

Nonparametric Estimation of Copula Functions For Dependence Modeling

Song Xi Chen¹

Tzeeming Huang²

Department of Statistics

Department of Statistics

Iowa State University

Iowa State University

November 1, 2005

Copulas are full measures of dependence among components of random vectors. Unlike the marginal and the joint distributions which are directly observable, a copula is a hidden dependence structure that couples the marginals and the joint distribution. This makes the task of proposing a parametric copula model non-trivial and is where a nonparametric estimator can play a significant role. In this paper, we investigate a kernel estimator which is mean square consistent everywhere in the support of the copula function. The kernel estimator is then used to formulate a goodness-of-fit test for parametric copula models.

KEY WORDS: Boundary Bias; Dependence; Goodness-of-fit tests; Kernel Estimator; Parametric copula models.

1 Introduction

Quantifying the dependence among two or more random variables has been an enduring task for statisticians. A rich set of dependence measures has been proposed, including the well-known Pearson's correlation coefficients, Kendall's tau and Spearman's rho for bivariate random variables, and in the case of more than two variables the partial and multiple correlation coefficients. While these measures are simple and can be easily computed, they are designed to capture only certain aspect of dependence. For instance the Pearson's correlation coefficient and its multivariate variates are catered for linear dependence. Indeed, it is rather unreasonable to expect a single scalar measure to have the capability to quantify all the dependence existed among the random variables.

Copula is a device that offers a complete description of the dependence among components of a random vector. Let $X = (X_1, X_2)^T$ be a bivariate random vector, and F be the distribution function of X with marginal distributions F_i for $i = 1$ and 2 . The Sklar's Theorem (Sklar, 1959; Schweizer and Sklar, 1974) assures the existence of a bivariate distribution function C on $[0, 1]^2$ such that

$$F(x_1, x_2) = C\{F_1(x_1), F_2(x_2)\}. \quad (1)$$

The function C is called the copula associated with X and couples the joint distribution F with its two marginals. If each marginal distribution F_i is continuous, then the C in (1) is unique. See Nelson (1998) for an comprehensive overview of copulas and their mathematical properties.

The implication of the Sklar's Theorem is that, after standardizing the effects of marginals, the dependence among components of X is fully described by the copula. Indeed, most conventional measures of dependence can be explicitly expressed in terms of the copula. For instance, the Kendall's tau between X_1 and X_2 is $4 \int_{[0,1]^2} C(u, v) dC(u, v) - 1$ and Spearman's rho is $12 \int_{[0,1]^2} C(u, v) dudv - 3$. The independence between X_1 and X_2 corresponds to $C(u, v) = uv$ and the positive quadrant dependence of Lehmann (1966) can be characterized as $C(u, v) \geq uv$. Copula can be also used to describe tail dependence, an important notion in risk management. The interest there is in the dependence between two extreme (risky) events. In particular, two random variables are said to be upper tail dependent if $\lim_{u \rightarrow 1} (1 - 2u + C(u, v))/(1 - u)$ has a limit in $(0, 1]$ (Joe, 1997).

Estimation of copulas can be achieved fully parametrically by assuming parametric models for both the copula and the marginals followed by the standard maximum likelihood estimation (Oakes, 1982 in the context of Clayton copula). Semiparametric estimation that specifies a parametric copula while leaving the marginals nonparametric is proposed in Oakes (1986) and Genest, Ghoudi and Rivest (1995). Estimation for the semiparametric family of Archmedian copulas as in Genest and Rivest (1993) can be considered as semiparametric too.

A nonparametric estimation of copula treats both the copula and the marginals parameter-free and offers the greatest generality. Unlike the marginal and the joint distributions which are directly observable, a copula is a hidden dependence structure. This makes the task of proposing a suitable parametric copula model non-trivial and is where a nonparametric estimator can play a significant role. Indeed, a nonparametric copula estimator can provide initial information needed in unlocking and subsequent formulation of a underlying parametric copula model.

The objectives of this paper are to proposed a nonparametric kernel copula estimator which are consistent everywhere in $[0, 1]^2$; and to evaluate its statistical properties. Recently, Fermanian and Scaillett (2004) proposed a kernel estimator via a bivariate kernel distribution estimator on the estimated marginals. However, as copulas are supported on a compact set $[0, 1]^2$, we need to exercise cares when formulating the kernel estimators. It is known in the kernel smoothing literature (Müller, 1991, 1993; Fan and Gijbels, 1992; and Jones, 1993) that kernel estimator encounters boundary bias due to a partial loss of kernel weight near the boundaries. An account on kernel estimation with multivariate boundary regions, which is the most relevant to the copula case, is given in Müller and Stadtmüller (1999). As it turns out the estimator of Fermanian and Scaillett (2004) is subject to the boundary bias which causes the estimator no longer consistent near all four edges of the unit square.

To remove the boundary bias we first employed local linear kernels to replace the standard kernel used in Fermanian and Scaillett (2004). However, employing the local linear kernel is not enough for the removal of the boundary bias near the upper and right edges of the unit square. We propose a new kernel copula estimator by subtracting certain mathematically known terms from the local linear kernel estimator. Explicit expressions for the bias and variance of this estimator are derived, which reveal that the overall effect

of the kernel smoothing is a net reduction of the variance and the mean square error when compared to another nonparametric estimator based on the empirical distributions. It is found that the largest variance reduction is achieved by carrying out undersmoothing in the first stage estimation of the marginal distributions. We then propose a practical bandwidth selection method for the kernel estimator. A goodness-of-fit test based on the kernel estimator is proposed for testing a parametric copula model. parametric copula models.

The paper is organized as follows. The kernel copula estimator is proposed in Section 2. Its bias and variance are reported and analyzed in Section 3. Section 4 considers the mean square error of estimation and bandwidth selection. Results from a simulation study are reported in Section 5. Section 6 analyzes an Urinuum exploration data, where the kernel copula estimator is used to form goodness-of-fit tests for several parametric copula models.

2 A Kernel Estimator

The basic thrust for a kernel copula estimator is the fact that, when the two marginal distributions are continuous, the copula C is the unique joint distribution of $F_1(X_1)$ and $F_2(X_2)$. As copulas are not directly observable, a nonparametric copula estimator has to be formed in two stages: estimate the two marginals $(F_1(X_1), F_2(X_2))$ first and then formulate the copula estimator based on these estimated marginals.

Let K be a symmetric kernel supported on $[-1, 1]$ and $G(x) = \int_{-\infty}^x K(t)dt$ be the distribution of K . In the first stage the marginal distribution F_l is estimated by

$$\hat{F}_l(x) = n^{-1} \sum_{i=1}^n G\left(\frac{x - X_{il}}{b_l}\right)$$

with a bandwidth b_l for $l = 1$ and 2 ; see Bowman, Hall and Prvan (1998) for more details on this kernel distribution function estimator.

To prevent the boundary bias, we use in the second stage

$$K_{u,h}(x) = \frac{K(x)(a_2(u, h) - a_1(u, h)x)}{a_0(u, h)a_2(u, h) - a_1^2(u, h)},$$

a local linear version of K , to smooth at a $u \in [0, 1]$ with a bandwidth $h > 0$, where

$a_l(u, h) = \int_{(u-1)/h}^{u/h} t^l K(t) dt$ for $l = 0, 1$ and 2 , which was proposed by Lejeune and Sarda (1992) and Jones (1993). It is easy to check that $K_{u,h} = K$ for $u \in [h, 1-h]$.

Let $G_{u,h}(t) = \int_{-\infty}^t K_{u,h}(x) dx$ and $T_{u,h} = G_{u,h}((u-1)/h)$. A seemingly natural estimator of $C(u, v)$ would be

$$n^{-1} \sum_{i=1}^n G_{u,h} \left(\frac{u - \hat{F}_1(X_{i1})}{h} \right) G_{v,h} \left(\frac{v - \hat{F}_2(X_{i2})}{h} \right).$$

However, it is readily derived from the bias expression given in (2) in the next section that this naive estimator incurs a bias $uT_{v,h} + vT_{u,h} + T_{u,h}T_{v,h}$ near $u = 1$ or $v = 1$ due to the fact that the marginal distributions assume value 1 at $u = 1$ or $v = 1$. Since $T_{u,h}$ is entirely known upon given the kernel and h , the bias can be readily removed by subtraction. This leads to the proposed kernel copula estimator

$$\begin{aligned} \hat{C}(u, v) &= n^{-1} \sum_{i=1}^n G_{u,h} \left(\frac{u - \hat{F}_1(X_{i1})}{h} \right) G_{v,h} \left(\frac{v - \hat{F}_2(X_{i2})}{h} \right) \\ &\quad - (uT_{v,h} + vT_{u,h} + T_{u,h}T_{v,h}). \end{aligned} \quad (1)$$

It is noted that a single bandwidth h is used to smooth $\hat{F}_l(X_{il})$ for $l = 1$ and 2 in the second stage, as the quantile transformation has already achieved a uniform data standardization. Nevertheless, different bandwidths can be used without altering the main results.

3 Main Results

The study of the statistical properties of the copula estimator (1) faces two challenges. One is that taking care of the boundary bias complicates the analysis. The other is that the estimator uses $\hat{F}_l(X_i)$ instead of $F_l(X_i)$, which largely increases the labor of derivation. However, despite the challenges we are able to obtain a quite simple expression for the mean integrated square error (MISE) of the copula estimator.

Let $C_u(u, v)$ and $C_v(u, v)$ be the first partial derivatives and $C_{uu}(u, v)$ and $C_{vv}(u, v)$ be the second partial derivatives of $C(u, v)$ with respect to u and v respectively. Put $\nu(u, h) = \int_{(u-1)/h}^{u/h} s^2 dG_{u,h}(s)$, which equals $\sigma_K^2 = \int s^2 K(s) ds$ if $u \in [h, 1-h]$. Derivations given in the appendix show that under conditions A1-A3 given in the appendix for any

$(u, v) \in [0, 1]^2$,

$$\begin{aligned}
& E\{\hat{C}(u, v)\} \\
&= C(u, v) + \frac{1}{2}h^2\{C_{uu}(u, v)\nu(u, h) + C_{vv}(u, v)\nu(v, h)\} + o(h^2) \\
&- \frac{1}{2}\sigma_K^2 \left[\{C_u(u, v) + T_{v,h}\}f_1^{(1)}\{F_1^{-1}(u)\}b_1^2 + \{C_v(u, v) + T_{u,h}\}f_2^{(1)}\{F_2^{-1}(v)\}b_2^2 \right] \quad (2)
\end{aligned}$$

and

$$\begin{aligned}
& Var\{\hat{C}(u, v)\} \\
&= n^{-1}Var\{I(U \leq u, V \leq v) - C_u(u, v)I(U \leq u) - C_v(u, v)I(V \leq v)\} \\
&- hn^{-1}[(C_u(u, v) + T_{v,h})^2\mu_1(u, h, b_1/h) + (C_v(u, v) + T_{u,h})^2\mu_2(v, h, b_2/h)] \\
&+ 2hn^{-1}\left[(C_v(u, v) + T_{u,h})^2\mu_2^*(v, h, b_2/h) + (C_u(u, v) + T_{v,h})^2\mu_1^*(u, h, b_1/h)\right] \\
&- hn^{-1}(C_u(u, v)(1 + 2T_{v,h}) + T_{v,h}^2) \int_{\frac{u-1}{h}}^{\frac{u}{h}} sdG_{u,h}^2(s) \\
&- hn^{-1}(C_v(u, v)(1 + 2T_{u,h}) + T_{u,h}^2) \int_{\frac{v-1}{h}}^{\frac{v}{h}} tdG_{v,h}^2(t) + o(hn^{-1}). \quad (3)
\end{aligned}$$

where, for $l = 1$ and 2

$$\begin{aligned}
\mu_l(v, h, \lambda) &= \int \int \int_{\frac{v-1}{h}}^{\frac{v}{h}} \int_{\frac{v-1}{h}}^{\frac{v}{h}} \max\{r_1 + f_l(F_l^{-1}(v))\lambda w_1, r_2 + f_l(F_l^{-1}(v))\lambda w_2\} \times \\
&\quad \times dG_{v,h}(r_2)dG_{v,h}(r_1)dG(w_1)dG(w_2)
\end{aligned}$$

and $\mu_l^*(v, h, \lambda) = \int \int_{\frac{v-1}{h}}^{\frac{v}{h}} \int_{\frac{v-1}{h}}^{\frac{v}{h}} \max\{t, r + f_l(F_l^{-1}(v))\lambda w\} dG_{v,h}(r)dG_{v,h}(t)dG(w)$.

The results in (2) and (3) can be easily extended to the case where the first stage kernel estimators \hat{F}_1 and \hat{F}_2 are replaced by the empirical distribution functions by setting $b_1 = b_2 = 0$.

While the bias conveys a simple story that smoothing in the two stages all contributes to the bias, the variance expression requires a further analysis. Let us concentrate our attention on $(u, v) \in [h, 1-h]^2$, the interior region. Let $x_u = f_1(F_1^{-1}(u))$, $y_v = f_2(F_2^{-1}(v))$, and $b_K = \int tdG^2(t)$. Then (3) can be simplified to

$$\begin{aligned}
Var\{\hat{C}(u, v)\} &= n^{-1}Var\{I(U \leq u, V \leq v) - C_u(u, v)I(U \leq u) - C_v(u, v)I(V \leq v)\} \\
&+ hn^{-1}\{C_u^2(u, v)\rho(x_u b_1/h) + C_v^2(u, v)\rho(y_v b_2/h)\} \\
&- hn^{-1}\{C_u(u, v) - C_u^2(u, v) + C_v(u, v) - C_v^2(u, v)\}b_K + o(hn^{-1}), \quad (4)
\end{aligned}$$

where $\rho(\lambda) = 2\mu^*(\lambda) - \mu(\lambda) - b_K$,

$$\mu(\lambda) = \int \int \int \int \max\{r_1 + \lambda w_1, r_2 + \lambda w_2\} dG(r_2) dG(r_1) dG(w_1) dG(w_2)$$

and $\mu^*(\lambda) = \int \int \int \max\{t, r + \lambda w\} dG(r) dG(t) dG(w)$.

A key fact needed in understanding (4) is that

$$\rho(\lambda) \geq 0 \quad \text{for any } \lambda \geq 0 \text{ and is minimized at } \lambda = 0. \quad (5)$$

In order to achieve the largest variance reduction, we need to minimize the second term on the RHS of (4) which involves the ρ -function and is positive. Our strategy is to choose $b_l = o(h)$ for $l = 1$ and 2 so that both $\rho(x_u b_1/h)$ and $\rho(y_v b_2/h)$ are $o(1)$, and hence

$$\begin{aligned} \text{Var}\{\hat{C}(u, v)\} &= n^{-1} \text{Var}\{I(U \leq u, V \leq v) - C_u(u, v)I(U \leq u) - C_v(u, v)I(V \leq v)\} \\ &\quad - hn^{-1}\{C_u(u, v) - C_u^2(u, v) + C_v(u, v) - C_v^2(u, v)\}b_K + o(hn^{-1}). \end{aligned}$$

This strategy leads to a second order variance reduction by the smoothing carried in the second stage. Despite this variance reduction happens in the interior region, it leads to a net reduction in the overall MISE over $[0, 1]^2$ as shown in the next section.

We note that the leading variance term of order n^{-1} coincides with that of an unsmoothed copula estimator

$$\tilde{C}(u, v) = n^{-1} \sum_{i=1}^n I(\hat{U}_i \leq u, \hat{V}_i \leq v) \quad (6)$$

where $\hat{U}_i = n^{-1} \sum_{j=1}^n I(X_{j1} \leq X_{i1})$ and $\hat{V}_i = n^{-1} \sum_{j=1}^n I(X_{j2} \leq X_{i2})$ are the marginal empirical distributions at X_{i1} and X_{i2} respectively. One drawback of $\tilde{C}(u, v)$ is its lacks of continuity, which makes it less attractive as a copula estimator for continuous random variables. A simulation study reported in Section 5 reveals that the MISE of \tilde{C} can be as twice as that of the kernel estimator, which indicates the variance and MISE reductions are significant in finite samples.

4 Bandwidth Selection

The findings of the previous section suggests that we should undersmooth in the first stage to reduce the variance, which reduces the bias due to the first stage smoothing too.

Therefore, b_l should be $o(h)$ for $l = 1$ and 2 by assigning smaller values relative to h . This strategy largely simplifies the expressions of (2) and (4) and leads to a tractable expression for the mean square error (MSE) for $(u, v) \in [h, 1 - h]^2$

$$\begin{aligned} MSE\{\hat{C}(u, v)\} &= n^{-1}Var\{I(U \leq u, V \leq v) - C_u(u, v)I(U \leq u) - C_v(u, v)I(V \leq v)\} \\ &\quad - hn^{-1}b_K[C_u(u, v)\{1 - C_u(u, v)\} + C_v(u, v)\{1 - C_v(u, v)\}] \\ &\quad + \frac{1}{4}h^4\sigma_k^4\{C_{uu}(u, v) + C_{vv}(u, v)\}^2 + o(h^4 + hn^{-1}). \end{aligned} \quad (7)$$

As the area of the boundary regions are of $O(h)$ and the leading variance term is valid throughout the entire $[0, 1]^2$, the MISE of \hat{C} is

$$\begin{aligned} MISE(\hat{C}) &= n^{-1} \int_0^1 \int_0^1 Var\{I(U \leq u, V \leq v) - C_u(u, v)I(U \leq u) \\ &\quad - C_v(u, v)I(V \leq v)\} dudv - hn^{-1}\alpha + \frac{1}{4}h^4\sigma_k^4\beta + o(hn^{-1} + h^4) \end{aligned}$$

where

$$\begin{aligned} \alpha &= b_K \int_0^1 \int_0^1 [C_u(u, v)\{1 - C_u(u, v)\} + C_v(u, v)\{1 - C_v(u, v)\}] dudv, \\ \beta &= \int_0^1 \int_0^1 \{C_{uu}(u, v) + C_{vv}(u, v)\}^2 dudv. \end{aligned}$$

The optimal h that minimizes the above MISE is then

$$h^* = \sigma_K^{4/3}(\alpha/\beta)^{1/3}n^{-1/3}. \quad (8)$$

Various plug-in bandwidth selection rules that have been used in kernel smoothing can be adopted here to attain an estimate for the optimal bandwidth. One approach is to estimate α and β and then substitute them to (8). A simple approach is to assume certain parametric family for the copula function which leads to parametric expressions for α and β , which is similar to the approach suggested by Silverman (1986) for kernel density estimation. The parameters of the parametric copula can be estimated by either the pseudo-maximum likelihood estimation of Genest, Ghoudi and Rivest (1995) or the method of moments of Genest and Rivest (1993).

We propose the following T -copula as the reference copula which is defined to be

$$C(u, v) = \int_{-\infty}^{t_m^{-1}(u)} \int_{-\infty}^{t_m^{-1}(v)} \frac{1}{2\pi(1 - \rho^2)^{1/2}} \left\{ 1 + \frac{s^2 - 2\rho st + t^2}{m(1 - \rho^2)} \right\}^{- (m+2)/2} dsdv \quad (9)$$

which has two parameters: the degree of freedom m and the correlation coefficient ρ . It contains the normal copula as its limit and accommodates a wide range of tail-thickness and tail-dependence. The estimation of the parameters in the T -copula can be done by the method of moment to avoid the intensive computation in the estimation of the degree of freedom parameter by the pseudo-maximum likelihood approach. The simulation studies in the next section demonstrate that this proposal work well for a range of copula models.

5 Simulation Studies

We report results from simulation studies which are designed to confirm the theoretical findings in Section 3 and the proposed bandwidth selection method in Section 4. To demonstrate the advantage of kernel smoothing, the kernel estimator is compared with the unsmoothed estimator \tilde{C} given in (6).

Three copulas are considered in the simulation study, which are respectively

$$C(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)}, \quad (10)$$

the Ali-Mikhail-Haq (AMH) family with $\theta = 1$;

$$C(u, v) = \exp(-[(-\log u)^\theta + (-\log v)^\theta]^{1/\theta}) \quad (11)$$

the Gumbel copula with $\theta = 2$; and

$$C(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right\} ds dt$$

the normal copula where Φ is the standard normal distribution function and ρ is the correlation coefficient and was set at 0.5 in the simulation.

For each copula model, we first generate independent and identically distributed uniform random variables $\{U_i\}_{i=1}^n$. Then, generate V_i from the conditional copula distribution given U_i , which is known under each model of simulation. The sample sizes considered are $n = 50$ and 100 . We choose $b_1 = b_2 = b$ in the first stage estimation as the marginals have already been standardized.

The first simulation study is designed to check on the effect of smoothing at each of the two stages. Twenty equally spaced bandwidths are chosen for b and h respectively,

whose ranges are displayed in Figures 1 to 3 respectively. For each given pair (b, h) , the MISE and mean integrated variance (MIV) of the kernel and the unsmoothed copula estimators are evaluated over 40×40 equally spaced grid points within $[0, 1]^2$. The MISEs and MIVs based on 1000 simulations are reported in Figures 1-3 for the three copulas respectively.

The results conveyed by Figures 1-3 can be summarized as follows. First of all, the smoothing at the first stage has little effect on the variance of the kernel estimator for all the three copulas and sample sizes considered. In particular, the shapes of the MISE and MIV contours coincide with our early predictions that (i) the role of first stage smoothing is in the bias and has little affect on the variance as ρ is slow varying and (ii) variance reduction is largely due to the second stage smoothing. The simulation also shows that kernel smoothing leads to a substantial improvement in estimation accuracy comparing with the unsmoothed estimator. Indeed, for $n = 50$ and each of the copula models considered, the MISE of the kernel estimator is nearly half of the unsmoothed estimator. There is still around 30% advantage for the kernel estimator when the sample size is 100.

To evaluate the practical performance of the proposed reference to the T -copula rule for selecting h -bandwidth, we conducted simulations for the same three copula models to obtain the MISEs of the kernel copula estimator using (i) the prescribed reference-rule and (ii) a set of fixed bandwidths, respectively, while setting $b_1 = b_2 = 10^{-4}$ to realize the strategy of undersmoothing in the first stage. The results of the simulation are displayed in Figure 4, which shows that the reference rule is able to achieve a level of MISE which is consistently close to the minimum MISE of the fixed bandwidth estimator. This is particularly encouraging as all the three copulas are not the T -copula and indicates the proposed rule is robust against mis-specifying the copula model in bandwidths selection.

6 Goodness-of-Fit Test and Empirical Study

We carry out an empirical study on a set of Uranium exploration data collected from water samples in the Montrose quadrangle in Colorado, which was originally studied in Cook and Johnson (1994). The same dataset has been analyzed by Genest and Rivest (1993) and Genest, Quessy and Remillard (2005) to demonstrate a semiparametric inference for Archmedian copulas. The dataset contains 655 log-concentrations of seven chemical

elements including Uranium, Caesium and Lithium. A primary interest is to understand the dependence in concentrations between an actinide metal Uranium and two alkali metals, Caesium and Lithium.

Figure 5 displays the original data in panel (a) for Uranium versus Caesium and in panel (c) for Uranium versus Lithium. The kernel copula estimators are displayed in panels (b) and (d) with the h -bandwidth chosen by the proposed reference rule which assigns $h = 0.176$ for Uranium versus Caesium $= 0.176$, and $h = 0.143$ for Uranium versus Lithium, whereas $b_1 = b_2 = 10^{-4}$. Figure 5 also displays the independent copula alone with the kernel copula estimates.

The objective of the empirical study is to find a copula model for the two pairs of chemical elements which best describes the underlying dependence structure. We considered four parametric copulas which are respectively the AMH copula (10), the Gumbel copula (11), the Clayton copula

$$C_\theta(u, v) = \max\{(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, 0\} \quad (12)$$

which was used in Cook and Johnson (1981)'s original study and the T -copula (9).

The parameter of each copula needs to be estimated before we can check on the adequacy of each copula model. The first three copulas are members of the Archmedian family (Nelson, 1998), which can be expressed as $C(u, v) = \phi^{-1}\{\phi(u) + \phi(v)\}$ for a convex decreasing function ϕ (the generator) such that $\phi(1) = 0$. The generator $\phi(t)$ is $\log\{\frac{1-\theta(1-t)}{t}\}$ for the AMH copula, $\{-\log(t)\}^\theta$ for the Gumbel copula and $(t^{-\theta} - 1)/\theta$ for the Clayton copula. We use the semiparametric approach proposed in Genest and Rivest (1993) which is based on the following moment equation regarding Kendall's tau

$$\tau(X_1, X_2) = 4 \int \int C(u, v) dC(u, v) - 1 = 4 \int_0^1 \frac{\phi_\theta(u)}{\phi'_\theta(u)} du + 1. \quad (13)$$

The parameter θ can be estimated by the method of moments after replacing $\tau(X_1, X_2)$ by its sample version. The parameter of the T -copula is estimated by the method of moments too.

Let $\hat{\theta}$ be the method of moments estimate and $C_{\hat{\theta}}$ be the estimated parametric copula function. Figures 6 and 7 display the four parametric copulas at $\hat{\theta}$ and the kernel copula estimate for the two pairs of chemical elements. Copulas are monotone non-decreasing with respect to each variable and in particular the contour curves $\{(u, v) | C(u, v) = t\}$

are all confined in a triangle with vertices (t, t) , $(1, t)$ and $(t, 1)$. These features make the copula estimates have similar shape and look similar, which are well reflected in Figures 6 and 7. Hence to check on the goodness-of-fit of a parametric copula model, a formal test procedure is needed as visual diagnostics are harder to detect any differences.

We propose the following Cramér-Von Mises type test statistic

$$T_n = n \int_0^1 \int_0^1 \{\hat{C}(u, v) - C_{\hat{\theta}}(u, v)\}^2 dudv$$

which is essentially a L_2 -distance between the kernel estimator \hat{C} and the estimated hypothesized parametric model $C_{\hat{\theta}}$.

Let c_α be the upper- α quantile of the test statistic T_n given a level of significance α , and \hat{c}_α be an estimate of c_α . Then, a goodness-of-fit test rejects H_0 if $T_n \geq \hat{c}_\alpha$. The following semiparametric bootstrap procedure is employed to obtain the critical value \hat{c}_α :

Step 1: Generate $\{X_{i1}^*\}_{i=1}^n$ from F_{n1} , the empirical distribution of $\{X_{n1}\}_{i=1}^n$ by sampling with replacement, and let $U_i^* = F_{n1}(X_{i1}^*)$ for $i = 1, \dots, n$;

Step 2: Generate V_i^* from $C_{\hat{\theta}, U_i^*}(v) = \frac{\partial C_{\hat{\theta}}(u, v)}{\partial u}|_{u=U_i^*}$, the conditional distribution of V given $U = U_i^*$, and let $X_{i2}^* = F_{n2}^{-1}(V_i^*)$ where F_{n2} is the empirical distribution of $\{X_{i2}\}_{i=1}^n$. Then $\{(X_{i1}^*, X_{i2}^*)\}_{i=1}^n$ constitutes a bootstrap resample which respects both the parametric copula and the two marginals.

Step 3: Construct $\hat{C}^*(u, v)$, the kernel estimator based on the bootstrap resample using the same h as in T_n and let $T_n^* = n \int_0^1 \int_0^1 \{\hat{C}^*(u, v) - C_{\hat{\theta}^*}(u, v)\}^2 dudv$ where $\hat{\theta}^*$ is the parameter estimate based on the resample.

Step 4: Repeat the above steps B times for a large integer B and obtain, without loss of generality, $T_{n1}^* \leq \dots \leq T_{nB}^*$. Compute $c_\alpha = T_{n[B(1-\alpha)]+1}^*$ be the upper α -th order statistic.

We apply the above procedure to test for the four copulas for Uranium versus Caesium and Uranium versus Lithium respectively with detail results summarized in Table 1. It is found that both AMH and Clayton copulas are overwhelmingly rejected for both pairs despite that Clayton copula was the one used in Cook and Johnson's original study. Gumbel copula is rejected for Uranium versus Caesium but not for Uranium versus Lithium. The T-copula seems to provide the best dependence description for both pairs of data especially for the pair of Uranium versus Lithium. The goodness-of-fit offered by the T-copula echoes some results in empirical finance (Embrechts, Lindskog and McNeil,

2001), which has been shown to be robust in fitting financial data which typically have heavy tails and tail dependence.

Acknowledgment: We thank Cheng Yong Tang for valuable computation assistance and Professor Christan Genest for making the Uranium exploration data available to us. We also acknowledge the support of an Iowa State University Start-up grant.

APPENDIX: TECHNICAL DETAILS

A.1 Derivation of (2) and (3)

The following conditions are assumed in our analysis:

A1: K is a symmetric and continuous probability density function supported on $[-1, 1]$, and the smoothing bandwidths satisfy $h = O(n^{-1/3})$ and $b_l = O(h)$ for $l = 1$ and 2 .

A2: For $l = 1$ or 2 , X_l is absolute continuous with a probability density function f_l which has bounded first two derivatives that vanishes at $\pm\infty$.

A3: The copula C has a probability density function f and there exists a C^∞ function g such that $f = g$ on $[0, 1]^2$.

Here we provide an outline on the derivations of (2) and (3) which is fundamentally based on a Taylor expansion of $\hat{C}(u, v)$ with negligible remainder terms. Detailed derivations can be found in a technical report (Chen and Huang, 2005).

For $i = 1, \dots, n$ and $k = 1$ and 2 , let $\Delta_{1,i} = F_1(X_{i1}) - \hat{F}_1(X_{i1})$ and $\Delta_{2,i} = F_2(X_{i2}) - \hat{F}_2(X_{i2})$ and for $j, k \geq 0$, let

$$I_{j,k}(s, t) = G_{u,h}^{(j)}\left(\frac{u-s}{h}\right) G_{v,h}^{(k)}\left(\frac{v-t}{h}\right).$$

Then a Taylor expansion for $\hat{C}(u, v)$ to the fifth order is

$$\hat{C}(u, v) = \sum_{(k,j) \in S} A_{k,j} + R_n - (vT_{u,h} + uT_{v,h} + T_{u,h}T_{v,h}), \quad (\text{A.1})$$

where $S = \{(j, k) : j \geq 0, k \geq 0, j + k \leq 5\}$,

$$A_{j,k} = n^{-1} \sum_{i=1}^n \frac{1}{j!k!} I_{j,k}(F_1(X_{i1}), F_2(X_{i2})) \left(\frac{\Delta_{1,i}}{h}\right)^j \left(\frac{\Delta_{2,i}}{h}\right)^k,$$

and

$$R_n = \sum_{j=0}^6 \frac{1}{(n)j!(6-j)!} \sum_{i=1}^n I_{j,6-j}(F_1(X_{i1}) - \theta\Delta_{1,i}, F_2(X_{i2}) - \theta\Delta_{2,i}) \left(\frac{\Delta_{1,i}}{h}\right)^j \left(\frac{\Delta_{2,i}}{h}\right)^{6-j}$$

for some $0 \leq \theta \leq 1$. Note that although we need to assume that the kernel K has high order derivatives in (A.1), it is argued in Chen and Huang (2005) that this assumption can be removed so that Assumption A.1 is sufficient.

It can be shown that $E(R_n^2) = o(h^2/n)$, $E(A_{j,k}) = o(h^2)$ for $j + k \geq 2$,

$$\begin{aligned} E(A_{0,0}) &= C(u, v) + \frac{1}{2}h^2 (C_{uu}(u, v)\nu(u, h) + C_{vv}(u, v)\nu(v, h)) \\ &\quad + vT_{u,h} + uT_{v,h} + T_{u,h}T_{v,h} + o(h^2), \\ E(A_{1,0}) &= -\frac{1}{2}\sigma_K^2 b_1^2 \{C_u(u, v) + T_{v,h}\} f_1^{(1)}(F_1^{-1}(u)) + o(h^2), \\ E(A_{0,1}) &= -\frac{1}{2}\sigma_K^2 b_1^2 \{C_v(u, v) + T_{u,h}\} f_2^{(1)}(F_2^{-1}(v)) + o(h^2). \end{aligned}$$

This leads to (2) since

$$\begin{aligned} E(\hat{C}(u, v)) &= E(A_{0,0}) + E(A_{0,1}) + E(A_{1,0}) - (vT_{u,h} + uT_{v,h} + T_{u,h}T_{v,h}) + o(h^2) \\ &= C(u, v) + \frac{1}{2}h^2 (C_{uu}(u, v)\nu(u, h) + C_{vv}(u, v)\nu(v, h)) \\ &\quad - \frac{1}{2}\sigma_K^2 b_1^2 \{C_u(u, v) + T_{v,h}\} f_1^{(1)}(F_1^{-1}(u)) \\ &\quad - \frac{1}{2}\sigma_K^2 b_1^2 \{C_v(u, v) + T_{u,h}\} f_2^{(1)}(F_2^{-1}(v)) + o(h^2). \end{aligned}$$

To establish (3), we note that

$$\begin{aligned} Cov(A_{j,k}, A_{\tilde{j},\tilde{k}}) &= o(h^4) \quad \text{if } j + k + \tilde{j} + \tilde{k} \geq 3, \\ Cov(A_{0,0}, A_{0,2}) &= o(h^4), \quad Cov(A_{0,0}, A_{1,1}) = o(h^4), \\ Cov(A_{1,0}, A_{0,1}) &= n^{-1} \{C(u, v) - uv\} \{C_v(u, v) + T_{u,h}\} \{C_u(u, v) + T_{v,h}\} + o(h^4); \end{aligned}$$

$$Var(A_{0,1}) = n^{-1}v(1-v)(C_v(u, v) + T_{u,h})^2 - hn^{-1}(C_v(u, v) + T_{u,h})^2 \mu_2(v, h, b_2/h) + o(h^4),$$

$$Var(A_{1,0}) = n^{-1}u(1-u)(C_u(u, v) + T_{v,h})^2 - hn^{-1}(C_u(u, v) + T_{v,h})^2 \mu_1(u, h, b_1/h) + o(h^4);$$

$$\begin{aligned} Cov(A_{0,0}, A_{0,1}) &= -n^{-1} \{C_v(u, v) + T_{u,h}\} \{(1-v)C(u, v) + v(1-v)T_{u,h} \\ &\quad + (C(u, v) - uv)T_{v,h}\} + hn^{-1}(C_v(u, v) + T_{u,h})^2 \mu_2^*(v, h, b_2/h) + o(h^4), \end{aligned}$$

$$\begin{aligned} Cov(A_{0,0}, A_{1,0}) &= -n^{-1} \{C_u(u, v) + T_{v,h}\} \{(1-u)C(u, v) + u(1-u)T_{v,h} \\ &\quad + (C(u, v) - uv)T_{u,h}\} + hn^{-1}(C_u(u, v) + T_{v,h})^2 \mu_1^*(u, h, b_1/h) + o(h^4); \end{aligned}$$

and

$$Var(A_{0,0}) = n^{-1}C(u, v)(1 + 2T_{u,h})(1 + 2T_{v,h})$$

$$\begin{aligned}
& + n^{-1} \left(T_{u,h}^2 v(1 + 2T_{v,h}) + T_{v,h}^2 u(1 + 2T_{u,h}) + T_{u,h}^2 T_{v,h}^2 \right) \\
& - n^{-1} h(C_u(u, v)(1 + 2T_{v,h}) + T_{v,h}^2) \int_{\frac{u-1}{h}}^{\frac{u}{h}} s dG_{u,h}^2(s) \\
& - n^{-1} h(C_v(u, v)(1 + 2T_{u,h}) + T_{u,h}^2) \int_{\frac{v-1}{h}}^{\frac{v}{h}} t dG_{v,h}^2(t) - n^{-1} (E(A_{0,0}))^2 + o(h^4).
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \text{Var}(\hat{C}(u, v)) \\
& = \text{Var}(A_{0,0}) + 2\text{Cov}(A_{0,0}, A_{0,1}) + 2\text{Cov}(A_{0,0}, A_{1,0}) + \text{Var}(A_{0,1}) + \text{Var}(A_{1,0}) \\
& \quad + 2\text{Cov}(A_{0,1}, A_{1,0}) + o(h^4) \\
& = n^{-1} C(u, v)(1 + 2T_{u,h})(1 + 2T_{v,h}) \\
& + n^{-1} \left(T_{u,h}^2 v(1 + 2T_{v,h}) + T_{v,h}^2 u(1 + 2T_{u,h}) + T_{u,h}^2 T_{v,h}^2 \right) \\
& - n^{-1} h(C_u(u, v)(1 + 2T_{v,h}) + T_{v,h}^2) \int_{\frac{u-1}{h}}^{\frac{u}{h}} s dG_{u,h}^2(s) \\
& - n^{-1} h(C_v(u, v)(1 + 2T_{u,h}) + T_{u,h}^2) \int_{\frac{v-1}{h}}^{\frac{v}{h}} t dG_{v,h}^2(t) \\
& - n^{-1} (C(u, v) + vT_{u,h} + uT_{v,h} + T_{u,h}T_{v,h})^2 \\
& - 2n^{-1} (C_v(u, v) + T_{u,h}) ((1 - v)C(u, v) + v(1 - v)T_{u,h} + (C(u, v) - uv)T_{v,h}) \\
& + 2hn^{-1} (C_v(u, v) + T_{u,h})^2 \mu_2^*(v, h, b_2/h) \\
& - 2n^{-1} (C_u(u, v) + T_{v,h}) ((1 - u)C(u, v) + u(1 - u)T_{v,h} + (C(u, v) - uv)T_{u,h}) \\
& + 2hn^{-1} (C_u(u, v) + T_{v,h})^2 \mu_1^*(u, h, b_1/h) \\
& + n^{-1} v(1 - v)(C_v(u, v) + T_{u,h})^2 - hn^{-1} (C_v(u, v) + T_{u,h})^2 \mu_2(v, h, b_2/h) \\
& + n^{-1} u(1 - u)(C_u(u, v) + T_{v,h})^2 - hn^{-1} (C_u(u, v) + T_{v,h})^2 \mu_1(u, h, b_1/h) \\
& + 2n^{-1} (C(u, v) - uv)(C_v(u, v) + T_{u,h})(C_u(u, v) + T_{v,h}) + o(h^4). \tag{A.2}
\end{aligned}$$

To see that (3) follows from (A.2), note that if both u and v are in $(0, 1)$ $T_{v,h}$ and $T_{u,h}$ are $o(1)$ for h sufficiently small and hence can be ignored. Thus, all the n^{-1} terms in (A.2) sum to

$$\begin{aligned}
& n^{-1} \left(C(u, v) - C^2(u, v) + u(1 - u)C_u^2(u, v) + v(1 - v)C_v^2(u, v) \right) \\
& + 2n^{-1} \{ C_v(u, v)C_u(u, v)(C(u, v) - uv) - C_v(u, v)C(u, v)(1 - v) \\
& \quad - C_u(u, v)C(u, v)(1 - u) \} \\
& = n^{-1} \text{Var}\{ I(U \leq u, V \leq v) - C_u(u, v)I(U \leq u) - C_v(u, v)I(V \leq v) \}, \tag{A.3}
\end{aligned}$$

where $U = F_1(X_{11})$ and $V = F_2(x_{12})$. Also, if one of u, v is 0 or 1, then direct calculation shows that all the n^{-1} terms in (A.2) are canceled. Therefore, the quantity in (A.3) can be used to replace the sum of all the n^{-1} terms in (A.2) regardless the values of u and v , which gives (3).

A.2 Derivation of (5)

Let R_1 and R_2 be two independent and identically distributed random variables with probability density function K , and let H_λ be the distribution function for $R_1 + \lambda R_2$. Then for $\lambda > 0$, $\mu(\lambda) = \int tdH_\lambda^2(t)$, $\mu^*(\lambda) = \int tdG(t)H_\lambda(t)$ and

$$\rho(\lambda) = - \int td(H_\lambda(t) - G(t))^2 = \int (H_\lambda(t) - G(t))^2 dt \geq 0.$$

This readily implies the conclusion in (5).

REFERENCES

- Bowman, A., Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika*, 85, 799–808.
- Chen, S. and Huang, T. (2005), Bias and variance calculation for a copula kernel estimator, Technical Report, Department of Statistics, Iowa State University.
- Cook, R. D. and Johnson, M. E. (1986), Generalized Burr-Pareto-logistic distributions with applications to a uranium exploration data set, *Technometrics*, 28, 123-131.
- Fan, J. Q. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics* **20** 2008-2036.
- Fermanian, J. D. and Scaillet, O. (2004). Nonparametric estimation of copulas for time series, *Journal of Risk*, 5, 25-54.
- Genest, C., Ghoudi, K. and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, 82, 543–552.
- Genest, C. Quessy, J.-F. and Rémillard, B. (2005). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, to appear.
- Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of American Statistical Association*, 88, 1034–1043.

- Joe, H. (1997). *Multivariate Models and Dependence Concepts*, London: Chapman and Hall.
- Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistic Computing*, 3, 135-146.
- Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, 37, 1137-1153.
- Lejeune, M. and Sarda, P. (1992) Smooth estimators of distribution and density functions. *Computational Statistics & Data Analysis*, 14, 457-471.
- Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints, *Biometrika*, 78, 521-530.
- Müller, H.-G. (1993). On the boundary kernel method for non-parametric curve estimation near endpoints. *Scandinavian Journal of Statistics*, 20, 313-328.
- Müller, H. G. and Stadtmüller, U. (1999). Multivariate boundary kernels and a continuous least squares principle. *Journal of the Royal Statistical Society, Series B*, 61 (1999), 439–458.
- Nelsen, R. B. (1998). *An Introduction to Copulas*, New York: Springer.
- Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B*, 44, 414–422.
- Oakes, D. (1986). Semiparametric inference in a model for association in bivariate survival data. *Biometrika*, 73, 353–361.
- Schweizer, B., Sklar, A. (1983). *Probabilistic metric spaces*, North-Holland Series in Probability and Applied Mathematics. New York: North-Holland Publishing.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8, 229–231.
- Silverman, B.W.(1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

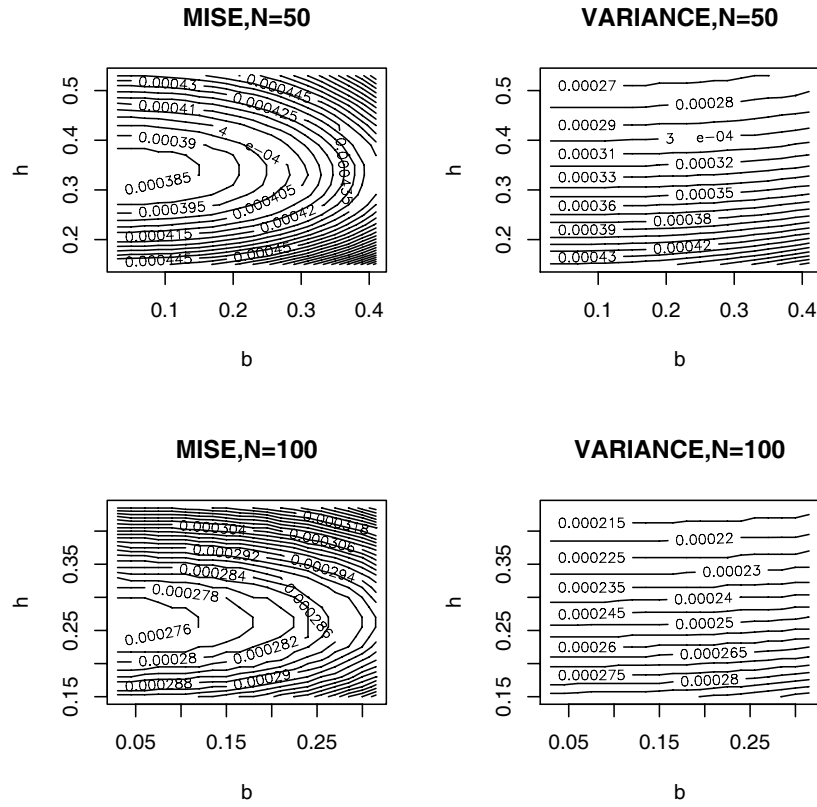


Figure 1. The MISE and Integrated Variance (IVAR) of kernel estimator as a function of two bandwidths (b, h) for AMH copula (10). The MISE and IVAR for the unsmoothed estimator (6) are 0.000715 and 0.000577 for $n = 50$; and 0.000391 and 0.000358 for $n = 100$, respectively.

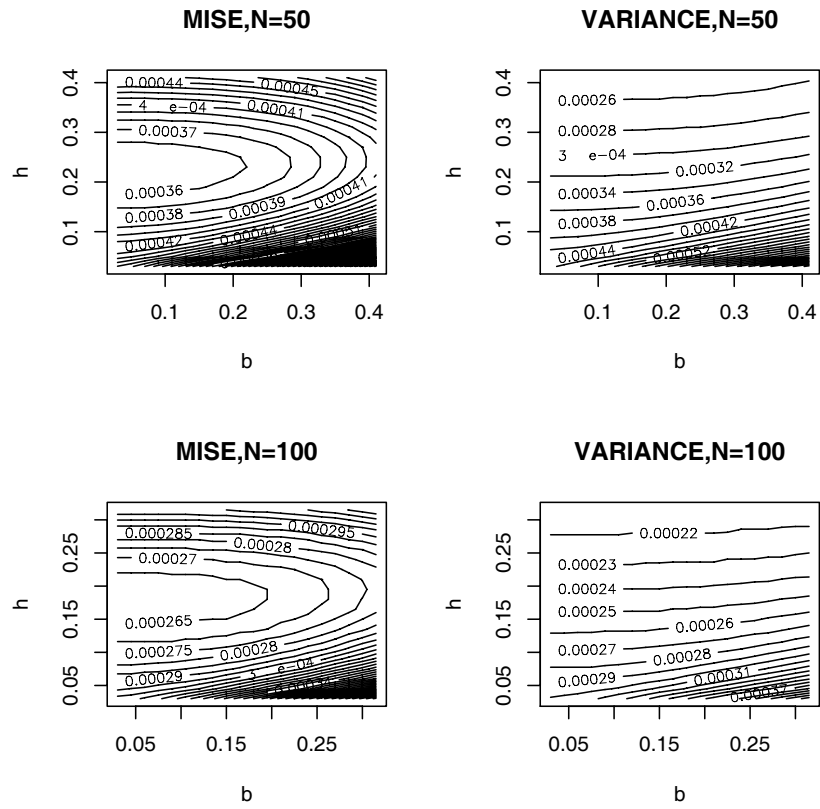


Figure 2. The MISE and Integrated Variance (IVAR) of kernel estimator as a function of two bandwidths (b, h) for Gumbel copula. The MISE and IVAR for the unsmoothed estimator (6) are 0.000649 and 0.000309 for $n = 50$; and 0.000427 and 0.000292 for $n = 100$, respectively.

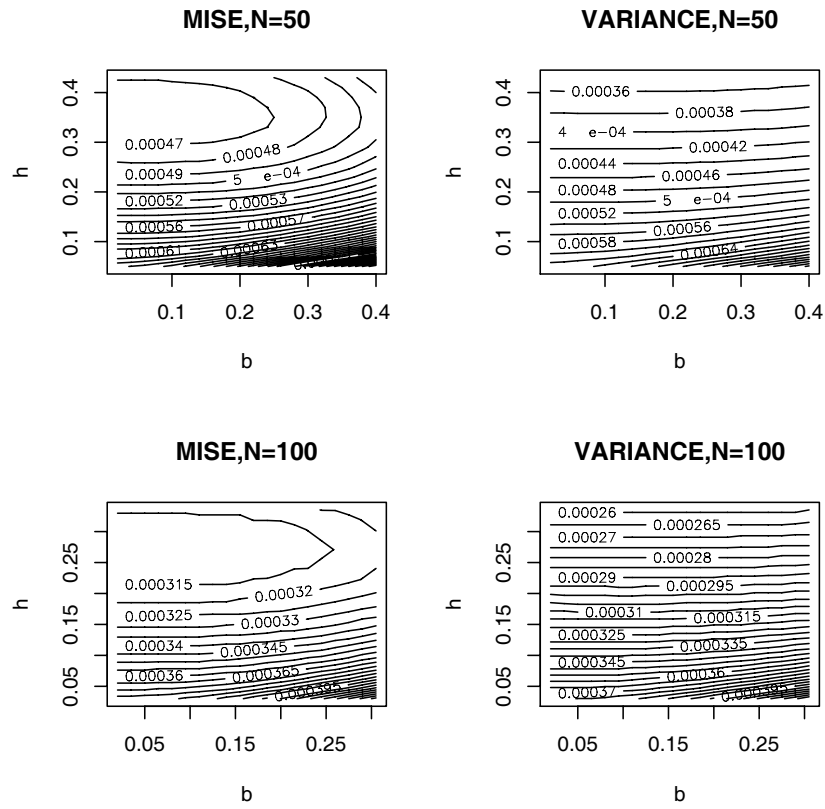


Figure 3. The MISE and Integrated Variance (IVAR) of kernel estimator as a function of two bandwidths (b, h) for the normal copula. The MISE and IVAR for the unsmoothed estimator (6) are 0.000854 and 0.000682 for $n = 50$; and 0.000443 and 0.000401 for $n = 100$, respectively.

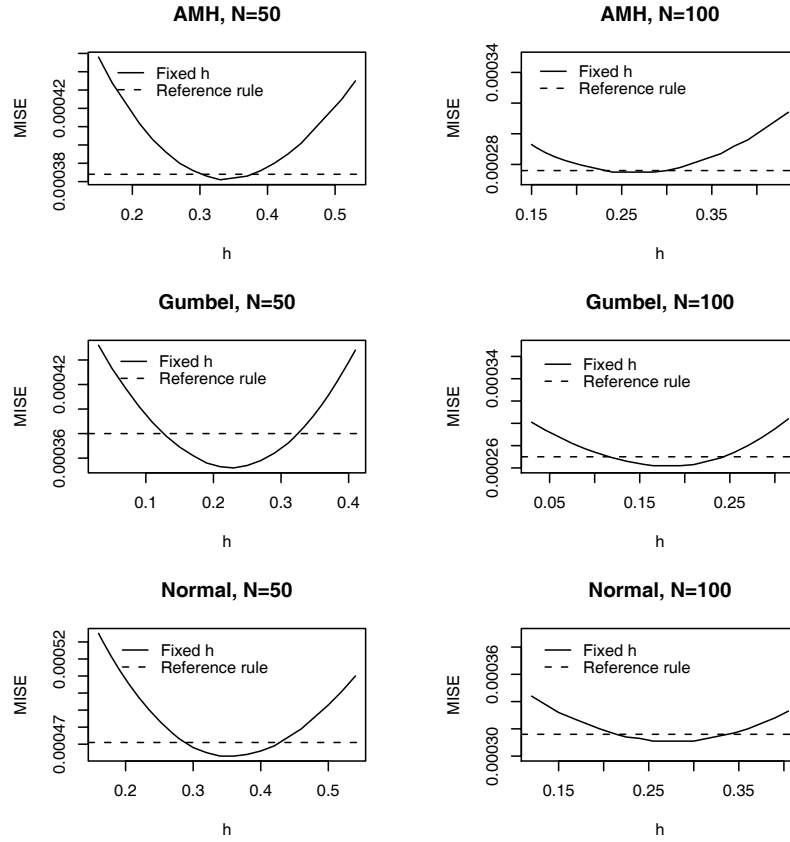


Figure 4. The MISE of the kernel estimators with the plug-in bandwidth (in dashed lines) and that with the fixed bandwidths (in solid lines). The two bandwidths b_1 and b_2 used in the first stage smoothing are set to be 10^{-4} respectively.

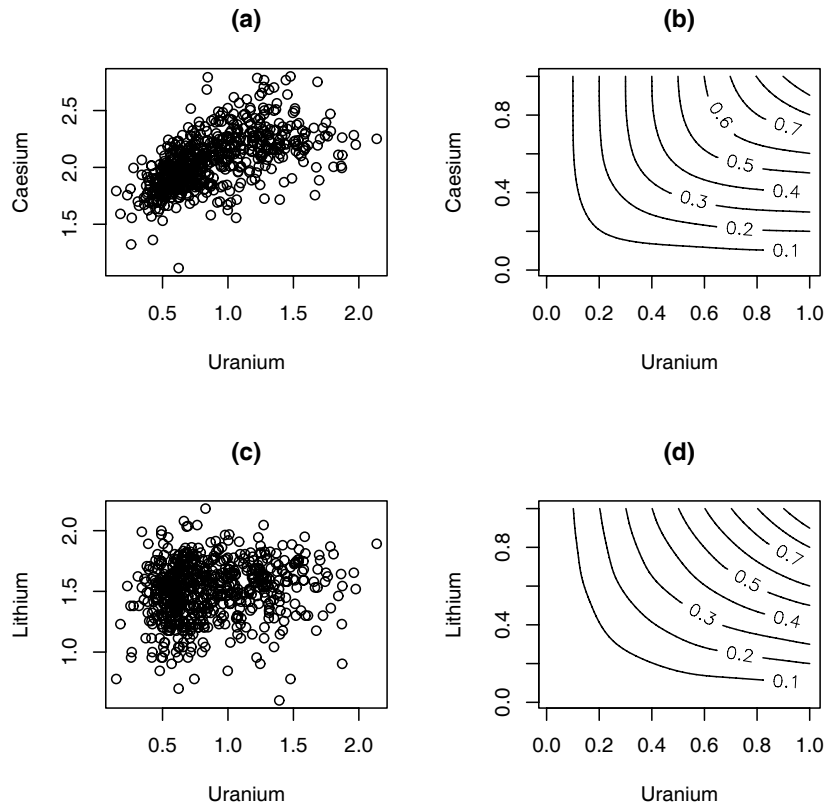


Figure 5. Scatter plots of log-concentrations of Uranium versus Caesium in (a) and of Uranium versus Lithium in (b); and the kernel copula estimators with the plug-in bandwidth (in solid lines) and the copula implied by independence (in dashed lines) of Uranium versus Caesium in (c) and of Uranium versus Lithium in (d).

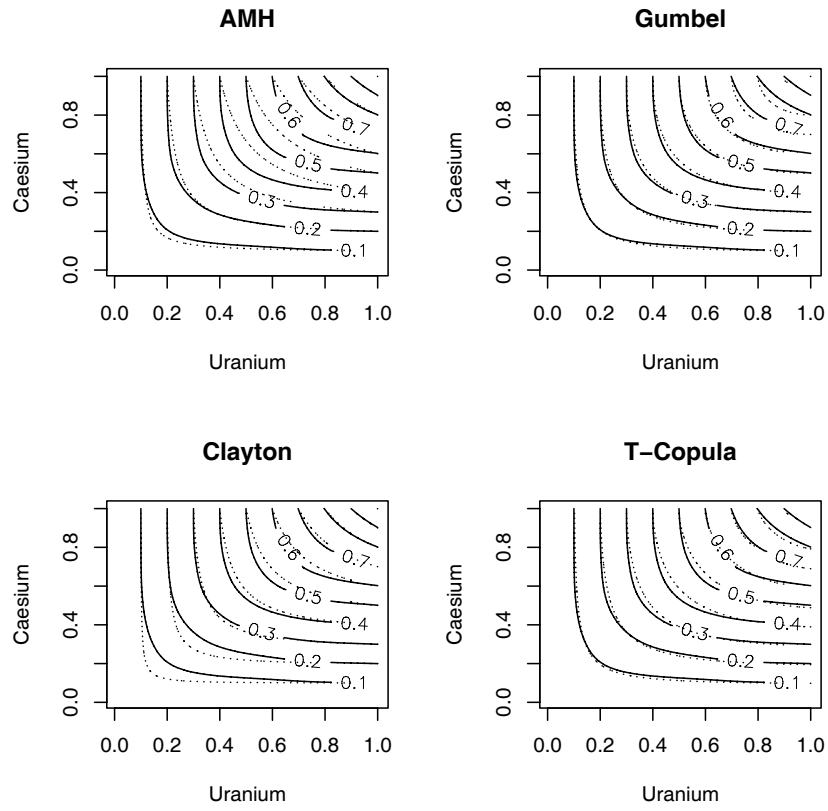


Figure 6. Copulas implied by the parametric models (in dashed lines) and the kernel estimator (in solid lines) for Uranium versus Caesium.

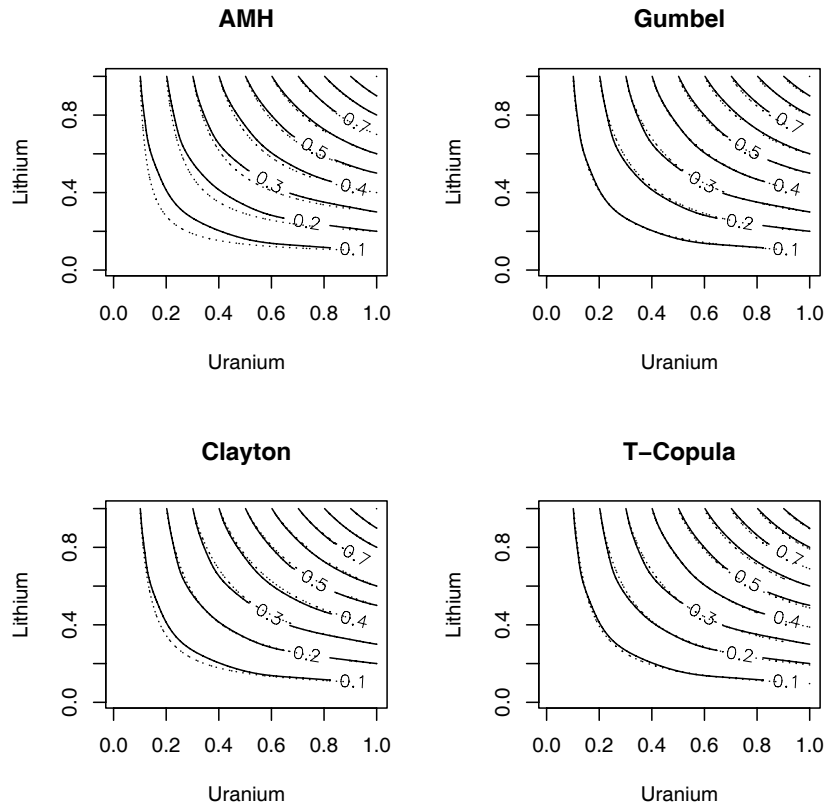


Figure 7. Copulas implied by the parametric models (in dashed lines) and the kernel estimator (in solid lines) for Uranium versus Lithium.

TABLE 1: Testing results for the Four Copula Models

(a) Uranium versus Caesium

Model	Test Statistic	5% critical value	p-value	parameter estimate
AMH	0.360	0.0254	< 0.001	$\theta = 1$
Gumbel	0.0484	0.0194	< 0.001	$\theta = 1.88$
Clayton	0.173	0.0215	< 0.001	$\theta = 1.76$
T	0.065	0.107	0.283	$\rho = .60, m = 59$

(b) Uranium versus Lithium

Model	Test Statistic	5% critical value	p-value	parameter estimate
AMH	0.1334	0.0689	< 0.001	$\theta = 0.7675$
Gumbel	0.0137	0.0210	0.221	$\theta = 1.1512$
Clayton	0.0338	0.0179	< 0.001	$\theta = 0.3024$
T	0.0212	0.0549	0.605	$\rho = 0.17, m = 59$