

Nonparametric estimation of copula functions for dependence modelling

Song Xi CHEN and Tzee-Ming HUANG

Key words and phrases: boundary bias; copula models; dependence; kernel estimator.
MSC 2000: Primary 62G05; secondary 62E20.

Abstract: 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 a joint distribution with its marginals. 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 propose a kernel estimator which is mean square consistent everywhere in the support of the copula function. The bias and variance of the copula estimator are derived which reveal the effects of kernel smoothing on the copula estimation. A smoothing bandwidth selection rule based on the derived bias and variance is proposed. The theoretical findings are confirmed by a simulation study. The kernel estimator is then used to formulate a goodness-of-fit test for parametric copula models.

Title in French: we can supply this

Résumé : Insérer votre résumé ici. 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 a joint distribution with its marginals. 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 propose a kernel estimator which is mean square consistent everywhere in the support of the copula function. The bias and variance of the copula estimator are derived which reveal the effects of kernel smoothing on the copula estimation. A smoothing bandwidth selection rule based on the derived bias and variance is proposed. The theoretical findings are confirmed by a simulation study. The kernel estimator is then used to formulate a goodness-of-fit test for 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. While these measures are simple and can be easily computed, they are designed to capture only certain aspects of dependence. Indeed, it is rather unreasonable to expect a single scalar measure to have the capability to quantify all the dependence existing among the random variables.

Copula is a device that fully quantifies the dependence among random variables. Let $X = (X_1, \dots, X_d)^\tau$ be a random vector, and F be the distribution function of X with marginal distributions F_i for $i = 1, \dots, d$. The Sklar's Theorem (Sklar 1959; Schweizer & Sklar 1983) assures the

existence of a multivariate distribution function C on $[0, 1]^d$ such that

$$F(x_1, \dots, x_d) = C\{F_1(x_1), \dots, F_d(x_d)\}.$$

The function C is called the copula associated with X and couples the joint distribution F with its marginals. See Nelsen (1998) for a 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 example, the Kendall's tau between X_1 and X_2 is $4 \int \int_{[0,1]^2} C(u, v) dC(u, v) - 1$, and Spearman's rho is $12 \int \int_{[0,1]^2} C(u, v) dudv - 3$. Copulas can also be used to describe tail dependence (Joe 1997), an important notion in risk management. The interest there is in the dependence between two extreme (risky) events.

An early statistical application of copulas can be found in Clayton (1978), where the dependence between two survival times in a multiple events study is modelled by $C(u, v) = \{u^{-1/\theta} + v^{-1/\theta} - 1\}^{1/(\theta-1)}$, the so-called Clayton copula. In later research into copulas, a driving force has been in financial risk management for modelling dependence among different assets in a portfolio; see Embrechts, Lindskog & McNeil (2003) for a comprehensive review.

Estimation of copulas can be achieved fully parametrically by assuming parametric models for both the copula and the marginals and then performing 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 Genest, Ghoudi & Rivest (1995) and Chen & Fan (2006). Estimation for Archimedean copulas in Genest & Rivest (1993) can be considered semiparametric as well. Recently, parametric copula models have been used for more general situations of inference. For instance, Fine & Jiang (2000) considered estimating the parameter of a Clayton copula with covariates in the marginal distributions in the context of life time data.

A nonparametric estimation of copula treats both the copula and the marginals parameter-free and thus 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 revealing and subsequent formulation of an underlying parametric copula model.

Nonparametric estimation of copulas dates back to Deheuvels (1979), who proposed an estimator based on a multivariate empirical distribution on the marginal empirical distributions. Smoother estimators based on the kernel method have been proposed in the literature. Gijbels & Mielnicuk (1990) proposed a kernel estimator for a bivariate copula density, which is employed by Fermanian (2005) in a goodness-of-fit test. Another approach of kernel estimation is to estimate a copula function directly as explored in Fermanian & Scaillet (2003). This approach is the focus of our investigation in this paper. The advantage of targeting directly on a copula instead of its density is that the estimation has a faster rate of convergence at $n^{-1/2}$, which is not affected by the dimension of the copula. This faster rate can lead to a less variable goodness-of-fit test statistic that is formulated based on the kernel copula estimator. Of course, whether to estimate a copula or its density depends on the particular statistical inference problem we encounter. One approach complements rather than rule out the prospects of the other.

A critical issue confronted with both kernel copula and copula density estimation is that a copula and its density are defined on a compact cube $[0, 1]^d$. This means that the boundary bias associated with kernel curve estimation will be present. Indeed, the multivariate boundaries require more cares than the standard univariate boundary region. An analysis for kernel estimation with multivariate boundary regions is given in Müller & Stadtmüller (1999). For the bivariate case, it is

necessary to ensure consistent estimation of a copula function over entire $[0, 1]^2$ especially near the corners $(0, 0)$ and $(1, 1)$. This is important in studying dependence between two extreme events, which is a target application of copula in risk management.

We propose in this paper a new bivariate kernel copula estimator based on local linear kernels and a simple mathematical correction that removes the boundary bias. We then derive the bias and variance of this estimator, which reveal that the kernel smoothing produces a second order reduction in both the variance and mean square error as compared with the unsmoothed empirical estimator of Deheuvels (1979). Our analysis also identifies which part of the kernel smoothing is the source of this variance reduction and provides a practical guideline for smoothing bandwidth selection.

The paper is organized as follows. The kernel estimator is proposed in Section 2. Its bias and variance are reported in Section 3. Section 4 considers bandwidth selection. Simulation results are reported in Section 5. Section 6 reports an empirical study. All the technical details are given in the appendix.

2. A KERNEL ESTIMATOR

The basic thrust for our 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 implied by Sklar's theorem. 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 estimate the copula based on the estimated marginals.

Let K be a symmetric probability density 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

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

with a bandwidth b_l for $l = 1$ and 2 ; see Bowman, Hall & Prvan (1998) for details on this kernel distribution 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 $h > 0$. Here $a_l(u, h) = \int_{(u-1)/h}^{u/h} t^l K(t)dt$ for $l = 0, 1$ and 2 , which was proposed by Lejeune and Sarda (1992) and Jones (1993) and was designed to remove the boundary bias in univariate density estimation. 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

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

The estimator considered by Fermanian & Scaillet (2003) is of the above form based on the original kernel K rather than the local linear kernels $K_{u,h}$.

It can be seen from the bias expression given in (2) in the next section that the use of the local linear kernel removes the boundary bias near $u = 0$ and $v = 0$. However, it still 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 each marginal distribution assumes value 1 at the right end point. Since both $T_{u,h}$ and $T_{v,h}$ are entirely known given the

kernel and h , the bias can be easily removed by subtraction. This leads to the proposed kernel copula estimator

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

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.

The above boundary bias correction is different from the one used in Gijbels & Mielniczuk (1990) for copula density estimation. The correction carried out there was rather involved via generating eight extra copies of the sample by reflecting the original data with respect to the four edges and four corners of the unit square $[0, 1]^2$ respectively, followed by applying the standard kernel density estimation technique to certain linear transformations of the estimated marginals on the extended sample.

3. MAIN RESULTS

The study of the copula estimator faces two challenges. One is that the estimator is based on the estimated marginal $\hat{F}_l(X_i)$ instead of $F_l(X_i)$. The other is that using the local linear kernels further increases the labor of derivations. However, we are able to obtain tractable expressions for the bias and variance, which add to the existing convergence results for copula density estimation (Gijbels & Mielniczuk 1990; Fermanian, Radulovic & Wegkamp 2004) and for copula estimation (Fermanian & Scaillet 2003), and provide a finer scale description of the sampling properties of the copula estimation.

The following conditions are assumed in our analysis:

A1: K is a symmetric and continuous probability density supported on $[-1, 1]$, and the 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 has a probability density function f_l such that $f_l^{(1)}$ and $f_l^{(2)}$, the first and second derivatives of f_l , are bounded and vanish at $\pm\infty$. Furthermore, each of the two limits $\lim_{x \rightarrow -\infty} f_l^{(2)}(x)/f_l^{(1)}(x)$ and $\lim_{x \rightarrow \infty} f_l^{(2)}(x)/f_l^{(1)}(x)$ is either a finite number or $\pm\infty$.

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

We note that Condition A3 does not imply that the copula density is infinitely differentiable. Rather there is an infinitely differentiable function g which is identical to f on $[0, 1]^2$. For instance, for the independent copula, g is the constant 1. However, this condition does put a strong restriction on the smoothness of copula near the boundaries, as pointed out by one referee. In particular, if the copula density is unbounded, then the condition does not hold. For instance, Condition A3 does not hold for the Gumbel copula used in our simulation study.

Let $C_u(u, v)$ and $C_v(u, v)$ be the first 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. Let $\nu(u, h) = \int_{(u-1)/h}^{u/h} s^2 dG_{u,h}(s)$, which equals $\sigma_K^2 =: \int s^2 K(s) ds$ for $u \in [h, 1-h]$. Also, for $l = 1$ and 2 , define

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

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

We have the following proposition, which quantifies the bias and variance of $\hat{C}(u, v)$.

PROPOSITION 1. *Under Conditions A1-A3, for any $(u, v) \in [0, 1]^2$,*

$$\begin{aligned}
& E\{\hat{C}(u, v)\} \\
&= C(u, v) + \frac{1}{2}h^2\{C_{uu}(u, v)\nu(u, h) + C_{vv}(u, v)\nu(v, h)\} + o(h^2) \tag{2} \\
&- \frac{1}{2}\sigma_K^2 \left[\{C_u(u, v) + T_{v,h}\}f_1^{(1)}\{F_1^{-1}(u)\}b_1^2 + \{C_v(u, v) + T_{u,h}\}f_2^{(1)}\{F_2^{-1}(v)\}b_2^2 \right]; \text{ and} \\
& \text{Var}\{\hat{C}(u, v)\} \\
&= n^{-1}\text{Var}\{I(U \leq u, V \leq v) - C_u(u, v)I(U \leq u) - C_v(u, v)I(V \leq v)\} \\
&- hn^{-1}[\{C_u(u, v) + T_{v,h}\}^2\mu_1(u, h, b_1/h) + \{C_v(u, v) + T_{u,h}\}^2\mu_2(v, h, b_2/h)] \\
&+ 2hn^{-1}[\{C_v(u, v) + T_{u,h}\}^2\mu_2^*(v, h, b_2/h) + \{C_u(u, v) + T_{v,h}\}^2\mu_1^*(u, h, b_1/h)] \\
&- hn^{-1}\{C_u(u, v)(1 + 2T_{v,h}) + T_{v,h}^2\} \int_{\frac{u-1}{h}}^{\frac{u}{h}} sdG_{u,h}^2(s) \\
&- hn^{-1}\{C_v(u, v)(1 + 2T_{u,h}) + T_{u,h}^2\} \int_{\frac{v-1}{h}}^{\frac{v}{h}} tdG_{v,h}^2(t) + o(hn^{-1}). \tag{3}
\end{aligned}$$

A sketch of proof of Proposition 1 is given in the appendix.

While the bias (2) conveys a simple story that both the first and the second stage smoothing contribute to the bias, the variance given in (3) requires a further analysis. Let us concentrate on $(u, v) \in [h, 1-h]^2$, the interior region as we use a compact kernel supported in $[-1, 1]$.

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

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

where $\rho^*(\lambda) = 2\mu^*(\lambda) - \mu(\lambda) - b_K$, $\mu^*(\lambda) = \int \int \int \max\{t, r + \lambda w\} dG(r) dG(t) dG(w)$ and

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

A key fact needed in understanding (4) is that

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

In order to achieve the largest variance reduction, we need to minimize the second term on the right hand side of (4) which involves the ρ^* -function. Due to (5), 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

$$\begin{aligned}
\text{Var}\{\hat{C}(u, v)\} &= n^{-1}\text{Var}\{I(U \leq u, V \leq v) - C_u(u, v)I(U \leq u) - C_v(u, v)I(V \leq v)\} \\
&- hn^{-1}b_K\{C_u(u, v) - C_u^2(u, v) + C_v(u, v) - C_v^2(u, v)\} + o(hn^{-1}). \tag{6}
\end{aligned}$$

This indicates a second order variance reduction by the second stage smoothing by noting that the n^{-1} order term on the right hand side is the variance of the following empirical estimator proposed

by Deheuvels (1979):

$$\tilde{C}(u, v) = n^{-1} \sum_{i=1}^n I(\hat{U}_i \leq u, \hat{V}_i \leq v) \quad (7)$$

where $\hat{U}_i = n^{-1} \sum_{j=1}^n I(X_{j1} \leq X_{i1})$ and $\hat{V}_i = n^{-1} \sum_{j=1}^n I(X_{j2} \leq X_{i2})$. Despite the variance reduction happens in the interior region only, it leads to a net reduction in the overall MISE over $[0, 1]^2$ as shown in the next section.

A drawback of $\tilde{C}(u, v)$ is its lack of continuity. This lack of continuity, as indicated in a simulation study reported in Section 5, can produce as twice the mean integrated square error (MISE) as that of the proposed kernel estimator. This shows that the variance and MISE reductions by the kernel estimator are significant in finite samples.

Finally, we note that the results in (2) and (3) apply to a kernel copula estimator that uses the marginal empirical distributions instead of the kernel distribution estimators in (1) by simply setting $b_1 = b_2 = 0$. In this case, (2) and (3) are simplified to

$$E\{\hat{C}(u, v)\} = C(u, v) + \frac{1}{2}h^2\{C_{uu}(u, v)\nu(u, h) + C_{vv}(u, v)\nu(v, h)\} + o(h^2)$$

and (6) respectively.

4. BANDWIDTH SELECTION

The findings of the previous section suggest that we should undersmooth in the first stage to reduce the variance, namely b_l should be $o(h)$ for $l = 1$ and 2 . This reduces the bias from the first stage smoothing too. This strategy largely simplifies the expressions of (2) and (4) and leads to a tractable expression for the mean square error (MSE) for $(u, v) \in [h, 1 - h]^2$

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

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 $\beta = \int_0^1 \int_0^1 \{C_{uu}(u, v) + C_{vv}(u, v)\}^2 dudv$ and

$$\alpha = b_K \int_0^1 \int_0^1 [C_u(u, v)\{1 - C_u(u, v)\} + C_v(u, v)\{1 - C_v(u, v)\}] dudv.$$

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 employed here to attain an estimate for the optimal bandwidth. A simple approach is to assume a parametric family for the copula function which then leads to parametric expressions for α and β . This is similar to the reference rule suggested by Silverman (1986) for kernel density estimation. The parameters of the parametric copula can be estimated by either the pseudo-maximum likelihood estimation or the method of moments. Then we plug-in the α and β estimates from the reference rule to (8) to obtain an estimate of the optimal bandwidth.

We propose the following T -copula as the reference copula:

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

It has two parameters: the degree of freedom ν and the correlation coefficient ρ . Here t_ν^{-1} is the marginal quantile function of the univariate T -distribution. It contains the normal copula as its limit and accommodates a wide range of tail-thickness and tail-dependence. The T -copula has been shown to be a popular parametric model in empirical finance applications as reported in Embrechts, Lindskog & McNeil (2003).

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 \hat{C} given in (7).

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

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

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. The sample sizes considered are $n = 50$ and 100 . We choose $b_1 = b_2 = b$ 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. 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 $[0, 1]^2$ based on 1000 simulations. We present in Figure 1 only the MISEs and MIVs for the Gumbel copula as those for the other two follow the same pattern.

The results conveyed by Figures 1 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. Indeed, this is shown for sample size as small as 50, which is the minimum sample size tried in the simulation. 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 as compared with the unsmoothed estimator. Indeed, for $n = 50$ and each of the copula models considered, the MISE of the kernel estimator is

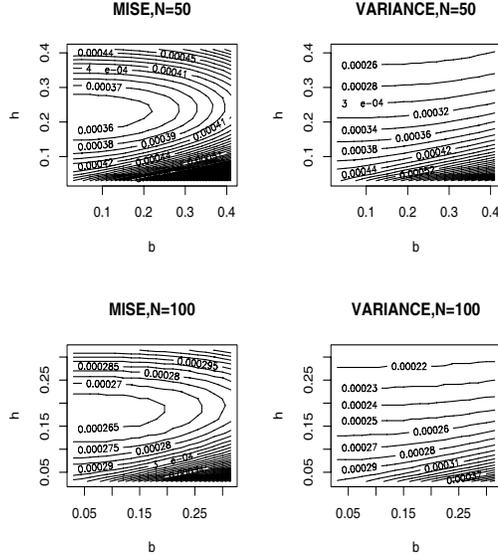


Figure 1: MISE and Integrated Variance (IVAR) of kernel estimator for Gumbel copula. The MISE and IVAR for the unsmoothed estimator (7) are 0.000649 and 0.000309 for $n = 50$; and 0.000427 and 0.000292 for $n = 100$, respectively.

nearly half of the unsmoothed estimator. Although the gap between the two estimators is reduced, 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, displayed in Figure 2, show 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 that the proposed rule is robust against mis-specifying the copula model in bandwidth selection. We also used the marginal empirical distributions instead of smoothing in the first stage, effectively setting $b_1 = b_2 = 0$. The performance of the copula estimator was almost identical to the first stage undersmoothed estimator (with $b_1 = b_2 = 10^{-4}$) in the interior of $[0, 1]^2$. However, it had some noticeable increase of bias near the upper and rights edges of the unit square, as compared to the undersmoothed estimator. This indicates some finite sample benefits by carrying out some smoothing in the first stage.

6. 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 & Johnson (1981). The same dataset has been analyzed by Genest & Rivest (1993) and Genest, Quessy & Rémillard (2006) to demonstrate a semiparametric inference for Archimedean copulas and a goodness-of-fit test. 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

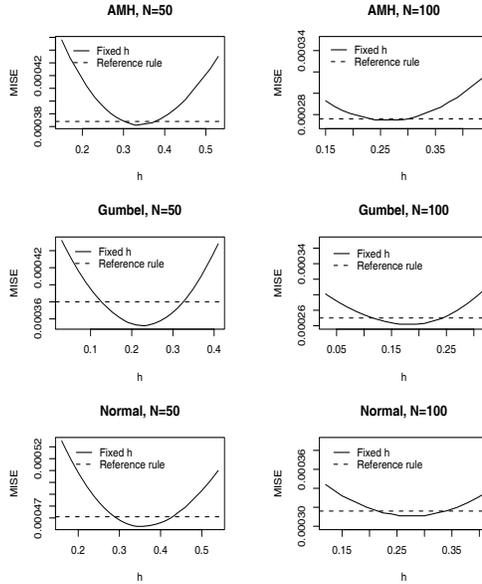


Figure 2: MISE of the kernel estimators with the plug-in bandwidth (dashed lines) and fixed bandwidths (solid lines).

between an actinide metal Uranium and two alkali metals, Caesium and Lithium.

Figure 3 displays the original data in panel (a) for Uranium versus Cesium 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, and $h = 0.143$ for Uranium versus Lithium, whereas $b_1 = b_2 = 10^{-4}$.

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 $C_\theta(u, v) = \max\{(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, 0\}$ 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 Archimedean 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 method of moment estimator proposed in Genest & Rivest (1993) which is based on the following equation regarding Kendall's tau

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

after replacing $\tau(X_1, X_2)$ by its sample version. The parameters of the T -copula are estimated by the method of moments too. Specifically, ρ is estimated by the sample correlation coefficient and $\hat{\nu} = \max\{\hat{\nu}_1, \hat{\nu}_2\}$ where $\hat{\nu}_i = (4m_{4i} + 6m_{2i}^2)(m_{4i} - 3m_{2i}^2)^{-1}$ and $m_{ki} = n^{-1} \sum_{j=1}^n (X_{ji} - \bar{X}_i)^k$ for $i = 1$ and 2 and $k = 2$ and 4 . Here $\hat{\nu}_i$ is the method of moment estimator of ν based on the i -th margin.

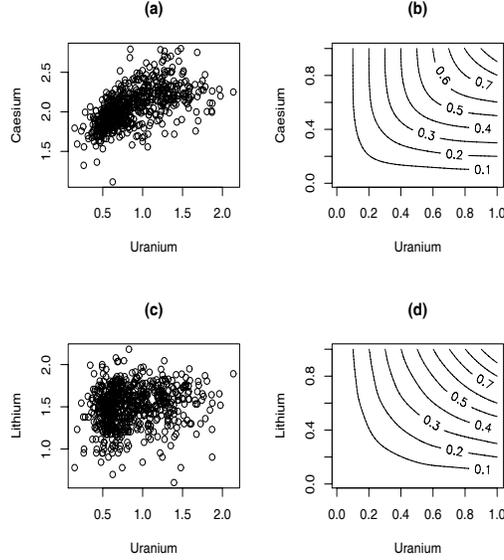


Figure 3: Log-concentrations of Uranium versus Caesium in (a) and of Uranium versus Lithium in (c); the kernel copula estimators with the plug-in bandwidth (solid lines) of Uranium versus Caesium in (b) and of Uranium versus Lithium in (d).

Let $\hat{\theta}$ be the method of moments estimator and $C_{\hat{\theta}}$ be the estimated parametric copula function. Figures 4 and 5 displays the four parametric copulas at $\hat{\theta}$ and the kernel copula estimate for Uranium versus Lithium. 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 look rather similar to each other. To check on the goodness-of-fit of a parametric copula model, a formal test procedure is needed as visual diagnostics can hardly detect the differences.

Let $C_{\hat{\theta}}$ be the estimated parametric copula model with a parameter estimate $\hat{\theta}$, and \hat{C} be the kernel estimator based on a smoothing bandwidth h . We propose the following Cramér-Von Mises type test statistic

$$\begin{aligned}
T_n &= n \int_0^1 \int_0^1 \{\hat{C}(u, v) - C_{\hat{\theta}}(u, v)\}^2 dudv \\
&= n^{-1} \int_0^1 \int_0^1 \sum_{i,j}^n \left\{ G_{u,h} \left(\frac{u - \hat{F}_1(X_{i1})}{h} \right) G_{v,h} \left(\frac{v - \hat{F}_2(X_{i2})}{h} \right) - C_{\hat{\theta}}(u, v) - B(u, v) \right\} \times \\
&\quad \times \left\{ G_{u,h} \left(\frac{u - \hat{F}_1(X_{j1})}{h} \right) G_{v,h} \left(\frac{v - \hat{F}_2(X_{j2})}{h} \right) - C_{\hat{\theta}}(u, v) - B(u, v) \right\} dudv
\end{aligned}$$

where $B(u, v) = (uT_{v,h} + vT_{u,h} + T_{u,h}T_{v,h})$. Clearly, T_n is a L_2 -distance between the kernel estimator \hat{C} and the hypothesized $C_{\hat{\theta}}$.

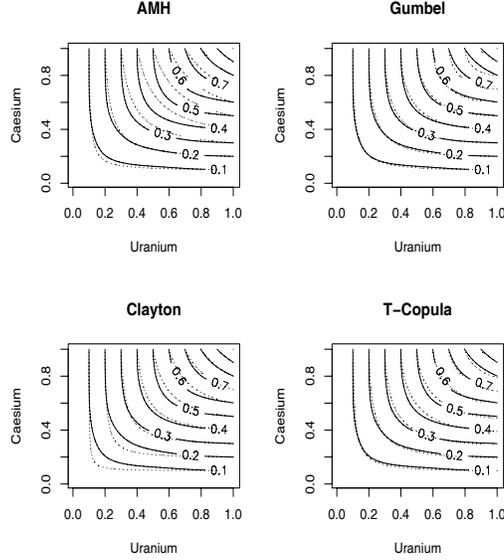


Figure 4: Copulas implied by the parametric models (dashed lines) and the kernel estimator (solid lines) for Uranium versus Caesium.

Let c_α be the upper- α quantile of the test statistic T_n at a level of significance α . The following bootstrap procedure is employed to obtain an estimate of 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 $C_{\hat{\theta}, U_i^*}(v) = \frac{\partial C_{\hat{\theta}}(u, v)}{\partial u} \Big|_{u=U_i^*}$, 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, $T_{n1}^* \leq \dots \leq T_{nB}^*$. Estimate c_α by $T_{n[B(1-\alpha)]+1}^*$, the upper α -th order statistic.

The proposed goodness-of-fit test rejects the parametric copula model as conveyed in H_0 if $T_n \geq \hat{c}_\alpha$. As the study of the kernel copula estimation has occupied large amount of space in the present paper, the theoretical properties of the proposed goodness-of-fit test will be reported in a future paper. The purpose of presenting the goodness-of-fit test is to demonstrate the usefulness of the kernel copula estimator.

We apply the above procedure to test for the four copulas for Uranium versus Caesium and Uranium versus Lithium respectively, and the results are 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 & Johnson (1981)'s original study. Gumbel copula is rejected

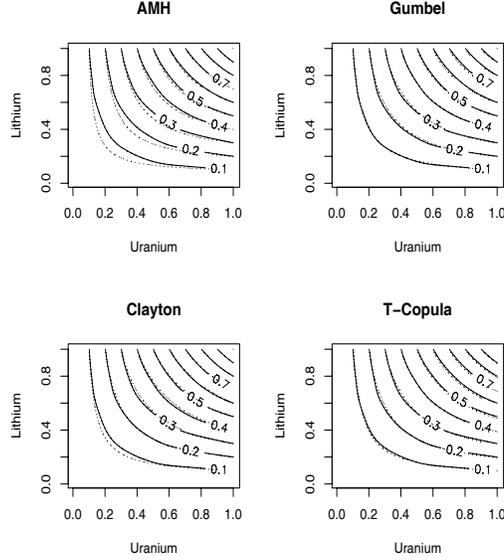


Figure 5: Copulas implied by the parametric models (dashed lines) and the kernel estimator (solid lines) for Uranium versus Lithium.

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 promising results in empirical finance (Embrechts, Lindskog & McNeil, 2003), which has been shown to be robust in fitting financial data, which typically have heavy tails and tail dependence.

APPENDIX

Proof of Proposition 1. We first outline some key steps in the proof of Proposition 1. A proof of (2) and a summary of the main results in the calculation of (3) are followed. Complete derivations can be found in a technical report (Chen and Huang, 2005).

For $i = 1, \dots, n$, let $\Delta_{1,i} = F_1(X_{i1}) - \hat{F}_1(X_{i1})$, $\Delta_{2,i} = F_2(X_{i2}) - \hat{F}_2(X_{i2})$, $I_{j,k}(s, t) = G_{u,h}^{(j)}\{(u - s)/h\}G_{v,h}^{(k)}\{(v - t)/h\}$ for $j, k \geq 0$ and $S = \{(j, k) : j \geq 0, k \geq 0, j + k \leq 5\}$. A Taylor expansion for $\hat{C}(u, v)$ is

$$\hat{C}(u, v) = \sum_{(k,j) \in S} A_{k,j} + R_n - (vT_{u,h} + uT_{v,h} + T_{u,h}T_{v,h}), \quad (12)$$

where $A_{j,k} = n^{-1} \sum_{i=1}^n \frac{1}{j!k!} I_{j,k}(F_1(X_{i1}), F_2(X_{i2})) (\Delta_{1,i}/h)^j (\Delta_{2,i}/h)^k$ and, for some $\theta \in [0, 1]$,

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

Derivations given in the technical report show that $E(R_n^2)$ is $o(h^2/n)$, which means that R_n is negligible and the derivations can be concentrated on $A_{j,k}$ terms for $(j, k) \in S$. Furthermore, it

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	$\rho = 0.17, m = 59$

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

$$\begin{aligned}
E(A_{0,0}) &= C(u, v) + \frac{1}{2}h^2\{C_{uu}(u, v)\nu(u, h) + C_{vv}(u, v)\nu(v, h)\} \\
&\quad + vT_{u,h} + uT_{v,h} + T_{u,h}T_{v,h} + o(h^2), \\
E(A_{1,0}) &= -\frac{1}{2}\sigma_K^2 b_1^2\{C_u(u, v) + T_{v,h}\}f_1^{(1)}\{F_1^{-1}(u)\} + o(h^2), \\
E(A_{0,1}) &= -\frac{1}{2}\sigma_K^2 b_2^2\{C_v(u, v) + T_{u,h}\}f_2^{(1)}\{F_2^{-1}(v)\} + o(h^2).
\end{aligned}$$

These lead to (2). To establish (3), we note that

$$Cov(A_{j,k}, A_{\tilde{j},\tilde{k}}) = o(h^4) \text{ if } j + k + \tilde{j} + \tilde{k} > 2.$$

Therefore,

$$\begin{aligned}
Var\{\hat{C}(u, v)\} &= Var(A_{0,0}) + 2Cov(A_{0,0}, A_{0,1}) + 2Cov(A_{0,0}, A_{1,0}) \\
&\quad + Var(A_{0,1}) + Var(A_{1,0}) + 2Cov(A_{0,1}, A_{1,0}) + o(h^4).
\end{aligned} \tag{13}$$

The variance expression is attained after deriving each covariance term above.

The proof of Proposition 1 relies on the Taylor expansion of $\hat{C}(u, v)$ in (12). To make the expansion legitimate, we assume in this section that

$$\sup_x |K^{(j)}(x)| < \infty \text{ for } j = 0, 1, \dots, 6. \tag{14}$$

However, (14) can be removed based on the following arguments. Write $\hat{C}_W(u, v)$ for the copula estimator based on a kernel W , and $L_{bias,W}(u, v)$ and $L_{var,W}(u, v)$ be the leading order terms of the bias and variance of $\hat{C}_W(u, v)$ respectively so that $E\{\hat{C}_W(u, v)\} = L_{bias,W}(u, v) + o(h^2)$ and $Var\{\hat{C}_W(u, v)\} = L_{var,W}(u, v) + o(h/n)$. Then, for any kernel K satisfying Condition A1, there is a symmetric bounded kernel K^* supported on $[-1, 1]$ that satisfies (14) and approximates K well enough in L_1 norm so that

$$\begin{aligned}
&\sup_{u,v,n,b_1,b_2,h} |L_{bias,K}(u, v) - L_{bias,K^*}(u, v)| + |L_{var,K}(u, v) - L_{var,K^*}(u, v)| \\
&+ |E\hat{C}_K(u, v) - E\hat{C}_{K^*}(u, v)| + |Var\{\hat{C}_K(u, v)\} - Var\{\hat{C}_{K^*}(u, v)\}| = o(n^{-2}).
\end{aligned}$$

Thus (2) and (3) hold for K if they hold true for K^* . Therefore, it is sufficient to prove (2) and (3) with the additional assumption (14).

Below we will give a detailed proof for (2) and a summary of key results in the proof for (3). We first state a lemma that gives the orders of the $\Delta_{1,i}$'s and $\Delta_{2,i}$'s.

LEMMA 1. For nonnegative integers a_1, b_1, a_2, b_2 ,

$$\sup_{x_1, y_1, x_2, y_2} \left| E \left(\Delta_{1,1}^{a_1} \Delta_{2,1}^{b_1} \Delta_{1,2}^{a_2} \Delta_{1,2}^{b_2} \middle| (X_{11}, X_{12}, X_{21}, X_{22}) = (x_1, y_1, x_2, y_2) \right) \right| = O \left(n^{-(a_1+b_1+a_2+b_2)/2} \right).$$

Proof of Lemma 1. Note that

$$\begin{aligned} & E \left(\Delta_{1,1}^{a_1} \Delta_{2,1}^{b_1} \Delta_{1,2}^{a_2} \Delta_{1,2}^{b_2} \middle| (X_{11}, X_{12}, X_{21}, X_{22}) = (x_1, y_1, x_2, y_2) \right) \\ &= E \prod_{j=1}^2 \left(A_j - n^{-1} \sum_{i=3}^n U_{j,i} \right)^{a_j} \left(B_j - n^{-1} \sum_{i=3}^n V_{j,i} \right)^{b_j}, \text{ where} \\ A_j &= F_1(x_j) - EG \left(\frac{x_j - X_{11}}{b_1} \right) + 2n^{-1} EG \left(\frac{x_j - X_{11}}{b_1} \right) - n^{-1} G(0) - n^{-1} G \left(\frac{x_j - x_{3-j}}{b_1} \right), \\ U_{j,i} &= G \left(\frac{x_j - X_{i1}}{b_1} \right) - EG \left(\frac{x_j - X_{11}}{b_1} \right), \\ B_j &= F_2(y_j) - EG \left(\frac{y_j - X_{12}}{b_2} \right) + 2n^{-1} EG \left(\frac{y_j - X_{12}}{b_2} \right) - n^{-1} G(0) - n^{-1} G \left(\frac{y_j - y_{3-j}}{b_2} \right), \text{ and} \\ V_{j,i} &= G \left(\frac{y_j - X_{i2}}{b_2} \right) - EG \left(\frac{y_j - X_{12}}{b_2} \right). \end{aligned}$$

Since $EG \left(\frac{x-X_{11}}{b_1} \right) - F_1(x) = \int F_1(x-sb_1)K(s)ds - F_1(x)$, the first derivative of f_1 is bounded, and $b_1 = O(n^{-1/3})$, we have

$$\sup_{x_1, y_1, x_2, y_2} A_j^{4a_j} = O(b_1^{8a_j}) = O(n^{-2a_j})$$

and for some positive constants C_1 and C_2 ,

$$\begin{aligned} & \sup_{x_1, y_1, x_2, y_2} E \left(A_j - n^{-1} \sum_{i=3}^n U_{j,i} \right)^{4a_j} \\ & \leq C_1 \left(\sup_{x_1, y_1, x_2, y_2} A_j^{4a_j} \right) + C_2 \left(\sup_{x_1, y_1, x_2, y_2} \left(E n^{-1} \sum_{i=3}^n U_{j,i} \right)^{4a_j} \right) = O(n^{-2a_j}). \end{aligned}$$

Similarly, $\sup_{x_1, y_1, x_2, y_2} E \left(B_j - n^{-1} \sum_{i=3}^n V_{j,i} \right)^{4b_j} = O(n^{-2b_j})$. From Cauchy Schwartz Inequality and the above calculation,

$$\begin{aligned} & \sup_{x_1, y_1, x_2, y_2} \left| E \left(\Delta_{1,1}^{a_1} \Delta_{2,1}^{b_1} \Delta_{1,2}^{a_2} \Delta_{1,2}^{b_2} \middle| (X_{11}, X_{12}, X_{21}, X_{22}) = (x_1, y_1, x_2, y_2) \right) \right| \\ & \leq \left(\prod_{j=1}^2 O(n^{-2a_j}) O(n^{-2b_j}) \right)^{1/4} = O(n^{-(a_1+b_1+a_2+b_2)/2}). \end{aligned}$$

This completes the proof of Lemma 1.

From (12), $E\{\hat{C}(u, v) - R_n\} = \sum_{(j,k) \in S} E(A_{j,k}) - vT_{u,h} - uT_{v,h} - T_{u,h}T_{v,h}$. From Lemma 1 and the fact that

$$\sup_{0 < h < 0.5} h^{-1} \int_0^1 \left| G_{u,h}^{(m)} \left(\frac{u-s}{h} \right) \right| ds < \infty \text{ for } u \in [0, 1] \text{ and for } 1 \leq m \leq 5. \quad (15)$$

$E(A_{j,k}) = o(h^2)$ for $j+k > 2$ or $(j, k) = (1, 1)$. Thus we will only compute $E(A_{j,k})$ for other cases to complete the proof of (2), which requires the following lemma.

LEMMA 2. *Suppose that G_1 and H_1 are absolutely continuous functions on $(-\infty, \infty)$ and g is a continuous function on $[0, 1]^2$. Let $C^*(u, v) = \int_0^u \int_0^v g(s, t) f(s, t) dt ds$, then for u, v in $[0, 1]$,*

(i)

$$\begin{aligned} & h^{-1} \int G_1 \left(\frac{u - F_1(x)}{h} \right) H_1' \left(\frac{v - F_2(y)}{h} \right) g(F_1(x), F_2(y)) dF(x, y) \\ &= \int_{\frac{u-1}{h}}^{\frac{u}{h}} \int_{\frac{v-1}{h}}^{\frac{v}{h}} C_v^*(u - sh, v - th) dH_1(t) dG_1(s) + \int_{-\infty}^{\frac{u-1}{h}} \int_{\frac{v-1}{h}}^{\frac{v}{h}} C_v^*(1, v - th) dH_1(t) dG_1(s) \end{aligned}$$

and

(ii) *if $\int_0^u g(s, 0) f(s, 0) ds = 0 = \int_0^u g(s, 1) f(s, 1) ds$ and H_1 and H_1' have compact supports, then $\int G_1 \left(\frac{u - F_1(x)}{h} \right) H_1' \left(\frac{v - F_2(y)}{h} \right) g(F_1(x), F_2(y)) dF(x, y) = o(h)$.*

Proof of (15) and Lemma 2. First of all, (15) can be derived by integration by parts and a fact that

$$\sup_{h \leq 0.5, 0 \leq c \leq 1} \frac{a_2(c, h) + |a_1(c, h)|}{|a_0(c, h)a_2(c, h) - a_1^2(c, h)|} < \infty. \quad (16)$$

The argument for (16) is as follows. By the symmetry of K , $a_0(c, h)$ and $a_2(c, h)$ are minimized at $c = 0, 1$ and maximized at $c = 0.5$, and $|a_1(c, h)|$ is maximized at $c = 0, 1$. Therefore,

$$\frac{a_2(c, h) + |a_1(c, h)|}{|a_0(c, h)a_2(c, h) - a_1^2(c, h)|} \leq \frac{\sigma_K^2 + \sigma_K}{a_0(1, 0.5)a_2(1, 0.5) - a_1^2(1, 0.5)}$$

and we have (16).

Lemma 2 is obtained by taking the partial derivative with respect to v of both sides of the following equation:

$$\begin{aligned} & \int G_1 \left(\frac{u - F_1(x)}{h} \right) H_1 \left(\frac{v - F_2(y)}{h} \right) g(F_1(x), F_2(y)) dF(x, y) \\ &= \int_{\frac{u-1}{h}}^{\frac{u}{h}} \int_{\frac{v-1}{h}}^{\frac{v}{h}} C^*(u - sh, v - th) dH_1(t) dG_1(s) + \int_{-\infty}^{(u-1)/h} \int_{\frac{v-1}{h}}^{\frac{v}{h}} C^*(1, v - th) dH_1(t) dG_1(s) \\ &+ \int_{\frac{u-1}{h}}^{\frac{u}{h}} \int_{-\infty}^{\frac{v-1}{h}} C^*(u - sh, 1) dH_1(t) dG_1(s) + \int_{-\infty}^{(u-1)/h} \int_{-\infty}^{\frac{v-1}{h}} C^*(1, 1) dH_1(t) dG_1(s). \end{aligned}$$

We are now ready to finish the proof of (2) by computing $E(A_{j,k})$ for $j+k \leq 2$ and $(j, k) \neq (1, 1)$. For $(j, k) = (0, 2)$,

$$2E(A_{0,2}) = h^{-2} \int I_{0,2}(F_1(x), F_2(y)) E(\Delta_{2,1}^2 | X_{11} = x, X_{12} = y) dF(x, y),$$

where

$$\begin{aligned}
& E(\Delta_{2,1}^2 | X_{11} = x, X_{12} = y) \\
&= E \left(F_2(y) - EG \left(\frac{y - X_{12}}{b_2} \right) + n^{-1} EG \left(\frac{y - X_{12}}{b_2} \right) - n^{-1} G(0) - n^{-1} \sum_{i=2}^n V_i \right)^2 \text{ and} \\
V_i &= G \left(\frac{y - X_{i2}}{b_2} \right) - EG \left(\frac{y - X_{12}}{b_2} \right).
\end{aligned}$$

Since

$$\begin{aligned}
F_2(y) - EG \left(\frac{y - X_{12}}{b_2} \right) &= F_2(y) - \int F_2(y - sb_2) K(s) ds \\
&= -\frac{\sigma_K^2 b_2^2 f_2^{(1)}(y)}{2} + o(b_2^3), \\
2E(A_{0,2}) &= o(h^4) + \frac{\sigma_K^4 b_2^4}{4h^2} \int I_{0,2}(F_1(x), F_2(y)) \left(f_2^{(1)}(y) \right)^2 dF(x, y) \\
&\quad + \frac{\sigma_K^2 b_2^2}{nh^2} \int I_{0,2}(F_1(x), F_2(y)) f_2^{(1)}(y) (G(0) - F_2(y)) dF(x, y) \\
&\quad + (nh^2)^{-1} \int I_{0,2}(F_1(x), F_2(y)) \text{Var} \left(G \left(\frac{y - X_{12}}{b_2} \right) \right) dF(x, y). \tag{17}
\end{aligned}$$

From (15) and the fact that

$$\text{Var} \left(G \left(\frac{y - X_{12}}{b_2} \right) \right) = \int F_2(y - sb_2) dG^2(s) - \left(\int F_2(y - sb_2) K(s) ds \right)^2,$$

(17) gives $2E(A_{0,2}) = o(h^2) + (nh^2)^{-1} \int I_{0,2}(F_1(x), F_2(y)) F_2(y) (1 - F_2(y)) dF(x, y)$, which is $o(h^2)$ by Lemma 2 part (ii).

Similar derivations yield $E(A_{2,0}) = o(h^2)$, and

$$\begin{aligned}
E(A_{0,0}) &= C(u, v) + \frac{1}{2} h^2 \{ C_{uu}(u, v) \nu(u, h) + C_{vv}(u, v) \nu(v, h) \} \\
&\quad + v T_{u,h} + u T_{v,h} + T_{u,h} T_{v,h} + o(h^2), \\
E(A_{1,0}) &= -\frac{1}{2} \sigma_K^2 b_1^2 \{ C_u(u, v) + T_{v,h} \} f_1^{(1)} \{ F_1^{-1}(u) \} + o(h^2), \\
E(A_{0,1}) &= -\frac{1}{2} \sigma_K^2 b_2^2 \{ C_v(u, v) + T_{u,h} \} f_2^{(1)} \{ F_2^{-1}(v) \} + o(h^2).
\end{aligned}$$

These lead to (2) since

$$E\{\hat{C}(u, v)\} = E(A_{0,0}) + E(A_{0,1}) + E(A_{1,0}) - (v T_{u,h} + u T_{v,h} + T_{u,h} T_{v,h}) + o(h^2).$$

Below we summarize the main results in the proof of (3). It can be shown that

$$\begin{aligned}
\text{Var}(A_{0,1}) &= n^{-1} v(1-v) \{ C_v(u, v) + T_{u,h} \}^2 - hn^{-1} \{ C_v(u, v) + T_{u,h} \}^2 \mu_2^*(v, h, b_2/h) + o(h^4), \\
\text{Var}(A_{1,0}) &= n^{-1} u(1-u) \{ C_u(u, v) + T_{v,h} \}^2 - hn^{-1} \{ C_u(u, v) + T_{v,h} \}^2 \mu_1^*(u, h, b_1/h) + o(h^4); \\
\text{Cov}(A_{0,0}, A_{0,1}) &= -n^{-1} \{ C_v(u, v) + T_{u,h} \} [(1-v)C(u, v) + v(1-v)T_{u,h} \\
&\quad + \{ C(u, v) - uv \} T_{v,h}] + hn^{-1} \{ C_v(u, v) + T_{u,h} \}^2 \mu_2^*(v, h, b_2/h) + o(h^4), \\
\text{Cov}(A_{0,0}, A_{1,0}) &= -n^{-1} \{ C_u(u, v) + T_{v,h} \} [(1-u)C(u, v) + u(1-u)T_{v,h} \\
&\quad + \{ C(u, v) - uv \} T_{u,h}] + hn^{-1} \{ C_u(u, v) + T_{v,h} \}^2 \mu_1^*(u, h, b_1/h) + o(h^4); \\
\text{Cov}(A_{1,0}, A_{0,1}) &= n^{-1} \{ C(u, v) - uv \} \{ C_v(u, v) + T_{u,h} \} \{ C_u(u, v) + T_{v,h} \} + o(h^4);
\end{aligned}$$

and

$$\begin{aligned}
\text{Var}(A_{0,0}) &= n^{-1}C(u, v)(1 + 2T_{u,h})(1 + 2T_{v,h}) \\
&+ n^{-1}\{T_{u,h}^2v(1 + 2T_{v,h}) + T_{v,h}^2u(1 + 2T_{u,h}) + T_{u,h}^2T_{v,h}^2\} \\
&- n^{-1}h\{C_u(u, v)(1 + 2T_{v,h}) + T_{v,h}^2\} \int_{\frac{u-1}{h}}^{\frac{u}{h}} sdG_{u,h}^2(s) \\
&- n^{-1}h\{C_v(u, v)(1 + 2T_{u,h}) + T_{u,h}^2\} \int_{\frac{v-1}{h}}^{\frac{v}{h}} tdG_{v,h}^2(t) - n^{-1}\{E(A_{0,0})\}^2 + o(h^4).
\end{aligned}$$

Substituting these into (13), (3) can be established.

ACKNOWLEDGEMENT

We thank two referees and the editor for constructive comments and suggestions that improve the presentation of the paper, and Cheng Yong Tang for valuable computation assistance and Professor Christian Genest for making the Uranium exploration data available. We also acknowledge the support of an Iowa State University Start-up grant.

REFERENCES

- A. Bowman, P. Hall & T. Prvan (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika*, 85, 799–808.
- R. D. Cook & M. E. Johnson (1981). A family of distributions for modelling nonelliptically symmetric multivariate data. *Journal of the Royal Statistical Society Series B*, 43, 210–218.
- Chen, S.X. and Huang, T. (2005), Bias and variance calculation for a copula kernel estimator, Technical Report 2005-08, Department of Statistics, Iowa State University.
- S. X. Chen & Y. Fan (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130, 307–335.
- D. G. Clayton (1978). A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incident. *Biometrika*, 65, 141–151.
- P. Deheuvels (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramérique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci.*, 65, 274–292.
- P. Embrechts, F. Lindskog, & A. McNeil (2003). Modelling dependence with copulas and applications to risk management. In *Handbook of Heavy Tailed Distributions in Finance*, ed. S. Rachev, Elsevier, 329–384.
- J. D. Fermanian (2005). Goodness-of-fit tests for copulas. *Journal of Multivariate Analysis*, 95, 119–152.
- J. Fermanian, D. Radulovic & M. Wegkamp (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10, 847–860.
- J. D. Fermanian & O. Scaillet (2003). Nonparametric estimation of copulas for time series. *Journal of Risk*, 5, 25–54.
- J. P. Fine & H. Jiang (2000). On association in a copula with time transformations. *Biometrika*, 87, 559–571.

- C. Genest, K. Ghoudi & L.-P. Rivest (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82, 543–552.
- C. Genest, J.-F. Quessy & B. Rémillard (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33, 337–366.
- C. Genest & L.-P. Rivest (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of American Statistical Association*, 88, 1034–1043.
- I. Gijbels & J. Mielniczuk (1990). Estimating the density of a copula function. *Comm. Statist. Theory Methods*, 19, 445–464.
- H. Joe (1997). *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.
- M. Lejeune & P. Sarda (1992). Smooth estimators of distribution and density functions. *Computational Statistics & Data Analysis*, 14, 457–471.
- M. C. Jones (1993). Simple boundary correction for kernel density estimation. *Statistic Computing*, 3, 135–146.
- H. G. Müller & U. Stadtmüller (1999). Multivariate boundary kernels and a continuous least squares principle. *Journal of the Royal Statistical Society Series B*, 61, 439–458.
- R. B. Nelsen (1997). *An Introduction to Copulas*, Springer, New York.
- D. Oakes (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society Series B*, 44, 414–422.
- B. Schweizer & A. Sklar (1983). *Probabilistic metric spaces*, North-Holland Series in Probability and Applied Mathematics, North-Holland Publishing, New York.
- A. Sklar (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8, 229–231.
- B. W. Silverman (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Received ???
Accepted ???

Song Xi CHEN: songchen@iastate.edu
Department of Statistics, Iowa State University
Ames, IA
USA, 50011-1210

Tzee-Ming HUANG: tmhuang@nccu.edu.tw
Department of Statistics, National Chengchi University
Taipei
Taiwan-ROC, 11605