) as \( n \to \infty \).

A critical component of data driven smooth tests is the selection of suitable set of functions \( \Psi_P \) from a candidate set \( \Psi \). In the spirit of Kallenberg and Ledwina (1999), we rearrange the candidate set to \( \Psi_p \) such that \( \Psi_{p,1} = \psi_{11} \) and the remaining elements, \( \{\psi_{i_1,i_2} : 1 \leq i_1, i_2 \leq M, (i_1, i_2) \neq (1,1)\} \), are ordered according to their corresponding entries in the vector \( \hat{\Omega}_P^{-1/2} |\sqrt{n} \hat{g}_P| \). The normalization of the moments is necessary as these moments are generally correlated and differ in variance.\(^6\)

Denote by \( |\Psi_p| \) the cardinality of \( \Psi_p \). Let \( \Psi_{p,(k)} = \{\psi_{p,1}, \ldots, \psi_{p,k}\}, k = 1, \ldots, |\Psi_p| \) and the corresponding \( g_{p,(k)}(X_{1t}, X_{2t}, \theta_0) \) and \( \Omega_{p,(k)} \), as given in (16) and Theorem 2, are similarly defined; their sample analogs are denoted by \( \hat{\Psi}_{p,(k)}, \hat{g}_{p,(k)}(\hat{\theta}_n) \) and \( \hat{\Omega}_{p,(k)} \), respectively. We then proceed to define \( \hat{Q}_{p,(k)} \) analogously to (20). For each \( k \), let \( \hat{u}_{p,(k)} = \hat{\Omega}_{p,(k)}^{-1/2} \hat{g}_{p,(k)}(\hat{\theta}_n) \). Following Inglot and Ledwina (2006), we use the following criterion to select a suitable \( \Psi_p \), whose cardinality is denoted by \( K_P \):

\[ K_P = \min \{k : \hat{Q}_{p,(k)} - J(k, n, \zeta) \geq \hat{Q}_{p,(r)} - J(r, n, \zeta), 1 \leq k, r \leq |\Psi_p|\}, \tag{21} \]

where \( J(r, n, \zeta) \) is a complexity penalty defined as:

\[ J(r, n, \zeta) = \begin{cases} r \log n & \text{if } \max_{1 \leq i \leq |\Psi_p|} |\sqrt{n} \hat{u}_{p,i}| \leq \sqrt{\zeta \log n}, \\ 2r & \text{if } \max_{1 \leq i \leq |\Psi_p|} |\sqrt{n} \hat{u}_{p,i}| > \sqrt{\zeta \log n}. \end{cases} \tag{22} \]

where \( \hat{u}_{p,i} \) is the \( i \)th element of \( \hat{u}_{p,(k)} \) and \( \zeta = 2.4 \). Note that the penalty is ‘adaptive’ such

\(^6\)In contrast, Kallenberg and Ledwina (1999) use ranks to construct a test of independence. The moments based on Legendre polynomials are free of nuisance parameters and orthonormal under the null hypothesis.
that either the AIC or BIC is adopted in a data driven manner, depending on the empirical evidence pertinent to the magnitude of deviation from the null copula distribution.

Next we present the asymptotic properties of the proposed test \( \hat{Q}_P \) based on a set of basis functions \( \Psi_P \) selected according to the procedure described above.

**Theorem 4.** Let \( K_P \) be selected according to (21). Suppose all conditions in Theorem 3 hold and \( M \to \infty \) as \( n \to \infty \) with \( M = o(\log n / \log \log n) \). (a) Suppose \((U_{1t}, U_{2t})\) follows the distribution \( C_0(u_1, u_2; \alpha_0) \). Then \( \Pr(K_P = 1) \xrightarrow{P} 1 \) and \( \hat{Q}_P \xrightarrow{d} \chi^2_1 \). (b) Let \( \mathcal{P} \) be an alternative. Suppose that under \( \mathcal{P} \), \( \hat{\theta}_n \) converges to some pseudo-parameter \( \theta \) such that \( \mathbb{E}_{\mathcal{P}}[g_{p,S}(X_{1t}, X_{2t}, \theta)] \neq 0 \) for some \( S \) in \( 1, \ldots, k \), where \( g_{p,S}(X_{1t}, X_{2t}, \theta) \) is the \( S \)th element of \( g_{p,k}(X_{1t}, X_{2t}, \theta) \) for \( k = 1, \ldots, |\Psi_p| \). Then \( \hat{Q}_P \to \infty \) as \( n \to \infty \).

Since the proposed test is constructed in a data driven fashion, the number of functions selected is random even under the null hypothesis. Consequently, the \( \chi^2_1 \) distribution does not provide an adequate approximation to the distribution of the test statistic for moderate sample sizes. Nevertheless since it is asymptotically distribution free, the asymptotic distribution of the test under the null can be easily approximated via simulations. Below we present a simple simulation procedure for this purpose.

- Step 1: Generate an \( n \)-by-2 i.i.d. random sample \( \{(U_{1t}, U_{2t})\}_{t=1}^n \) from the Uniform\([0,1]^2\).
- Step 2: Select a set of basis functions \( \Psi_P \) according to (21); Calculate the test statistic \( \hat{Q}^*_P \).
- Step 3: Repeat steps 1 and 2 \( L \) times; denote the results by \( \{\hat{Q}^*_P,l\}_{l=1}^L \).
- Step 4: Use the \( q \)th percentile of \( \{\hat{Q}^*_P,l\}_{l=1}^L \) to approximate the \( q \)th percentile critical value for the distribution of \( \hat{Q}_P \).

### 4.2 Tests with nonparametric marginal distributions

In this section, we eschew parametric assumptions on the marginal distributions. Let \( U_{jt} = S_j(X_{jt}), \ j = 1, 2 \). We estimate \( S_j(\cdot) \) by the Kaplan-Meier estimator

\[
\tilde{S}_j(x) = \prod_{X_{jt} \leq x} \left(1 - \frac{1}{n - t + 1}\right)^{\delta_{j(t)}},
\]

where for \( j = 1, 2 \), \( X_{j(1)} \leq X_{j(2)} \leq \ldots \leq X_{j(n)} \) are the order statistics of \( \{X_{jt}\}_{t=1}^n \), and \( \{\delta_{j(t)}\}_{t=1}^n \) are similarly defined. Lai and Ying (1991) established the consistency of \( \tilde{S}_j(\cdot) \) under the assumption of independent censoring.
Define \((\tilde{U}_{1t}, \tilde{U}_{2t}) = (\tilde{S}_1(X_{1t}), \tilde{S}_2(X_{2t}))\). Given \(\{(\tilde{U}_{1t}, \tilde{U}_{2t})\}_{t=1}^n\), we proceed to estimate the copula parameters via the maximum likelihood estimation

\[
\hat{\alpha}_n = \arg\max_{\alpha} \frac{1}{n} \sum_{t=1}^n l(\tilde{U}_{1t}, \tilde{U}_{2t}; \alpha),
\]

where \(l(u_{1t}, u_{2t}; \alpha)\) is defined in (14).

Let \(\Psi_S\) be a non-empty subset of basis functions \(\{\psi_{i_1i_2} : 1 \leq i_1, i_2 \leq M\}\). Denote the cardinality of \(\Psi_S\) by \(K_S = |\Psi_S|\). Similarly to (5), we consider a smooth alternative distribution given by

\[
c_S(u_1, u_2; \alpha, \lambda_S) = c_0(u_1, u_2; \alpha_0) \exp \left( \sum_{i_1=1}^{K_S} \lambda_{S,i} \Psi_{S,i}(u_1, u_2) - \lambda_{S,0} \right),
\]

where \(\lambda_S = (\lambda_{S,1}, \ldots, \lambda_{S,K_S})'\) and \(\lambda_{S,0}\) is a normalization constant. Similar to the test with parametric margins, the test with nonparametric margins is based on the score function

\[
\tilde{g}_S(\hat{\alpha}_n) = \frac{1}{n} \sum_{t=1}^n g_s(\tilde{U}_{1t}, \tilde{U}_{2t}; \hat{\alpha}_n),
\]

which is analogous to (9) with \(\Psi\) replaced by \(\Psi_S\), \(U_{jt}\) replaced by \(\tilde{U}_{jt}\), \(j = 1, 2\) and \(\alpha_0\) by \(\hat{\alpha}_n\).

We need to account for the influence of the nuisance parameters, including the finite dimensional copula parameter \(\alpha\) and the infinite-dimensional marginal survival function \(S_j\), \(j = 1, 2\). We start with the asymptotic distribution of \(\hat{\alpha}_n\), which is studied by Shih and Louis (1995) and Chen et al. (2010). We denote

\[
W_{jt} \equiv W_j(X_{jt}, \delta_{jt}; \alpha_0) = E\{I_{\alpha}(U_{1s}, U_{2s}; \alpha_0) I_j(X_{jt}, \delta_{jt})(X_{js})|X_{jt}, \delta_{jt}\},
\]

\[
I_j(X_{jt}, \delta_{jt})(X_{js}) = -S_j(X_{js}) \left[ \int_{-\infty}^{X_{js}} \frac{dN_{jt}(u)}{P_{nj}(u)} - \int_{-\infty}^{X_{jt}} \frac{I\{X_{jt} \geq u\} dH_j(u)}{P_{nj}(u)} \right],
\]

where \(H_j(x) \equiv -\log(S_j(x))\) is the cumulative hazard function of \(X_j\), \(N_{jt}(x) = \delta_{jt} I(X_{jt} \leq x)\), \(dN_{jt}(x) = N_{jt}(x) - N_{jt}(x-)\), and \(P_{nj}(x) = n^{-1} \sum_{k=1}^n P(X_{jrk} \geq x)\). Also define

\[
B_n = -\frac{1}{n} \sum_{t=1}^n E\{I_{\alpha}(U_{1t}, U_{2t}; \alpha_0)\}, \quad \Sigma_n = \frac{1}{n} \sum_{t=1}^n \var\{I_{\alpha}(U_{1t}, U_{2t}; \alpha_0)\} + \sum_{j=1}^2 W_{jt}.
\]

The asymptotic distribution of \(\hat{\alpha}_n\) is given by the following theorem.

**Theorem 5.** (a) Under Conditions **C1-C2** and **C5-C7** given in the Appendix, \(\|\hat{\alpha}_n - \alpha_0\| = o_p(1)\). (b) Furthermore, if conditions **A3** and **A5-A7** hold, then \(\sqrt{n}B_n\Sigma_n^{-1/2}(\hat{\alpha}_n - \alpha_0) \xrightarrow{d} \)
Chen et al. (2010) established the equivalence between these two expressions.

\[ \sum \] where \( \tilde{\alpha} \)

As indicated in Chen et al. (2010), an alternative expression for \( \tilde{\alpha} \) is uncorrelated with \( W_{jt}, j = 1, 2 \). Therefore, the asymptotic variance of \( \sqrt{n}(\tilde{\alpha}_n - \alpha_0) \) can be simplified to

\[ B_n^{-1} + \frac{1}{n} B_n^{-1} \sum_{t=1}^{n} \text{var}\{W_{1t} + W_{2t}\} B_n^{-1}. \]

If the marginal distributions are known, the asymptotic variance of \( \tilde{\alpha}_n \) is reduced to \( B_n^{-1} \).

Next define

\[ B_n = \frac{1}{n} \sum_{t=1}^{n} l_{\alpha}(\tilde{U}_{1t}, \tilde{U}_{2t}; \tilde{\alpha}_n) l_{\alpha}(\tilde{U}_{1t}, \tilde{U}_{2t}; \tilde{\alpha}_n)', \]

\[ \tilde{W}_{jt} = \frac{1}{n} \sum_{s=1,s \neq t}^{n} l_{\alpha_j}(\tilde{U}_{1s}, \tilde{U}_{2s}; \tilde{\alpha}_n) \tilde{I}_j(X_{jt}, \delta_{jt}) (X_{js}), \]

where for \( j = 1, 2 \),

\[ \tilde{I}_j(X_{jt}, \delta_{jt})(X_{js}) = -\tilde{S}_j(X_{js}) \left[ \frac{I(X_{jt} \leq X_{js}, \delta_{jt})}{n^{-1} \sum_{k=1}^{n} I(X_{jk} \geq X_{jt})} - \frac{1}{n} \sum_{t=1}^{n} I(X_{js} \geq X_{jt}) I(X_{jt} \geq X_{jt}) \delta_{jt} \right]. \]  

As indicated in Chen et al. (2010), an alternative expression for \( \tilde{I}_j(X_{jt}, \delta_{jt})(X_{js}) \) is

\[ \tilde{I}_j(X_{jt}, \delta_{jt})(X_{js}) = -\tilde{S}_j(X_{js}) \left[ \frac{I\{X_{jt} \leq X_{js}, \delta_{jt} = 1\}}{\tilde{P}_{nj}(X_{jt})} - \sum_{X_{jt} \leq X_{js}} I(X_{jt} \geq X_{jt}) \Delta \tilde{H}_j(X_{jt}) \right], \]

where \( \tilde{P}_{nj}(x) = n^{-1} \sum_{k=1}^{n} I(X_{jk} \geq x) \) and \( \Delta \tilde{H}_j(x) = \frac{I(Y_{j}(x) > 0)}{Y_{j}(x)} d\tilde{N}_j(x) \), in which \( \tilde{Y}_{j}(x) = \sum_{k=1}^{n} I\{X_{jk} \geq x\} \) and \( \tilde{N}_j(x) = \sum_{k=1}^{n} N_{jk}(x) \). \( \Delta \tilde{H}_j(x) \) is the so-called Nelson’s estimator. Chen et al. (2010) established the equivalence between these two expressions.

The asymptotic variance of \( \tilde{\alpha}_n \) can be consistently estimated by

\[ B_n^{-1} + \tilde{B}_n^{-1} \left\{ \frac{1}{n} \sum_{t=1}^{n} (\sum_{j=1}^{2} \tilde{W}_{jt}) (\sum_{j=1}^{2} \tilde{W}_{jt})' \right\} \tilde{B}_n^{-1}, \]

where \( \tilde{B}_n^{-1} \) is the generalized inverse of \( \tilde{B}_n \).
Further define
\[ G_{S,\alpha} = \frac{1}{n} \sum_{t=1}^{n} E[g_{S,\alpha}(U_{1t}, U_{2t}; \alpha_0)], \]
\[ Z_{jt} = Z_j(X_{jt}, \delta_{jt}; \alpha_0) = E[g_{S,j}(U_{1s}, U_{2s}; \alpha_0) I_j(X_{jt}, \delta_{jt})(X_{js}), j = 1, 2, \quad (27) \]
where \( g_{S,\alpha}(u_1, u_2; \alpha) \) and \( g_{S,j}(u_1, u_2; \alpha) \) are the partial derivatives of \( g_S(u_1, u_2; \alpha) \) with respect to \( \alpha \) and \( u_j \). We obtain the following asymptotic distribution of \( \tilde{g}_S(\tilde{\alpha}_n) \).

**Theorem 6.** Suppose that all conditions of Theorem 5 are satisfied and conditions T5-T7 given in the Appendix hold. Under null hypothesis (12), \( \tilde{g}_S(\tilde{\alpha}_n) \overset{p}{\rightarrow} 0 \) and \( \sqrt{n}V_S^{-1/2} \tilde{g}_S(\tilde{\alpha}_n) \overset{d}{\rightarrow} N(0, I_K) \), where
\[ V_S = \frac{1}{n} \sum_{t=1}^{n} \text{var} \left[ g_S(U_{1t}, U_{2t}; \alpha_0) + \sum_{j=1}^{2} Z_{jt} + \mathcal{G}_{S,\alpha} B_n^{-1} \left\{ I_\alpha(U_{1t}, U_{2t}; \alpha_0) + \sum_{j=1}^{2} W_{jt} \right\} \right]. \quad (28) \]

**Remark 3.** The various simplifications to the variance for the parametric test discussed in Remark 2 apply to its semiparametric counterpart (28) as well.

Next we present a consistent estimator for \( V_S \). Define
\[ \tilde{G}_{S,\alpha} = \frac{1}{n} \sum_{t=1}^{n} g_{S,\alpha}(\tilde{U}_{1t}, \tilde{U}_{2t}; \tilde{\alpha}_n), \quad (29) \]
\[ \tilde{Z}_{jt} = \frac{1}{n} \sum_{s=1, s \neq t}^{n} g_{S,j}(\tilde{U}_{1s}, \tilde{U}_{2s}; \tilde{\alpha}_n) \tilde{I}_j(X_{jt}, \delta_{jt})(X_{js}), j = 1, 2. \quad (30) \]
Let
\[ \tilde{\varphi}_t = g_S(\tilde{U}_{1t}, \tilde{U}_{2t}; \tilde{\alpha}_n) + \sum_{j=1}^{2} \tilde{Z}_{jt} + \tilde{\mathcal{G}}_{S,\alpha} \tilde{B}_n^{-1} \left\{ I_\alpha(\tilde{U}_{1t}, \tilde{U}_{2t}; \tilde{\alpha}_n) + \sum_{j=1}^{2} \tilde{W}_{jt} \right\}. \]
It follows that \( V_S \) can be estimated consistently by
\[ \tilde{V}_S = \frac{1}{n} \sum_{t=1}^{n} (\tilde{\varphi}_t - \bar{\varphi})(\tilde{\varphi}_t - \bar{\varphi})', \]
where \( \bar{\varphi} = n^{-1} \sum_{t=1}^{n} \tilde{\varphi}_t \).

We can now construct a semiparametric smooth test of copula specification as follows.
Theorem 7. Suppose all conditions of Theorem 5 and 6 are satisfied. The semiparametric smooth test of copula specification for censored data is given by

\[ \hat{Q}_S = n\hat{g}_S(\hat{\alpha}_n)\hat{V}_S^{-1}\hat{g}_S(\hat{\alpha}_n). \]  

Under null hypothesis (12), \( \hat{Q}_S \xrightarrow{d} \chi^2_{K_S} \).

Similarly to the tests with the parametric margins, we use data driven method to select a suitable \( \Psi_S \) to capture possible deviations from the null distribution. We start with a candidate set \( \Psi = \{\psi_{i_1,i_2} : 1 \leq i_1, i_2 \leq M\} \) and rearrange \( \Psi \) to \( \Psi_s \) such that \( \Psi_{s,1} = \psi_{11} \) and the remaining elements, \( \{\psi_{i_1,i_2} : 1 \leq i_1, i_2 \leq M, (i_1, i_2) \neq (1,1)\} \), are arranged in the descending order according to their corresponding entries in the vector \( \hat{V}_\Psi^{-1/2} |\sqrt{n}\hat{g}_\Psi| \).

Denote by \( |\Psi_s| \) the cardinality of \( \Psi_s \). Let \( \Psi_s(k) = \{\Psi_{s,1}, ..., \Psi_{s,k}\} \), \( k = 1, ..., |\Psi_s| \) and the corresponding \( g_{s,(k)}(U_{1t}, U_{2t}, \alpha_0) \) and \( V_{s,(k)} \), as given in (24) and Theorem 6 are similarly defined; their sample analogs are denoted by \( \tilde{\Psi}_{s,(k)}, \tilde{g}_{s,(k)}(\hat{\alpha}_n) \) and \( \tilde{V}_{s,(k)} \), respectively. We then proceed to define \( \tilde{Q}_{s,(k)} \) analogously to (31). For each \( k \), let \( \tilde{u}_{s,(k)} = \tilde{V}_{s,(k)}^{-1/2}\tilde{g}_{s,(k)}(\hat{\alpha}_n) \).

Following Inglot and Ledwina (2006), we use the following criterion to select a suitable \( \Psi_s \), whose cardinality is denoted by \( K_S \):

\[ K_S = \min\{k : \tilde{Q}_{s,(k)} - J(k, n, \zeta) \geq \tilde{Q}_{s,(r)} - J(r, n, \zeta), 1 \leq k, r \leq |\Psi_s|\}, \]  

where the complexity penalty \( J(r, n, \zeta) \) is the same as (22) with \( \tilde{u}_{s,i} \), the \( i \)th element of \( \tilde{u}_{s,(k)} \), taking the place of \( \hat{u}_{p,i} \).

The asymptotic properties of the data driven test \( \tilde{Q}_S \) and its consistency can be established similarly as those of the parametric test \( \hat{Q}_P \).

Theorem 8. Let \( K_S \) be selected according to (32). Suppose all conditions in Theorem 7 hold and \( M \to \infty \) as \( n \to \infty \) with \( M = o(\log n/\log \log n) \). (a) Suppose \( (U_{1t}, U_{2t}) \) follows the distribution \( C_0(u_1, u_2, \alpha_0) \). Then \( \Pr(K_S = 1) \xrightarrow{p} 1 \) and \( \hat{Q}_S \xrightarrow{d} \chi^2_1 \). (b) Let \( \mathcal{P} \) be an alternative. Suppose that under \( \mathcal{P} \), \( \hat{\alpha}_n \) converges to some pesudo-parameter \( \alpha \) such that \( E_{\mathcal{P}}[g_{s,K}(U_{1t}, U_{2t}, \alpha)] \neq 0 \) for some \( K \) in 1, ..., \( k \), where \( g_{s,K}(U_{1t}, U_{2t}, \alpha) \) is the \( K \)th element of \( g_{s,(k)}(U_{1t}, U_{2t}, \alpha) \), \( k \) in 1, ..., |\( \Psi_s | \). Then \( \hat{Q}_S \to \infty \) as \( n \to \infty \).

The same simulation procedure, as described in the previous section for the parametric tests, can be used to calculate the critical values of the tests with nonparametrically estimated marginal distributions.
5 Simulations

We conduct a series of Monte Carlo simulations to assess the finite-sample performance of the proposed tests. In align with the existing literature, we consider four commonly used parametric Archimedean copulas as the null copula distributions: the Clayton, Frank, log- and Gumbel copula. They share a common structure $C(v_1, v_2) = \phi^{-1}\{\phi(v_1) + \phi(v_2)\}$, $0 < v_1, v_2 < 1$, where the function $\phi(\cdot)$ is often referred to as the generator of an Archimedean copula. For the log-copula, $\phi(v) = (1 - \log(v)/\alpha \gamma)^{\alpha + 1} - 1$. We assume $\alpha \gamma = 1$ as in Wang and Wells (2000) and Chen and Fan (2007). We refer readers to Wang and Wells (2000) for a detailed description of these copulas. Regarding the alternative copulas, we include the same four Archimedean copulas, plus the Gaussian copula and $t$ copula with 4 degrees of freedom due to their popularity in practice. For each copula, we consider two levels of dependence by setting its parameter such that the corresponding Kendall’s tau $\tau = 0.3$ and 0.6.

In our simulations, we generate the censoring variable $D_1$ and $D_2$ from the standard exponential distribution. The marginal distributions of $T_1$ and $T_2$ are generated from the exponential distributions, whose parameters are set such that the censoring rates $P(T_1 < D_1) = 0.2$ and $P(T_2 < D_2) = 0.4$. For each experiment, we generate 1,000 pseudo-random samples of size $n = 250$.

We select the basis functions from the normalized Legendre polynomials, which are orthonormal with respect to the Lebesgue measure on $[0, 1]$. Our $\psi$ functions consist of elements of the tensor products of these basis functions. Let $\psi_j$ denote the $j$th normalized Legendre polynomial and $\psi_{ij}(u_1, u_2) = \psi_i(u_1)\psi_j(u_2)$. We consider the candidate set \{\psi_{11}, \psi_{12}, \psi_{22}, \psi_{13}, \psi_{23}, \psi_{33}\}, which is the upper triangle of a 3 by 3 matrix of basis functions. We apply the information criterion prescribed in the previous sections to select a suitable set of basis functions, based on which the test statistic is calculated. We have experimented with larger candidate sets and found little difference in terms of size and power across experiment designs and sample sizes. The critical value is computed following the prescribed simulation procedure. We use the MATLAB functions integral and integral2 for one- and two-dimensional integrations in the calculation of the test statistics. When the sample size $n = 250$, the critical value at the 5% significance level is equal to 7.8533.

For each experiment, we estimate the marginal distributions parametrically and non-parametrically and calculate their corresponding test statistics of copula specification. The empirical sizes and powers of the tests with parametric and nonparametric margins are reported in Table 1 under the headers $\hat{Q}_P$ and $\hat{Q}_S$ respectively. The bold entries reflect the empirical size. The proposed tests show the desirable size, centering about 5%, under var-
ious null copula distributions. The size appears to be rather stable across different levels of dependence. The regular font entries show the power of the tests. Non-trivial powers are observed for all combinations of null and alternative distributions under either level of dependence. The performance of the parametric and nonparametric tests are largely comparable. Overall, the nonparametric tests seem to have higher powers; at the same time, their sizes are also slightly higher.

In most cases, the powers are well above the 5% nominal level. When \( \tau = 0.3 \), the powers for distinguishing among the Frank, log- and Gaussian copulas are relatively low; the power for distinguishing between the Clayton copula and log-copula is also conservative. However, these are not unexpected given the well-documented evidence on the close resemblance shared by these copulas. Several existing studies report similar difficulties in testing among these copulas when the dependence is low; see, e.g., Wang and Wells (2000), Chen and Fan (2007) and Kojadinovic and Yan (2011). It is noted that regarding the same set of experiments, the powers of the proposed tests increase considerably when \( \tau = 0.6 \).

Table 1: Empirical sizes and powers of the proposed tests (%)

<table>
<thead>
<tr>
<th>True Copula</th>
<th>( \tau )</th>
<th>( \hat{Q}_P )</th>
<th>( \hat{Q}_S )</th>
<th>( \hat{Q}_P )</th>
<th>( \hat{Q}_S )</th>
<th>( \hat{Q}_P )</th>
<th>( \hat{Q}_S )</th>
<th>( \hat{Q}_P )</th>
<th>( \hat{Q}_S )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>0.3</td>
<td>6.1</td>
<td>7.1</td>
<td>27.2</td>
<td>21.7</td>
<td>11.6</td>
<td>18.4</td>
<td>78.4</td>
<td>64.7</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>6.4</td>
<td>6.9</td>
<td>25.7</td>
<td>70.6</td>
<td>14.9</td>
<td>29.2</td>
<td>99.1</td>
<td>100</td>
</tr>
<tr>
<td>Frank</td>
<td>0.3</td>
<td>27.3</td>
<td>32.6</td>
<td>4.1</td>
<td>6.7</td>
<td>6.7</td>
<td>10.4</td>
<td>47.5</td>
<td>38.8</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>66.9</td>
<td>69.9</td>
<td>4.3</td>
<td>7.2</td>
<td>10.3</td>
<td>23.7</td>
<td>75.6</td>
<td>84.5</td>
</tr>
<tr>
<td>log-copula</td>
<td>0.3</td>
<td>10.8</td>
<td>10.1</td>
<td>10.4</td>
<td>7.4</td>
<td>5.2</td>
<td>4.6</td>
<td>56.2</td>
<td>47.5</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>15.1</td>
<td>21.7</td>
<td>10.7</td>
<td>23.6</td>
<td>5.9</td>
<td>7.0</td>
<td>94.0</td>
<td>95.8</td>
</tr>
<tr>
<td>Gumbel</td>
<td>0.3</td>
<td>74.9</td>
<td>84.0</td>
<td>43.8</td>
<td>43.3</td>
<td>54.1</td>
<td>66.4</td>
<td>6.3</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>100</td>
<td>100</td>
<td>55.2</td>
<td>88.4</td>
<td>84.0</td>
<td>99.1</td>
<td>6.6</td>
<td>8.6</td>
</tr>
<tr>
<td>Gaussian</td>
<td>0.3</td>
<td>27.9</td>
<td>36.6</td>
<td>8.3</td>
<td>7.9</td>
<td>12.2</td>
<td>17.1</td>
<td>25.6</td>
<td>17.4</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>89.2</td>
<td>95.6</td>
<td>24.0</td>
<td>38.7</td>
<td>38.0</td>
<td>68.2</td>
<td>26.1</td>
<td>20.5</td>
</tr>
<tr>
<td>t_4</td>
<td>0.3</td>
<td>50.7</td>
<td>64.5</td>
<td>41.4</td>
<td>29.7</td>
<td>41.9</td>
<td>50.0</td>
<td>14.1</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>93.3</td>
<td>98.7</td>
<td>52.3</td>
<td>66.1</td>
<td>62.0</td>
<td>80.7</td>
<td>15.5</td>
<td>16.3</td>
</tr>
</tbody>
</table>

Note: The bold entries reflect the empirical sizes. The nominal size is 5%. The sample size is 250. The number of replications is 1,000.
We conclude this section with some illustration of the impact of misspecified marginal distributions on the parametric test. In particular, we calculate the empirical sizes when the marginal distributions are mistakenly estimated under the assumption of a Weibull distribution with a shape parameter 0.9 and a free scale parameter. Note that when the shape parameter equals 1, the Weibull distribution reduces to the exponential distribution – the null distribution. In the case of Clayton null copula, the empirical sizes for $\tau = (0.3, 0.6)$ are (38.8%, 43.4%). The corresponding sizes for the Frank, log- and Gumbel copulas are respectively (32.7%, 40.7%), (37.2%, 41.2%) and (38.7%, 32.4%). The tests under misspecified marginal distributions are apparently over-sized, implying the sensitivity of parametric tests with respect to distributional assumption of the marginal densities.

6 Empirical examples

In this section, we apply the proposed tests to two real world examples. To save space, we report only results of the nonparametric tests, which are suggested to share the same desirable performance with the correctly specified parametric tests and at the same time are immune from potential misspecification risk.

6.1 Couples’ wage and salary earnings

Our first example concerns the dependence structure of married couples’ total wage and salary earnings. We randomly selected the earnings of 1,500 couples from the March 2001 Supplement to the Current Population Survey (CPS). The data were restricted to dual-earner couples in which both husband and wife were aged 25 to 55 at the time of the survey, eligible to labor force, and reporting non-negligible individual annual earnings ($1,000 or higher). In this dataset, there are 142 couples with either husband’s or wife’s earnings right censored at some top-coded values, and 8 couples with both spouses’ earnings censored.\footnote{For details on topcoding rules used by the CPS, see the documentation of March 2001 Current Population Survey (Bureau of the Census) at \url{http://www.census.gov/prod/techdoc/cps/cpsmar01.pdf}.}

We estimate the marginal distributions by the Kaplan-Meier estimator. The scatterplot of the couples’ earnings on the logarithm scale is displayed in Figure 1a. Only uncensored observations are plotted. Figure 1a shows that more data points are clustered below the diagonal line, suggesting that a husband is more likely to earn more than his wife. It also shows apparent positive right-tail dependence between couples’ wage and salary earnings: men with high earnings tend to match with women with high earnings, which is consistent
with the theory of “positive assortative mating” (cf. Lam (1988)). At the same time, the left-tail dependence is also not negligible.

We consider the null hypothesis of Frank copula, which is well suited for empirical application due to its desirable properties, see Frees and Valdez (1998), Nelsen (1986) and Genest (1987). This copula is employed by Frees et al. (1996) to study the dependence of couples’ lifetimes, which are subject to heavy censoring. Our preliminary analysis also prefers the Frank copula over the Clayton, log- and Gumbel copulas.8

The semi-parametric estimate of the Frank copula coefficient is \( \hat{\alpha}_n = 1.8980 \). The test statistic with the data-driven dimension \( K = 3 \) is equal to 29.7036. The simulated critical value for the sample size \( n = 1,500 \) at the 5% level is 5.6612. Thus the proposed test decisively rejected the null hypothesis of Frank copula at the 5% level.

A useful feature of the proposed tests is that they provide useful guidance in the construction of alternative copulas. If the null hypothesis is rejected, we can augment the null copula distribution with extra features suggested by the selected moment functions of a data-driven test. In the present case, given the hypothesized Frank copula and the selected moment functions \( \{ \psi_{11}, \psi_{12}, \psi_{22} \} \), the implied alternative copula takes the form

\[
c_a(u_1, u_2; \hat{\alpha}_n, \hat{\lambda}) = c_F(u_1, u_2; \hat{\alpha}_n) \exp(\tilde{\lambda}_1 \psi_{11}(u_1, u_2) + \tilde{\lambda}_2 \psi_{12}(u_1, u_2) + \tilde{\lambda}_3 \psi_{22}(u_1, u_2) - \tilde{\lambda}_0),
\]

where \( c_F \) is the Frank copula density function, and \( \tilde{\lambda}_0 \) is a normalization constant. It is not a trivial task to calculate the maximum likelihood estimates \( \hat{\alpha}_n, \tilde{\lambda}_1, ..., \tilde{\lambda}_k \) due to the complicate functional form of (6) and the presence of censoring. We adopt a two-stage estimation approach. In the first step, we estimate the parameter \( \alpha \) by restricting \( \tilde{\lambda}_k = 0 \), \( k = 1, 2, 3 \). In the second step, we fix the estimated \( \hat{\alpha}_n \) and use the EM algorithm to evaluate the log-likelihood function. At the maximization step, we use a Newton-type algorithm (see Wu, 2010) to update \( \tilde{\lambda}'s.\)

Figure 1b and 1c report the contour plots of the Frank copula density \( c_F(\alpha = 1.8980) \) and the alternative copula density function \( c_a \). The Frank copula allows only symmetric dependence. In contrast, the alternative copula density captures a salient feature of the data: the density at the upper-right corner is clearly higher than that at the lower-left tail corner, reflecting the asymmetric tail behaviors revealed in Figure 1a.

8For the Clayton, Frank, log- and Gumbel copulas, their estimated log-likelihoods are \((-597, -353, -390, -358)\). Their corresponding AIC and BIC are respectively \((1196, 708, 782, 718)\) and \((1201, 713, 787, 724)\). We therefore choose to focus on the Frank copula in this investigation.
Our second example concerns the dependence structure of the indemnity payment (LOSS) and the allocated loss adjustment expense (ALAE) of insurance companies. The data set consists of 1,500 observations, of which 34 observations of the losses are right censored because the amount of claim cannot exceed the policy limit. The ALAE data are not censored. The data were collected by the US Insurance Services Office and have been analyzed by, among others, Klugman and Parsa (1999), Denuit et al. (2006) and Chen et al. (2010).\footnote{We thank Professors Frees and Valdez for kindly providing the loss-ALAE data.}

All evidence in the existing studies suggested that these data can be best fitted by the Gumbel copula. We therefore conduct the specification test under the Gumbel copula null hypotheses. The semiparametric estimator yields $\tilde{\alpha}_n = 1.4440$. The smooth test statistic, with a data-driven dimension of $K = 1$, is calculated to be $\tilde{Q}_S = 0.0151$, considerably smaller than the critical value of 5.6612. In accordance with the previous studies, the proposed smooth test fails to reject the hypothesis of Gumbel copula.

### 7 Concluding remarks

In this paper, we propose a family of flexible specification tests of copulas under general censorship. Our tests can be characterized as score tests on moment conditions of empirical copula distributions under the null hypothesis. We use data-driven methods to select a set of suitable moment functions to construct consistent tests. We investigate two possible strategies of estimating the marginal distributions – parametric and nonparametric – and establish...
their theoretical properties. Monte Carlo simulations show outstanding finite sample performance of the proposed tests. In particular, the tests with nonparametric margins are shown to be compatible with the parametric tests with correctly specified marginal distributions and at the same time are immune from the risk of misspecification of marginal distributions. We use two real world examples to demonstrate the usefulness of the proposed methods.

References


Appendix

To simplify notation, we let \( \sum_t = \sum_{t=1}^n \), \( \sum_j = \sum_{j=1}^2 \). We first list some conditions needed to establish the large sample properties of the proposed tests. For \( j = 1, 2 \), denote \( U_{jt} = S_j(X_{jt}, \beta_{j0}) \) or \( S_j(X_{jt}) \), for the conditions of the tests with the parametric margins and with the nonparametric margins, respectively.

Conditions **C1 - C7** are sufficient to ensure the convergence of the maximum likelihood estimators \( \hat{\theta}_n \) to \( \theta_0 \) in Theorem 1 and \( \hat{\alpha}_n \) to \( \alpha_0 \) in Theorem 5.

**C1** (i) The sequence of survival variables, \( \{(T_{1t}, T_{2t})\}_{t=1}^n \), is an i.i.d. sample from an unknown survival function \( S(t_1, t_2) \) with continuous marginal survival functions \( S_j(\cdot) \), \( j = 1, 2 \).

(ii) The sequence of censoring variables \( \{(D_{1t}, D_{2t})\}_{t=1}^n \), is an independent sample with joint survival functions \( \{G_t(x_1, x_2)\}_{t=1}^n = \{P(D_{1t} > x_1, D_{2t} > x_2)\}_{t=1}^n \) and marginal survival functions \( \{G_{jt}(\cdot)\}_{t=1}^n, j = 1, 2 \).

(iii) The censoring variables \( (D_{1t}, D_{2t}) \) are independent of survival variables \( (T_{1t}, T_{2t}) \) and there is no mass concentration at 0 in the sense that \( \limsup_{n \to \infty} n^{-1} \sum_t (1 - G_{jt}(\eta)) \to 0 \) as \( \eta \to 0 \).

**C2** The true (unknown) copula function \( C(u_1, u_2) \) has continuous partial derivatives.

**C3** Let \( \Theta \) be a compact subset of \( \mathbb{R}^{q_1+q_2+p} \). For every \( \epsilon > 0 \),

\[
\liminf_{\theta \in \Theta, ||\theta - \theta_0|| \geq \epsilon} \frac{1}{n} \sum_t \{E[l^*(X_{1t}, X_{2t}; \theta_0)] - E[l^*(X_{1t}, X_{2t}; \theta)] \} > 0.
\]

**C4** (i) \( l^*(x_1, x_2; \theta) \) is a continuous function of \( \theta \in \Theta \).

(ii) \( n^{-1} \sum_t E\{\sup_{\theta \in \Theta, ||\theta - \theta_0|| = o(1)} |l^*(X_{1t}, X_{2t}; \theta)|\} < \infty \).
C5 Let $\mathcal{A}$ be a compact subset of $\mathbb{R}^p$. For every $\epsilon > 0$,
\[
\liminf_{\alpha \in \mathcal{A} : ||\alpha - \alpha_0|| \geq \epsilon} \frac{1}{n} \sum_{t} \{ E[l(U_{1t}, U_{2t}; \alpha)] - E[l(U_{1t}, U_{2t}; \alpha_0)] \} > 0.
\]

C6 (i) For any $(u_1, u_2) \in (0, 1)^2$, $l(u_1, u_2; \alpha)$ is a continuous function of $\alpha \in \mathcal{A}$.
(ii) Let $L_t = \sup_{\alpha \in \mathcal{A}} |l(U_{1t}, U_{2t}; \alpha)|$ and $L_{ta} = \sup_{\alpha \in \mathcal{A}} |l_{\alpha}(U_{1t}, U_{2t}; \alpha)|$. Then
\[
\lim_{M \to \infty} \limsup_{n \to \infty} \frac{1}{n} \sum_{t} E\{L_tI(L_t \geq M) + L_{ta}I(L_{ta} \geq M)\} = 0.
\]
(iii) For any $\eta > 0$, $\epsilon > 0$, there is $M > 0$ such that $|l(u_1, u_2; \alpha)| \leq M|l(\bar{u}_1, \bar{u}_2; \alpha)|$ for all $\alpha \in \mathcal{A}$ and all $u_j \in [\eta, 1)$ such that $1 - u_j \geq \epsilon(1 - \bar{u}_j)$, $j = 1, 2$.

C7 If $\{T_{jt}\}_{t=1}^n$ are subject to non-trivial censoring (i.e., $D_{jt} \neq \infty$), then $\tilde{S}_j$ is truncated at the tail in the sense that for some $\tau_j$, $\tilde{S}_j(x_j) = \tilde{S}_j(\tau_j)$ for all $x_j \geq \tau_j$ and $\liminf n^{-1} \sum_{t=1}^n G_{jt}(\tau_j)S_j(\tau_j) > 0$.

In contrast to the censoring mechanism in Shih and Louis (1995), Condition C1(ii) allows the censoring variables $\{(D_{1t}, D_{2t})\}_{t=1}^n$ to be non-identically distributed. In addition, no assumptions are made on the joint survival function $G_t(x_1, x_2)$ of the censoring variables $(D_{1t}, D_{2t})$. Therefore, general censoring mechanism is allowed in this framework, for example, both variables are censored, one censored and the other uncensored, or one random censoring and the other fixed censoring. Conditions C3 and C5 are identifiability conditions for the tests with the parametric and nonparametric margins, respectively. Condition C7 is imposed to handle the tail instability of the Kaplan-Meier estimator, specifically for non-identically distributed censoring times.

Conditions A1 - A7 are sufficient to ensure the asymptotic normality of $\hat{\theta}_n$ in Theorem 1 and $\tilde{\alpha}_n$ in Theorem 5.

A1 $\Sigma_n^*$ defined in Theorem 1 is finite and positive definite.

A2 Functions $l_{i\alpha}(x_1, x_2; \theta)$, $l_{i\beta, \alpha}(x_1, x_2; \theta)$ and $l_{i\beta, \beta}(x_1, x_2; \theta)$, $i, j = 1, 2$, are well-defined and continuous functions of $\theta$, where $\theta \in \Theta$.

A3 (i) $|l_i(u_1, u_2; \alpha_0)| \leq q \prod_{j=1}^2 \{u_j(1 - u_j)\}^{-a_j}$ for some $q > 0$ and $a_j \geq 0$ such that $\limsup n^{-1} \sum_t E[\prod_{j=1}^2 \{U_{jt}(1 - U_{jt})\}^{-2a_j}] < \infty$;
(ii) $|l_{ij}(u_1, u_2; \alpha_0)| \leq \text{constant} \times \{u_j(1-u_j)\}^{-b_j} \{u_k(1-u_k)\}^{-a_k}$ for some $b_j, a_k$ and $j \neq k$ such that $\lim sup n^{-1} \sum_t E[\{U_{jt}(1-U_{jt})\}^{\xi_j-b_j} \{U_{kt}(1-U_{kt})\}^{-a_k}] < \infty$ for some $\xi_j \in (0, 1/2)$.

A4 (i) $n^{-1} \sum_t E\{\sup_{\theta \in \Theta; \|\theta-\theta_0\|=o(1)} l_{i,j}^*(X_{1t}, X_{2t}; \theta)\} < \infty$, \(i, j = 1, 2;\)

(ii) $n^{-1} \sum_t E\{\sup_{\theta \in \Theta; \|\theta-\theta_0\|=o(1)} l_{i,j}^*(X_{1t}, X_{2t}; \theta)\} < \infty$, \(j = 1, 2;\)

(iii) $n^{-1} \sum_t E\{\sup_{\theta \in \Theta; \|\theta-\theta_0\|=o(1)} l_{i,j}^*(X_{1t}, X_{2t}; \theta)\} < \infty$.

A5 (i) $B_n = -n^{-1} \sum_{t=1}^n E\{l_{\alpha\alpha}(U_{1t}, U_{2t}; \alpha_0)\}$ is positive definite;

(ii) $\Sigma_n = n^{-1} \sum_{t=1}^n \text{var}\{l_{\alpha}(U_{1t}, U_{2t}; \alpha_0) + \sum_j W_{jt}\}$ is finite and positive definite, where $W_{jt}$ is defined in (25);

(iii) $\{l_{\alpha}(U_{1t}, U_{2t}; \alpha_0) + \sum_j W_{jt}\}_{t=1}^n$ satisfies Lindeberg condition.

A6 $l_{\alpha\alpha}(u_1, u_2; \alpha)$ and $l_{\alpha j}(u_1, u_2; \alpha)$, \(j = 1, 2\) are well-defined and continuous in $(u_1, u_2; \alpha) \in (0, 1)^2 \times \mathcal{A}$.

A7 (i) Let $L_{\alpha\alpha} = \sup_{\alpha \in \mathcal{A}} |l_{\alpha}(U_{1t}, U_{2t}; \alpha)|$ and $L_{\alpha\alpha} = \sup_{\alpha \in \mathcal{A}} |l_{\alpha}(U_{1t}, U_{2t}; \alpha)|$.

Then, $\lim_{M \to \infty} \lim_{n \to \infty} n^{-1} \sum_t E\{L_{\alpha\alpha}I(L_{\alpha\alpha} \geq M) + L_{\alpha\alpha}I(L_{\alpha\alpha} \geq M)\} = 0$;

(ii) For any $\eta > 0$ and any $\epsilon > 0$, there is $M > 0$, such that $|l_{\alpha}(u_1, u_2; \alpha)| + |l_{\alpha\alpha}(u_1, u_2; \alpha)| \leq M\{|l_{\alpha}(\bar{u}_1, \bar{u}_2; \alpha)| + |l_{\alpha\alpha}(\bar{u}_1, \bar{u}_2; \alpha)|\}$ for all $\alpha \in \mathcal{A}$ and all $u_j \in [\eta, 1)$ such that $1-u_j \geq \epsilon(1-\bar{u}_j)$, $j = 1, 2$.

Conditions A3 and A7 allow the score function and its partial derivatives with respect to $u_j$, $j = 1, 2$ to blow up at the boundaries but require them be dominated by a weighted function which satisfies certain moment conditions. These conditions characterize many popular copula functions such as the Gaussian, $t$ and Clayton copulas (see Chen et al. (2010) for details).

Proof of Theorem 1. Since $\hat{\theta}_n$ is the maximum likelihood estimate of $\theta_0$, it’s straightforward to show that $\hat{\theta}_n \overset{p}{\to} \theta_0$ under the listed regularity conditions.

Now we prove the asymptotic normality of $\hat{\theta}_n$. Applying Taylor’s expansion to the score
function $L^*_{\beta_1}(\hat{\theta}_n), L^*_{\beta_2}(\hat{\theta}_n)$ and $L^*_{\alpha}(\hat{\theta}_n)$ yields,

$$
L^*_{\beta_1}(\hat{\theta}_n) = 0 = L^*_{\beta_1}(\theta_0) - L^*_{\beta_1,\beta_1}(\theta_0)(\hat{\beta}_1 - \beta_1) - L^*_{\beta_1,\beta_2}(\theta_0)(\hat{\beta}_2 - \beta_2) - L^*_{\beta_1,\alpha}(\theta_0)(\hat{\alpha}_n - \alpha_0) + o_p(n^{-1/2})
$$

(A.1)

$$
L^*_{\beta_2}(\hat{\theta}_n) = 0 = L^*_{\beta_2}(\theta_0) - L^*_{\beta_2,\beta_1}(\theta_0)(\hat{\beta}_1 - \beta_1) - L^*_{\beta_2,\beta_2}(\theta_0)(\hat{\beta}_2 - \beta_2) - L^*_{\beta_2,\alpha}(\theta_0)(\hat{\alpha}_n - \alpha_0) + o_p(n^{-1/2})
$$

(A.2)

$$
L^*_{\alpha}(\hat{\theta}_n) = 0 = L^*_{\alpha}(\theta_0) - L^*_{\alpha,\beta_1}(\theta_0)(\hat{\beta}_1 - \beta_1) - L^*_{\alpha,\beta_2}(\theta_0)(\hat{\beta}_2 - \beta_2) - L^*_{\alpha,\alpha}(\theta_0)(\hat{\alpha}_n - \alpha_0) + o_p(n^{-1/2})
$$

(A.3)

where for $i, j = 1, 2$, $L^*_{\beta_i,\beta_j}(\theta_0) = -\partial L^*(\theta)/\partial \beta_i \partial \beta_j|_{\theta = \theta_0}$, $L^*_{\alpha,\beta_j}(\theta_0) = -\partial L^*(\theta)/\partial \alpha \partial \beta_j|_{\theta = \theta_0}$ and $L^*_{\alpha,\alpha}(\theta_0) = -\partial L^*(\theta)/\partial \alpha \partial \alpha|_{\theta = \theta_0}$.

Since $||\hat{\theta}_n - \theta_0|| = o_p(1)$, we can obtain that,

$$
\begin{bmatrix}
L^*_{\beta_1,\beta_1}(\theta_0) & L^*_{\beta_1,\beta_2}(\theta_0) & L^*_{\beta_1,\alpha}(\theta_0) \\
L^*_{\beta_2,\beta_1}(\theta_0) & L^*_{\beta_2,\beta_2}(\theta_0) & L^*_{\beta_2,\alpha}(\theta_0) \\
L^*_{\alpha,\beta_1}(\theta_0) & L^*_{\alpha,\beta_2}(\theta_0) & L^*_{\alpha,\alpha}(\theta_0)
\end{bmatrix}
\xrightarrow{p} \Sigma^*_n
$$

(A.4)

where $\Sigma^*_n$ is defined in Theorem 1. Rewriting (A.1), (A.2) and (A.3) into a matrix form and using (A.4), we can obtain,

$$
\begin{bmatrix}
L^*_{\beta_1}(\theta_0) \\
L^*_{\beta_2}(\theta_0) \\
L^*_{\alpha}(\theta_0)
\end{bmatrix} = \Sigma^*_n(\hat{\theta}_n - \theta_0) + o_p(n^{-1/2}),
$$

By central limit theorem, $(L^*_{\beta_1}(\theta_0), L^*_{\beta_2}(\theta_0), L^*_{\alpha}(\theta_0))'$ converges to multivariate normal with mean zero and variance covariance matrix $\Sigma^*_n$. Thus, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges to multivariate normal with mean zero and variance covariance matrix $\Sigma^*_n^{-1}$.

Consistency and asymptotic normality of the semiparametric copula parameter, $\hat{\alpha}_n$, as given in Theorem 5, can be proved similarly. The main difficulty involved with the tests with nonparametric margins lies in the appropriate control of the tail behaviors of Kaplan-Meier estimates.

**Proof of Theorem 5.** Under Conditions **C1-C2** and **C5-C7**, the consistence of $\hat{\alpha}_n$ is readily obtained according to Proposition 3.1 of Chen et al. (2010). Next under Conditions **A3** and **A5-A7**, the asymptotic normality of $\hat{\alpha}_n$ is given by Proposition 3.2 of Chen et al. (2010).
Conditions T1 - T7 are sufficient to ensure the asymptotic normality of the test statistics $\hat{Q}_P$ defined in (20) and $\hat{Q}_S$ in (31).

T1 (i) $g_P(x_1, x_2; \theta)$ is a continuous function of $\theta$;
(ii) $n^{-1} \sum_t E[\sup_{\theta \in \Theta : \|\theta - \theta_0\| = o(1)} |g_P(X_{1t}, X_{2t}; \theta)|] < \infty$.

T2 (i) $\Omega_P = n^{-1} \sum_t \text{var}[g_P(X_{1t}, X_{2t}; \theta_0) + \mathcal{G}'_P \Sigma_n^{-1} l'_\theta(X_{1t}, X_{2t}; \theta_0)]$ is finite and positive definite, where $\mathcal{G}_P$ and $\Sigma_n$ are defined in (17) and Theorem 2, respectively; (ii) $\Omega_{P(k)}$ defined in (21) has all its eigenvalues bounded below and above by some finite positive constants.

T3 (i) $g_{P,\beta_j}(x_1, x_2; \theta)$ is a continuous function of $\theta$ in a neighborhood of $\theta_0$;
(ii) $n^{-1} \sum_t E[\sup_{\theta \in \Theta : \|\theta - \theta_0\| = o(1)} |g_{P,\beta_j}(X_{1t}, X_{2t}; \theta)|] < \infty$.

T4 (i) $g_{P,\alpha}(x_1, x_2; \theta)$ is a continuous function of $\theta$ in a neighborhood of $\theta_0$;
(ii) $n^{-1} \sum_t E[\sup_{\theta \in \Theta : \|\theta - \theta_0\| = o(1)} |g_{P,\alpha}(X_{1t}, X_{2t}; \theta)|] < \infty$.

T5 (i) For any $(u_1, u_2) \in (0, 1)^2$, $g_S(u_1, u_2; \alpha)$ is a continuous function of $\alpha$, $g_S(u_1, u_2; \alpha)$ has continuous partial derivatives with respect to $u_1, u_2$;
(ii) $n^{-1} \sum_t E[\sup_{\alpha \in \mathcal{A} : \|\alpha - \alpha_0\| = o(1)} |g_S(U_{1t}, U_{2t}; \alpha)|] < \infty$.

T6 (i) For any $(u_1, u_2) \in (0, 1)^2$, $g_{S,\alpha}(u_1, u_2; \alpha)$ is a continuous function of $\alpha$ in a neighborhood of $\alpha_0$;
(ii) $n^{-1} \sum_t E[\sup_{\alpha \in \mathcal{A} : \|\alpha - \alpha_0\| = o(1)} |g_{S,\alpha}(U_{1t}, U_{2t}; \alpha)|] < \infty$.

T7 (i) $V_S = n^{-1} \sum_t \text{var}\{g_S(U_{1t}, U_{2t}; \alpha_0) + \sum_j Z_{jt} + \mathcal{G}'_S \Sigma_n^{-1} |l_\alpha(U_{1t}, U_{2t}; \alpha_0) + \sum_j W_{jt}|\}$ is finite and positive definite, where $W_{jt}$ and $Z_{jt}$ are defined in (25) and (27), respectively; (ii) $V_{S,(k)}$ defined in (32) has all its eigenvalues bounded below and above by some finite positive constants.

Proof of Theorem 2. Expanding $\hat{g}_P(\hat{\theta}_n)$ defined in (16) in a Taylor series around $\theta_0$ yields

$$
\hat{g}_P(\hat{\theta}_n) = \frac{1}{n} \sum_t \left\{ g_P(X_{1t}, X_{2t}; \theta_0) + g'_{P,\alpha}(X_{1t}, X_{2t}; \tilde{\theta}_n)(\hat{\alpha}_n - \alpha_0) \\
+ g'_{P,\beta_1}(X_{1t}, X_{2t}; \tilde{\theta}_n)(\hat{\beta}_{1n} - \beta_{10}) + g'_{P,\beta_2}(X_{1t}, X_{2t}; \tilde{\theta}_n)(\hat{\beta}_{2n} - \beta_{20}) \right\} \quad (A.5)
$$

where $\tilde{\theta}_n$ lies between $\theta_0$ and $\hat{\theta}_n$. 

32
Under Conditions T3 and T4, we have

\[
\sup_{\theta \in \Theta : \|\theta - \theta_0\| = o(1)} \left\| \frac{1}{n} \sum_t g_{P,\beta_j}(X_{1t}, X_{2t}; \theta) - \mathcal{G}_{P,\beta_j} \right\| = o(1),
\]

\[
\sup_{\theta \in \Theta : \|\theta - \theta_0\| = o(1)} \left\| \frac{1}{n} \sum_t g_{P,\alpha}(X_{1t}, X_{2t}; \theta) - \mathcal{G}_{P,\alpha} \right\| = o(1),
\]

where \(\mathcal{G}_{P,\alpha}\) and \(\mathcal{G}_{P,\beta_j}\) are defined in (17). Recall that \(\mathcal{G}_P = (\mathcal{G}'_{P,\beta_1}, \mathcal{G}'_{P,\beta_2}, \mathcal{G}'_{P,\alpha})'\), it follows that (A.5) can be written as

\[
\hat{g}_P(\hat{\theta}_n) = n^{-1} \sum_t g_P(X_{1t}, X_{2t}; \theta_0) + \mathcal{G}'_P(\hat{\theta}_n - \theta_0) + o_p(1).
\] (A.6)

It’s easy to see that \(E[\hat{g}_P(\hat{\theta}_n)] = 0\).

Next Theorem 1 indicates that \(\hat{\theta}_n\) can be expressed as an asymptotically linear estimator such that

\[
\hat{\theta}_n - \theta_0 = \Sigma^* - \frac{1}{n} \sum_{t=1}^n l_0^*(X_{1t}, X_{2t}; \theta_0) + o_p(1)
\] (A.7)

where \(\Sigma^*_n\) and \(l_0^*(X_{1t}, X_{2t}; \theta)\) are defined in Theorem 1 and Theorem 2, respectively.

Plugging (A.7) into (A.6) yields

\[
\hat{g}_P(\hat{\theta}_n) = \frac{1}{n} \sum_t [g_P(X_{1t}, X_{2t}; \theta_0) + \mathcal{G}'_P \Sigma^*_n l_0^*(X_{1t}, X_{2t}; \theta_0)] + o_p(1).
\]

It follows that \(\text{var}(\sqrt{n}\hat{g}_P(\hat{\theta}_n)) = \Omega_P\), where \(\Omega_P\) is defined in Theorem 1.

**Proof of Theorem 3.** By Theorem 1, we have \(\hat{\theta}_n \overset{p}{\rightarrow} \theta_0\). Using the boundedness of \(\psi_i, 1 \leq i \leq M\), it follows that \(\hat{\Sigma}^*_n \overset{p}{\rightarrow} \Sigma^*_n, \hat{\mathcal{G}}_P \overset{p}{\rightarrow} \mathcal{G}_P\) and \(\hat{\Omega}_P \overset{p}{\rightarrow} \Omega_P\). The result of this theorem then follows readily from the asymptotic normality of \(\hat{g}_P(\hat{\theta}_n)\) given in Theorem 2.

**Proof of Theorem 4.** Define the simplified BIC rule as follows

\[
SK_P = \min \{k : \hat{Q}_{P,(k)} - k \log n \geq \hat{Q}_{P,(s)} - s \log n, 1 \leq k, s \leq |\Psi_p|\},
\] (A.8)

In order to prove Theorem 4(a), we need to prove that, under the null hypothesis,

\[
\lim_{n \to \infty} \Pr(K_P = SK_P) = 1,
\] (A.9)

33
and that
\[
\lim_{n \to \infty} \Pr(SK_P = 1) = 1, \quad (A.10)
\]
We start by proving (A.9). Recalling that \( \hat{u}_{p, (k)} = \hat{\Omega}_{p, (k)}^{-1/2} \hat{g}_{p, (k)}(\hat{\theta}_n) \), which is a \( k \times 1 \) vector. For \( x \in \mathbb{R}^k \), denote \( \|x\|_\infty = \max_{1 \leq i \leq k} |x_i| \). Define the event
\[
A_n(\zeta) = \left\{ \sqrt{n}\|\hat{u}_{p, (k)}\|_\infty > \sqrt{\zeta \log n} \right\},
\]
By (22), in order to prove (A.9), it suffices to prove \( P(A_n(\zeta)) \to 0 \).
Define
\[
\hat{u}_{p, (k)} = \Omega_{p, (k)}^{-1/2} \sum_t g_{p, (k)}(X_{1t}, X_{2t}; \theta_0).
\]
We can rewrite \( \hat{u}_{p, (k)} \) as,
\[
\sqrt{n}\hat{u}_{p, (k)} = \sqrt{n}\hat{\Omega}_{p, (k)}^{-1/2} \Omega_{p, (k)}^{1/2} \hat{g}_{p, (k)} + R_{n(k)},
\]
where
\[
R_{n(k)} = \sqrt{n}\hat{\Omega}_{p, (k)}^{-1/2} \left\{ \hat{g}_{p, (k)}(\hat{\theta}_n) - \frac{1}{n} \sum_t g_{p, (k)}(X_{1t}, X_{2t}; \theta_0) \right\}. \quad (A.11)
\]
By (A.6) and \( \|\hat{\theta}_n - \theta_0\| = O_p(n^{-1/2}) \), we have
\[
\hat{g}_{p, (k)}(\hat{\theta}_n) = \frac{1}{n} \sum_t g_{p, (k)}(X_{1t}, X_{2t}; \theta_0) + O_p(n^{-1/2}).
\]
By (A.11), we can obtain \( \|R_{n(k)}\|_\infty = O_p(1) \). Therefore, for any \( \epsilon > 0 \), we have \( P(\|R_{n(k)}\|_\infty > 2^{-1}\sqrt{\zeta \log n}) \leq \epsilon/2 \) for all large \( n \).
Also define the event
\[
A_M = \left\{ \|\hat{\Omega}_{p, (k)}^{1/2} \Omega_{p, (k)}^{1/2}\|_2 > M \right\},
\]
where \( M \) is a positive constant to be defined later.
By the consistency of \( \hat{\theta}_n \) and the continuous mapping theorem, we have,
\[
\|\hat{\Omega}_{p, (k)} - \Omega_{p, (k)}\| = o_p(1)
\]
Under Condition \( \text{T2} \) (ii), it’s straightforward to show that there exists a positive constant \( M \) such that \( P(A_M) < \epsilon/2 \) for all \( \epsilon > 0 \).
Notice that

\[ P(A_n(\zeta)) \leq P(A_n(\zeta) \cap A_M^c) + P(A_M) \]

\[ \leq P \left( \sqrt{n} \| \hat{Q}_{p,(k)}^{1/2} \hat{u}_{p,(k)} \|_\infty > 2^{-1} \sqrt{\zeta \log n}, A_M^c \right) \]

\[ + P \left( \| R_n(k) \|_\infty > 2^{-1} \sqrt{\zeta \log n} \right) + \epsilon/2 \]

\[ \leq P \left( \sqrt{n} \| \hat{Q}_{p,(k)}^{1/2} \hat{u}_{p,(k)} \|_2 > 2^{-1} \sqrt{\zeta \log n}, A_M^c \right) + \epsilon, \]

\[ \leq P \left( \sqrt{n} \| \hat{u}_{p,(k)} \|_2 > (2M)^{-1} \sqrt{\zeta \log n} \right) + \epsilon, \]

where \( A_M^c \) denotes the complementary set of \( A_M \).

Now since \( \sqrt{n} \| \hat{u}_{p,(k)} \|_2 = O_p(1) \), it follows that

\[ P \left( \sqrt{n} \| \hat{u}_{p,(k)} \|_2 > (2M)^{-1} \sqrt{\zeta \log n} \right) \to 0, \]

as \( n \to \infty \). Hence, since \( \epsilon > 0 \) was arbitrary we conclude that (A.9) holds.

Next, we prove (A.10) also holds. In view of (A.8), we have

\[ \text{Pr}(SK_P = 1) = 1 - \sum_{k=2}^{[q_p]} \text{Pr}(SK_P = k). \]

Because \( SK_P = k \) implies that dimension \( k \) “beats” dimension 1 and \( \hat{Q}_{p,(1)} \geq 0 \), we obtain by definition of \( SK_P \),

\[ \text{Pr}(SK_P = k) \leq \text{Pr}(\hat{Q}_{p,(k)} - k \log n \geq \hat{Q}_{p,(1)} - \log n) = \text{Pr}(\hat{Q}_{p,(k)} \geq (k - 1) \log n). \]

We have

\[ \text{Pr}(\hat{Q}_{p,(k)} \geq (k - 1) \log n) \leq Q_{1n} + Q_{2n}, \]

where \( Q_{1n} = \text{Pr}(Q_{p,(k)} \geq (k - 1) \log n/2) \) and \( Q_{2n} = \text{Pr}(\| \sqrt{n} \hat{u}_{p,(k)} - \sqrt{n} u_{p,(k)} \|^2 \geq (k - 1) \log n/2) \).

Since under the null \( Q_{p,(k)} \) converges to a non-degenerate (\( \chi_k^2 \) distributed) variable for any \( k \), it’s easy to prove that \( Q_{1n} \to 0 \) as \( n \to \infty \). Similarly, because \( \| \sqrt{n} \hat{u}_{p,(k)} - \sqrt{n} u_{p,(k)} \| = O_p(1), Q_{2n} \to 0 \) as \( n \to \infty \).

Further we have

\[ \text{Pr}(\hat{Q}_P \leq x) = \text{Pr}(\hat{Q}_{p,(1)} \leq x) - \text{Pr}(\hat{Q}_{p,(1)} \leq x, SK_P \geq 2) + \text{Pr}(\hat{Q}_P \leq x, SK_P \geq 2). \]

Because \( \hat{Q}_{p,(1)} \overset{d}{\to} \chi_1^2 \) under \( H_0 \) and \( \text{Pr}(SK_P \geq 2) = 0 \), it follows immediately that \( \hat{Q}_P \overset{d}{\to} \chi_1^2. \)
We now prove Theorem 4(b). Define the simplified AIC rule as follows

\[ AK_P = \min \{ k : \hat{Q}_{p,(k)} - 2k \geq \hat{Q}_{p,(s)} - 2s, 1 \leq k, s \leq |\Psi_p| \}. \tag{A.12} \]

In order to prove Theorem 4(b), we need to prove that under the alternative distribution \( P \),

\[ \lim_{n \to \infty} \Pr(K_P = AK_P) = 1, \tag{A.13} \]

and that

\[ \Pr(AK_P \geq S) \to 1, \tag{A.14} \]

We define the event

\[ D_n(\zeta) = \left\{ \sqrt{n} \| u_{p,(k)} \|_\infty < \sqrt{\zeta \log n} \right\}, \]

Under the alternative distribution \( \mathcal{P} \), there exists \( S \leq |\Psi_p| \) such that \( \mathbb{E}_P[g_{p,S}(X_{1t}, X_{2t}, \theta)] \neq 0 \), therefore \( u_{p,S} \neq 0 \). Then since \( S \leq |\Psi_p| \),

\[ P(D_n(\zeta)) \leq P\left( \sqrt{n} |\hat{u}_{p,S}| < \sqrt{\zeta \log n} \right) \to 0, \]

since \( \hat{u}_{p,S} \to u_{p,S} \neq 0 \) by the continuous mapping theorem. Hence, (A.13) holds.

Now, according to the continuous mapping theorem, we have \( \|\hat{u}_{p,(k)}\|^2 \to 0 \) and \( \|\hat{u}_{p,(S)}\|^2 \to |u_{p,S}|^2 > 0 \). For \( k = 1, \ldots, S - 1 \), we can obtain,

\[ P(AK_P = k) \leq P(\hat{Q}_{p,(k)} - 2k \geq \hat{Q}_{p,(S)} - 2S) \]
\[ \leq P(n\|\hat{u}_{p,(k)}\|^2 \geq 2(k - S) + n\|\hat{u}_{p,(S)}\|^2) \to 0 \]

Therefore, (A.14) also holds. Therefore, for each \( M > 0 \),

\[ \Pr(\hat{Q}_P \leq M) = \Pr(\hat{Q}_P \leq M, K_P \geq S) + o(1) \leq P(n\|\hat{u}_{p,(S)}\|^2 \leq M) + o(1) = o(1), \]

where the last equality follows the continuous mapping theorem. Then, \( \hat{Q}_P \to \infty \) as \( n \to \infty \). \( \Box \)

**Proof of Theorem 6.** In the proofs that follow we make use of Lemma 1 of Chen et al. (2010). We present the results below for the reader’s convenience.

**Lemma (Chen et al., 2010)** Suppose that Conditions C1 and C7 are satisfied. Then: (i) the marginal Kaplan-Meier estimators are uniformly strongly consistent: \( \sup_{x \leq \tau_j} |\hat{S}_j(x) - \hat{S}_j(x)| \to 0 \) almost surely as \( \tau_j \to \infty \).
\( S_j(x) \to 0 \) a.s. for \( j = 1, 2 \); (ii) they can be expressed as martingale integrals:
\[
\tilde{S}_j(x) - S_j(x) = -S_j(x)\int_{-\infty}^{x} \frac{\tilde{S}_j(u-)}{S_j(u)} \sum_t dM_j(t) \sum_t I(X_{jt} \geq u)
\]
where \( M_j(t) = N_j(t) - \int_{-\infty}^{x} J_j(t) dH_j(u) \), in which \( N_j(t) = \delta_j t I(X_{jt} \leq x) \), \( J_j(t) = I(X_{jt} \geq x) \), and \( H_j(x) = -\log S_j(x) \).

By Theorem 5, we have
\[
\|\tilde{\alpha}_n - \alpha_0\| = o_p(1).
\]
Applying Taylor’s expansion to \( \tilde{g}(\tilde{\alpha}_n) \) with respect to \( \tilde{\alpha}_n \) at \( \alpha_0 \) yields
\[
\tilde{g}_S(\tilde{\alpha}_n) = \frac{1}{n} \sum_t \left\{ g_S(U_{1t}, U_{2t}; \alpha_0) + g'_S(U_{1t}, U_{2t}; \tilde{\alpha}_n)(\tilde{\alpha}_n - \alpha_0) \right\},
\]
where \( \tilde{\alpha}_n \) is between \( \alpha_0 \) and \( \tilde{\alpha}_n \).

Under Condition T6, we have
\[
\sup_{\alpha \in A: \|\alpha - \alpha_0\| = o(1)} \left\| \frac{1}{n} \sum_t g_{S,\alpha}(U_{1t}, U_{2t}; \alpha) - G_{S,\alpha} \right\| = o(1).
\]
It follows that (A.15) can be written as,
\[
\tilde{g}_S(\tilde{\alpha}_n) = \frac{1}{n} \sum_t g_S(U_{1t}, U_{2t}; \alpha_0) + G'_S(\tilde{\alpha}_n - \alpha_0) + o_p(n^{-1/2}).
\]

Next Theorem 5 indicates that \( \tilde{\alpha}_n \) can be expressed as an asymptotically linear estimator such that
\[
\tilde{\alpha}_n - \alpha_0 = B_n^{-1} \frac{1}{n} \sum_t l_\alpha(U_{1t}, U_{2t}; \alpha_0) + o_p(1).
\]
Plugging (A.17) into (A.16) yields,
\[
\tilde{g}_S(\tilde{\alpha}_n) = \frac{1}{n} \sum_t \left\{ g_S(U_{1t}, U_{2t}; \alpha_0) + G'_{S,\alpha} B_n^{-1} l_\alpha(U_{1t}, U_{2t}; \alpha_0) \right\} + o_p(1)
\]
By the mean-value theorem, we expand (A.18) to obtain
\[
\tilde{g}_S(\tilde{\alpha}_n) = \frac{1}{n} \sum_t \left\{ g_S(U_{1t}, U_{2t}; \alpha_0) + G'_{S,\alpha} B_n^{-1} l_\alpha(U_{1t}, U_{2t}; \alpha_0) \right\} + J_n + o_p(1)
\]
where

\[ J_n = \frac{1}{n} \sum_j \sum_t \left[ g_{S,j}(\bar{U}_{1t}, \bar{U}_{2t}; \alpha_0) + G'_{S,\alpha}B^{-1}_n l_{\alpha_j}(\bar{U}_{1t}, \bar{U}_{2t}; \alpha_0) \right](\tilde{U}_{jt} - U_{jt}), \]

in which \((\bar{U}_{1t}, \bar{U}_{2t})\) lie on the line segment between \((U_{1t}, U_{2t})\) and \((\tilde{U}_{1t}, \tilde{U}_{2t})\).

By Lemma 1, we have

\[ J_n = -\frac{1}{n} \sum_j \sum_t \left\{ g_{S,j}(\bar{U}_{1t}, \bar{U}_{2t}; \alpha_0) + G'_{S,\alpha}B^{-1}_n l_{\alpha_j}(\bar{U}_{1t}, \bar{U}_{2t}; \alpha_0) \right\} \]

\[ S_j(X_{jt}) \int_{-\infty}^{X_{jt}} \frac{\tilde{S}_j(u-)}{\tilde{S}_j(u-)} \sum_s dM_{js}(u) \sum_s I(X_{js} \geq u). \]

Then following the same logic of the proof in Proposition 3.2 of Chen et al. (2010), we can see that \(\tilde{g}_S(\tilde{\alpha}_n)\) is asymptotically a sum of independent zero-mean random vectors. Given Condition T7 (i), Theorem 6 now follows from the standard multivariate central limit theorem for independent but non-identically distributed random variables.

\[ \square \]

**Proof of Theorem 7.** This theorem can be derived by following the proofs of Theorem 3 in this paper. For brevity, the proof is not reproduced here.

\[ \square \]

**Proof of Theorem 8.** The proof of this theorem is essentially identical to that of Theorem 4 and therefore is omitted.

\[ \square \]