Kernel Smoothing for Bounded Copula Density Functions

Mathias Muia[1,*], Olivia Atutey[1], Mahmud Hasan[2]

[1]*Department of Mathematics and Statistics, University of South Alabama, Mobile, AL 36688, USA*
mnmuia@southalabama.edu, oatutey@southalabama.edu
[2]*Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA*
hasanm10@vcu.edu
*Correspondence: mnmuia@southalabama.edu

ABSTRACT. Non–parametric estimation of copula density functions presents significant challenges. One issue is the unboundedness of certain copula density functions and their derivatives at the corners of the unit square. Another is the boundary bias inherent in kernel density estimation. This paper presents a kernel–based method for estimating bounded copula density functions, addressing boundary bias through the mirror reflection technique. Optimal smoothing parameters are derived via Asymptotic Mean Integrated Squared Error (AMISE) minimization and cross-validation, with theoretical guarantees of consistency and asymptotic normality. Two kernel smoothing strategies are proposed: the rule–of–thumb approach and least squares cross-validation (LSCV). Simulation studies highlight the efficacy of the rule–of–thumb method in bandwidth selection for copulas with unbounded marginal supports. The methodology is further validated through an application to the Wisconsin Breast Cancer Diagnostic Dataset (WBCDD), where LSCV is used for bandwidth selection.

1. INTRODUCTION

The study of dependence between random variables is a mainstay of statistical analysis. It is a critical, yet challenging task in multivariate statistical modeling, as it requires specifying complex joint distributions of random variables to fully capture their dependence structure. This complexity can be overcome by using copula models, which disentangle the marginal distributions from the dependence structure of the joint distribution. A multivariate distribution can be fully characterized by its marginal distributions and an associated copula, making copulas an indispensable tool for statistical investigations (Darsow *et al.* (1992) [8]; Nelsen (2006) [20]). For further details on copulas, see Joe (1997) [15], Nelsen (2006) [20], and Durante and Sempi (2015) [10]. In what follows, we focus on the bivariate case only for simplicity, the results being extendable to more than two dimensions.

Sklar's theorem (Sklar, 1959 [27]) is central to the theoretical foundation of copulas. It states that if (X, Y) is a pair of random variables with a joint distribution function H and continuous marginal distribution functions F_1 and F_2, then there exists a unique copula C such that for all x, y, $H(x, y) = C(F_1(x), F_2(y))$. The copula C fully describes the dependence structure between X and Y.

Formally, a function $C : [0, 1]^2 \rightarrow [0, 1]$ is called bivariate copula if it satisfies the following conditions:

   i. $C(0, u) = C(u, 0) = 0, \quad C(u, 1) = C(1, u) = u, \quad \forall \, u \in [0, 1]$;

   ii. $C(u_1, v_1) + C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) \geq 0, \quad \forall \, [u_1, u_2] \times [v_1, v_2] \subset [0, 1]^2$.

If C is absolutely continuous (see Nelsen (2006) [20]), then it admits a joint density given by:

$$c(u, v) = \frac{\partial^2 C}{\partial u \partial v} = \frac{\partial^2 C}{\partial v \partial u}.$$

In this case, such a copula is said to possess a density. The primary objective of this paper is to propose an estimator for c, given an i.i.d. sample $\{(X_i, Y_i)\}_{i=1}^n$ from the distribution function H, and analyze its properties.

Methods for estimating copulas and copula densities generally depend on the assumptions made about the joint distribution function H. Commonly used approaches include parametric methods (Iyengar *et al.*, 2011 [14]), semiparametric methods (Chen and Fan, 2006 [5]), and nonparametric methods (Chen and Huang, 2007 [6]; Wang *et al.*, 2012 [32]). In fully parametric models, where both the copula and the marginal distributions are explicitly specified, maximum likelihood estimation is the method of choice. Semiparametric approaches, on the other hand, often specify a parametric copula while estimating the marginals nonparametrically, providing greater flexibility.

Striking a balance between accuracy and computational efficiency is essential in copula estimation. Trivedi and Zimmer (2005) [30] and Choros *et al.* (2010) [7] provided a discussion of these trade–offs. However, parametric methods can suffer significant underestimation when the marginal distributions are unknown or misspecified, as noted by Charpentier *et al.* (2007) [4]. In such cases, non–parametric methods emerge as a more robust and flexible alternative. For example, Behnen *et al.* (1985) [1] proposed using a rank–based estimator, which modifies kernel estimators for application to rank data, to estimate the copula density. Later, Gijbels and Mielniczuk (1990) [11] introduced a kernel–type estimator for bivariate copula densities and demonstrated its consistency and asymptotic normality under various conditions related to bandwidth selection and kernel smoothness.

These advancements underscore the effectiveness of nonparametric methods in capturing complex dependence structures without imposing restrictive assumptions on the copula or its marginals (Charpentier *et al.* (2007) [4]; Chen and Huang, 2007 [6]). Building on these techniques, this paper employs a two–stage kernel density estimation procedure to estimate the copula density. In the first stage, the marginal distribution functions $F_1(X)$ and $F_2(Y)$ are estimated nonparametrically using

empirical CDFs. These marginal estimates are then utilized in the second stage to construct a flexible, non-parametric estimate of the copula density. This approach provides a robust framework for modeling dependence structures, particularly when the underlying distribution functions are complex or unknown.

To formalize the methods, we introduce several notations that will be used throughout the paper. The squared L_2 norm of a function g is defined as

$$R(g) \equiv \|g\|_2^2 = \int_{-\infty}^{\infty} g(x)^2 \, dx, \tag{1.1}$$

where $R(g)$ can be interpreted as a measure of the roughness of g. The square of the pth derivative of g is denoted by $g^{(p)}(x)^2$. Moments of the kernel K are represented by μ_k, where

$$\mu_k = \int s^k K(s) \, ds. \tag{1.2}$$

A symmetric kernel function is characterized by the property $k(u) = k(-u)$ for all u. In this case, all odd moments are equal to zero. Most nonparametric estimation techniques utilize symmetric kernels, and our discussion will focus on them. For a kernel K, we assume that $\mu_0 = 1$, $\mu_1 = \cdots = \mu_{p-1} = 0$, and $0 < |\mu_p| < \infty$ for some even p. These moment conditions play a pivotal role in determining the theoretical properties of kernel-based estimators. Unless otherwise stated, integrals are taken over the entire real line.

The remainder of the paper is organized as follows. Section 2 introduces the proposed estimator. In Section 3, we examine the properties of the estimator, including its bias and asymptotic behavior. Section 4 discusses bandwidth selection procedures, detailing two main methods: the rule-of-thumb approach and least squares cross-validation (LSCV). Section 5 presents a simulation study using data generated from the Frank copula to validate our methodology. Finally, the results are applied to the Wisconsin Breast Cancer dataset in Section 6.

2. Estimating the Copula Density

Let $\{(X_i, Y_i)\}_{i=1}^n$ be an i.i.d. sample from the joint distribution function H. To estimate the copula density, we consider the transformed sample $\{(\hat{F}_1(X_i), \hat{F}_2(Y_i))\}_{i=1}^n$, where $\hat{F}_1$ and $\hat{F}_2$ are the empirical distribution functions of the marginal distributions. Using empirical estimators for the margins has distinct advantages and trade-offs. On the one hand, this approach introduces zero bias into the copula density estimator and provides an exact representation of the data, unlike kernel-based margin estimators as discussed in Chen and Huang (2007) [6]. On the other hand, it results in higher variance, which may affect the overall accuracy of the estimator in the case of small samples.

Based on the transformed sample, the natural kernel-type estimator of $c(u, v)$ is (Wand and Jones (1995) [31]; Silverman (2018) [26])

$$\frac{1}{nh_n^2} \sum_{i=1}^{n} K\left(\frac{u - \hat{F}_1(X_i)}{h_n}, \frac{v - \hat{F}_2(Y_i)}{h_n}\right), \tag{2.1}$$

where K is a bivariate density function that is symmetric, unimodal at $(0, 0)$, and has support on $[-1, 1]$. The parameter h_n represents a sequence of bandwidths that converges to zero as the sample size increases. Commonly used kernel functions include the uniform, Epanechnikov, biweight, triweight, and Gaussian kernels, each of which affects the smoothness and accuracy of the estimate differently. Detailed discussions on kernel density estimation can be found in Wand and Jones (1995) [31], Bowman *et al.* (1998) [2], and Silverman (2018) [26].

However, the estimator (2.1) has a notable limitation. It is inconsistent at points on the boundary of the unit square in which c has a jump, as noted by Gijbels and Mielniczuk (1990) [11]. This arises because, near the boundary, the summands in (2.1) allocate a significant amount of mass outside the unit square, leading to an issue known as boundary bias. Three methods have been proposed to address this problem, all of which were initially developed in the context of univariate kernel density estimation on the unit line. The first approach employs mirror reflection techniques, as discussed by Deheuvels and Hominal (1979) [9], Schuster (1985) [25], and Behnen *et al.* (1985) [1]. In two dimensions, this method reflects data points across all edges and corners of the unit square, generating an expanded dataset from which the kernel estimate is constructed. By redistributing the kernel's mass within the unit square, this approach mitigates boundary bias. The second method uses a local linear variant of the kernel estimator, as introduced by Chen and Huang (2007) [6], to address the bias near boundaries. The third method involves employing kernels whose support aligns with the support of the target density. These kernels adjust their shape based on the location where the density is being estimated. They are known as *boundary kernels* and are exemplified by the *beta kernel*, proposed by Charpentier *et al.* (2007) [4].

In this work, we employ the mirror reflection technique, which yields the following estimate of $c(u, v)$. Denoting $\hat{F}_1(X_i)$ and $\hat{F}_2(Y_i)$ by $\hat{U}_i$ and $\hat{V}_i$, respectively, we define the estimator for $c(u, v)$ as follows (Behnen *et al.* (1985) [1]):

Definition 2.1. *Let $c(u, v)$ be a copula density. The mirror-reflection kernel estimate of c with smoothing parameter $h_n > 0$ is defined by*

$$\hat{c}(u, v) = \begin{cases} \frac{1}{nh_n^2} \sum_{i=1}^{n} \sum_{l=1}^{9} K\left(\frac{u - \hat{U}_{il}}{h_n}, \frac{v - \hat{V}_{il}}{h_n}\right), & \text{for } 0 \le u, v \le 1, \\ 0, & \text{otherwise}, \end{cases} \tag{2.2}$$

where $\{(\hat{U}_{il}, \hat{V}_{il}), i = 1, \ldots, n; l = 1, \ldots, 9\} = \{(\pm\hat{U}_i, \pm\hat{V}_i), (\pm\hat{U}_i, 2 - \hat{V}_i), (2 - \hat{U}_i, \pm\hat{V}_i), (2 - \hat{U}_i, 2 - \hat{V}_i), i = 1, \ldots, n\}$.

The following conditions are assumed in the analysis

Assumptions 2.2. A_1: *H has continuous marginal distribution functions and the copula density*
        c has bounded derivatives up to the fourth order on $[0, 1]^2$;
    A_2: *K is a symmetric, bounded, sufficiently smooth continuous probability density supported on*
        $[-1, 1]$;
    A_3: *The bandwidths satisfy $h_n \longrightarrow 0$ as $n \longrightarrow \infty$ and $nh_n^2 \longrightarrow \infty$ as $n \longrightarrow \infty$.*

Some widely applied copula densities, such as the Clayton, Gumbel, Gaussian, and Student's t copulas, as well as their first- and second-order partial derivatives, exhibit unboundedness at the boundaries of the unit square. This behavior introduces complexities in the analysis of these derivatives near the corners and edges of the unit square. To address this challenge in their work on copula estimation, Omelka *et al.* (2009) [21] used the "bandwidth shrinking" method when approaching the borders of $[0, 1]^2$. This approach involves substituting the bandwidth h_n with a bandwidth function $r(w)h_n$, where $w = u$ or v, effectively "shrinking" the bandwidth towards zero at the corners of the unit square.

In this work, we focus on the estimation of copula densities that are bounded and possess bounded second-order partial derivatives, exemplified by families such as Ali-Mikhail-Haq, Frank, Plackett, Farlie-Gumbel-Morgenstern, among others.

3. Properties of the Estimator

In this section, we analyze the properties of the estimator defined in (2.2). For this purpose, we utilize the multiplicative kernel, which is specified as follows:

$$\hat{c}(u, v) = \frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{9} K_{h_n}\left(u - \hat{U}_{il}\right) K_{h_n}\left(v - \hat{V}_{il}\right), \tag{3.1}$$

where $K_{h_n}(x) = 1/h_n K(x/h_n)$. This kernel function will be employed consistently throughout the remainder of the paper.

3.1. **Bias and variance.** The following proposition establishes the expressions for the estimator's bias and variance.

Proposition 3.1. *Let $c(u, v)$ be a copula density function that is twice-continuously differentiable on $[0, 1]^2$. Assume that the smoothing parameter $h_n \to 0$ and $nh_n^2 \to \infty$ as $n \to \infty$. Then, for all points $(u, v) \in [0, 1]^2$, the bias and variance of the estimator $\hat{c}(u, v)$, defined in (3.1), are given by:*

$$\text{Bias}[\hat{c}(u, v)] = \frac{\mu_2(K)h_n^2}{2}\left[c_{uu}(u, v) + c_{vv}(u, v)\right] + o(h_n^2),$$

$$\text{Var}[\hat{c}(u, v)] = \frac{R(K)^2}{nh_n^2}c(u, v) + o\left(\frac{1}{nh_n^2}\right),$$

where $R(K)$ and $\mu_2(K)$ are as defined in (1.1) and (1.2) respectively. Here, c_{uu} and c_{vv} denote the second-order partial derivatives of c with respect to u and v, respectively.

Proof.   i. The bias is given by

$$Bias[\hat{c}(u,v)] = \sum_{l=1}^{9} E\left[K_{h_n}(u - \hat{U}_{1l})K_{h_n}(v - \hat{V}_{1l})\right] - c(u,v) \tag{3.2}$$

To show that this bias is uniformly $O(h_n^2)$, we utilize the Taylor expansion of $c(u - h_n s, v - h_n t)$. This expansion is justified by the assumption that $c(u,v)$ has bounded first and second-order derivatives on $[0,1]^2$.

The simplest case occurs when $(u,v) \in [h_n, 1 - h_n]^2$. In this case, Equation (3.2) reduces to

$$Bias[\hat{c}(u,v)] = E\left[K_{h_n}(u - \hat{U}_1)K_{h_n}(v - \hat{V}_1)\right] - c(u,v).$$

Substituting $s = (u - x)/h_n$ and $t = (v - y)/h_n$ and using Assumptions 2.2, we obtain

$$Bias[\hat{c}(u,v)] = \int_{-1}^{1}\int_{-1}^{1} c(u - h_n s, v - h_n t)K(s)K(t)dsdt - c(u,v)$$

$$= \int_{-1}^{1}\int_{-1}^{1}\{c(u,v) - h_n s c_u(u,v) - h_n t c_v(u,v) + \frac{h_n^2 s^2}{2}c_{uu}(u,v) +$$

$$h_n^2 st c_{uv}(u,v) + \frac{h_n^2 t^2}{2}c_{vv}(u,v)\}K(s)K(t)dsdt + o(h_n^2) - c(u,v)$$

$$= \frac{\mu_2(K)h_n^2}{2}[c_{uu}(u,v) + c_{vv}(u,v)] + o(h_n^2)$$

where $\mu_2(K)$ denotes the second moment of K, and $c_u, c_v, c_{uu}, c_{uv},$ and c_{vv} are the first- and second-order partial derivatives with respect to u and v.

Regarding the remaining cases, we will show only for the case when $(u,v) \in [1 - h_n, 1]^2$. The other cases may be handled in a similar way. Note that the Taylor expansion together with Assumptions 2.2 imply that $c(u,v) = c(1,1) + (u - sh_n)c_u(1,1) + (v - th_n)c_v(1,1) + O(h_n^2)$ uniformly in $(u,v) \in [1 - 2h_n, 1]^2$. Further routine algebra shows that (3.2) simplifies to

$$Bias[\hat{c}(u,v)] = EK_{h_n}(u - U_1)K_{h_n}(v - V_1)$$

$$+ EK_{h_n}(u - U_1)K_{h_n}(v + V_1 - 2)$$

$$+ EK_{h_n}(u + U_1 - 2)K_{h_n}(v - V_1)$$

$$+ EK_{h_n}(u + U_1 - 2)K_{h_n}(v + V_1 - 2) - c(u,v)$$

Let $G(z) = \int_{-\infty}^{z} K(t)dt$ and $T_{w,h_n} = G((w - 1)/h)$ for $w = u$ or v. We compute

$$EK_{h_n}(u - U_1)K_{h_n}(v - V_1)$$

$$= \int\int K_{h_n}(u - x)K_{h_n}(v - y)c(x,y)dxdy$$

$$= \int_{\frac{v-1}{h_n}}^{1} \int_{\frac{u-1}{h_n}}^{1} c(u - sh_n, v - th_n)K(s)K(t)dsdt$$

$$= \int_{\frac{v-1}{h_n}}^{1} \int_{\frac{u-1}{h_n}}^{1} \{c(1,1) + uc_u(1,1) - sh_nc_u(1,1) - c_u(1,1) + vc_v(1,1) - th_nc_v(1,1) - c_v(1,1)\}$$

$$K(s)K(t)dsdt + O(h_n^2)$$

$$= c(1,1)[1 - T_{v,h_n}][1 - T_{u,h_n}] + uc_u(1,1)[1 - T_{v,h_n}][1 - T_{u,h_n}] - c_u(1,1)[1 - T_{v,h_n}] \times$$

$$[1 - T_{u,h_n}] - h_nc_u(1,1)[1 - T_{v,h_n}]\int_{\frac{u-1}{h_n}}^{1} sK(s)ds + vc_v(1,1)[1 - T_{v,h_n}][1 - T_{u,h_n}] -$$

$$c_v(1,1)[1 - T_{v,h_n}][1 - T_{u,h_n}] - h_nc_v(1,1)[1 - T_{u,h_n}]\int_{\frac{v-1}{h_n}}^{1} tK(t)dt + O(h_n^2). \tag{3.3}$$

Similarly,

$$EK_{h_n}(u - U_1)K_{h_n}(v + V_1 - 2) = \int\int K_{h_n}(u - x)K_{h_n}(v + y - 2)c(x,y)dxdy$$

$$= \int_{-1}^{\frac{v-1}{h_n}} \int_{\frac{u-1}{h_n}}^{1} c(u - sh_n, 2 + th_n - v)K(s)K(t)dsdt$$

$$= c(1,1)[1 - T_{u,h_n}]T_{v,h_n} + uc_u(1,1)[1 - T_{u,h_n}]T_{v,h_n} - h_nc_u(1,1)\int_{\frac{u-1}{h_n}}^{1} sK(s)dsT_{v,h_n} -$$

$$c_u(1,1)[1 - T_{u,h_n}]T_{v,h_n} + c_v(1,1)[1 - T_{u,h_n}]T_{v,h_n} + h_nc_v(1,1)\int_{-1}^{\frac{v-1}{h_n}} tK(t)dt[1 - T_{u,h_n}] -$$

$$vc_v(1,1)[1 - T_{u,h_n}]T_{v,h_n} + O(h_n^2). \tag{3.4}$$

Taking a similar approach as above gives

$$EK_{h_n}(u + U_1 - 2)K_{h_n}(v - V_1) = \int\int K_{h_n}(u + x - 2)K_{h_n}(v - y)c(x,y)dxdy$$

$$= c(1,1)[1 - T_{v,h_n}]T_{u,h_n} - uc_u(1,1)[1 - T_{v,h_n}]T_{u,h_n} + h_nc_u(1,1)\int_{-1}^{\frac{u-1}{h_n}} sK(s)ds[1 - T_{v,h_n}] +$$

$$c_u(1,1)[1 - T_{v,h_n}]T_{u,h_n} - c_v(1,1)[1 - T_{v,h_n}]T_{u,h_n} - h_nc_v(1,1)\int_{\frac{v-1}{h_n}}^{1} tK(t)dtT_{u,h_n} +$$

$$vc_v(1,1)[1 - T_{v,h_n}]T_{u,h_n} + O(h_n^2). \tag{3.5}$$

and

$$EK_{h_n}(u + U_1 - 2)K_{h_n}(v + V_1 - 2) = \int\int K_{h_n}(u + x - 2)K_{h_n}(v + y - 2)c(x,y)dxdy$$

$$= c(1,1)T_{v,h_n}T_{u,h_n} + c_u(1,1)T_{v,h_n}T_{u,h_n} + h_nc_u(1,1)\int_{-1}^{\frac{u-1}{h_n}} sK(s)dsT_{v,h_n} - uc_u(1,1)T_{v,h_n}T_{u,h_n} +$$

$$c_v(1,1)T_{v,h_n}T_{u,h_n} + h_nc_v(1,1) + \int_{-1}^{\frac{v-1}{h_n}} tK(t)dtT_{u,h_n} - vc_v(1,1)T_{v,h_n}T_{u,h_n} + O(h_n^2) \tag{3.6}$$

Combining (3.3) with (3.4), (3.5) and (3.6) gives us

$$Bias[\hat{c}(u,v)] = c(1,1) + (u - sh_n)c_u(1,1) + (v - th_n)c_v(1,1) + O(h_n^2) - c(u,v) = O(h_n^2),$$

which was to be proved.

ii. The prove for the variance relies on results on the asymptotic normality for the mirror–reflection estimator by Gijbels and Mielniczuk (1990) [11], Theorem 3.2.

□

3.2. **Consistency and asymptotic normality of the estimator.** Behnen *et al.* (1985) [1], in Theorem 2.1, showed that the estimator (2.2) is a uniformly strongly consistent estimator of $c(u, v)$. The asymptotic normality of the estimator is established in Theorem 3.2 of Gijbels and Mielniczuk (1990) [11].

4. Bandwidth Selection

The estimation process is highly sensitive to the choice of smoothing parameters, h_n. In this section, we discuss three main methods for selecting bandwidth. Determining the "optimal" values for these parameters requires a specific criterion function.

The first method identifies the optimal bandwidth by minimizing the AMISE. The other two methods involve cross–validation (CV) approaches aimed at minimizing the MISE, following the works of Bowman (1984) [3], Scott and Terrell (1987) [24], and Terrell and Scott (1985) [29]. In the context of multivariate kernel estimation, these data–driven CV methods seek to estimate the Integrated Squared Error (ISE) or MISE directly from the data, identifying the smoothing parameter that minimizes the estimated ISE or MISE.

Stone (1984) [28] demonstrated a significant theoretical finding that, when the underlying multivariate density and its one–dimensional marginals are bounded, the smoothing parameters selected using multivariate least-squares cross-validation (LSCV) are asymptotically optimal. For more information on cross–validation methods, refer to Rudemo (1982) [22], Bowman (1984) [3], Scott and Terrell (1987) [24], and Terrell and Scott (1985) [29]. Higher–order plug–in algorithms for multivariate density estimation have been explored by Wand and Jones (1995) [31].

4.1. **Rule–of–thumb approach.** A common method for selecting bandwidth involves establishing a rule–of–thumb based on a specific parametric family. Given that our interest lies in estimating density over the entire unit square, our focus is on choosing a global bandwidth. Consequently, we adopt the strategy of minimizing the asymptotic mean integrated squared error (AMISE).

The mean square error (MSE) for $(u, v) \in [h_n, 1 - h_n]^2$ is

$$MSE(\hat{c}(u, v)) = Var(\hat{c}(u, v)) + Bias(\hat{c}(u, v))^2$$

$$= \frac{R(K)^2}{nh_n^2}c(u, v) + [\frac{\mu_2(K)h_n^2}{2}[c_{uu}(u, v) + c_{vv}(u, v)]]^2 + o((nh_n^2)^{-1} + h^4) \quad (4.1)$$

The MISE is obtained by integrating (4.1) above using the integrability assumption on $c(u, v)$.

$$MISE(\hat{c}(u, v))$$

$$=(nh_n^2)^{-1}R(K)^2 \int_0^1 \int_0^1 c(u,v)\ du\ dv + \frac{h_n^4 \mu_2(K)^2}{4} \int_0^1 \int_0^1 [c_{uu}(u,v) + c_{vv}(u,v)]^2\ du\ dv$$

$$+\ o((nh_n^2)^{-1} + h^4)$$

$$=(nh_n^2)^{-1}R(K)^2 + \frac{h_n^4 \mu_2(K)^2}{4}\beta + o((nh_n^2)^{-1} + h^4), \tag{4.2}$$

where $\beta = \int_0^1 \int_0^1 [c_{uu}(u,v) + c_{vv}(u,v)]^2\ du\ dv$. This leads to the asymptotic MISE

$$\text{AMISE}(h_n) = (nh_n^2)^{-1}R(K)^2 + \frac{h_n^4 \mu_2(K)^2}{4}\beta, \tag{4.3}$$

with β as defined above. The optimal bandwidth that minimizes the MISE is therefore given by

$$h_n^* = \left[\frac{2R(K)^2}{n\mu_2(K)^2\beta}\right]^{1/6}. \tag{4.4}$$

This expression still relies on the unknown copula density $c(u,v)$. In practice, we select a parametric reference copula family and adjust the parameter to match the dependence strength observed in the data, for instance, by inverting Kendall's tau to estimate the unknown parameter in Frank's copula density. The optimal bandwidth h_n^* is then determined numerically. It is important to note that this approach is feasible only if $R(K)^2$ and β are finite.

4.2. **Cross-validation (CV).** We now shift our focus to an alternative approach for estimating bandwidth. Notably, h_n^*, defined in (4.4), is obtained by minimizing the AMISE, where estimates replace the unknown curvature components of c. Alternatively, we may directly minimize the MISE using cross-validation approaches, as recommended by Scott and Terrell (1987) [24]. The CV approach involves using the sample twice: once to compute the KDE and again to assess its accuracy in estimating $c(u,v)$. To avoid dependence on the same data for both computation and evaluation, the CV approach partitions the sample in a cross-validatory manner, ensuring that the data used to compute the KDE is excluded from its evaluation.

4.2.1. *Least squares cross-validation (LSCV).* Least squares cross-validation (LSCV) is a statistical technique for evaluating a model's predictive accuracy and optimizing model parameters by dividing data into training and validation subsets. This method is particularly useful in selecting parameters that prevent overfitting, ensuring the model generalizes well to new data. Rudemo (1982) [22] discussed LSCV for selecting smoothing parameters in density estimation, while Bowman (1984) [3] introduced foundational approaches to LSCV for bandwidth selection. Wand and Jones (1995) [31] provided an overview of LSCV for kernel smoothing, covering both theoretical and practical applications.

Second-stage smoothing

In the second stage of the estimation process, we employ the estimated marginal distributions to estimate the copula density $c(u,v)$, where $u = F(x)$ and $v = G(y)$. Here, the Epanechnikov kernel,

$K(x) = 3/4(1 - x^2)_+$, where the subscript "+" denotes the positive part. Our goodness-of-fit criterion between $c(u, v)$ and $\hat{c}(u, v)$ is the usual ISE, defined as follows:

$$\text{ISE} = \int_0^1 \int_0^1 [\hat{c}(u, v) - c(u, v)]^2 \; du \, dv. \tag{4.5}$$

The ISE measures the discrepancy between the true copula density and the estimated density over the unit square. By taking the expectation of the ISE, we obtain the MISE, which represents the average error in our estimation over repeated samples (see Rudemo (1982) [22] and Bowman (1984) [3]). The MISE serves as a criterion for selecting an optimal bandwidth by minimizing the estimation error. Replacing $\hat{c}(u, v)$ with the generalized estimator $\hat{c}(u, v; h_n)$ in (4.5) and expanding yields

$$ISE(\hat{c}(\cdot; h_n)) = R(\hat{c}(\cdot; h_n)) - 2 \int_0^1 \int_0^1 \hat{c}(u, v; h_n) c(u, v) \, du \, dv + \int_0^1 \int_0^1 c(u, v)^2 \, du \, dv,$$

where

$$\hat{c}(u, v; h_n) = \frac{1}{n} \sum_{i=1}^n \sum_{l=1}^9 K_{h_n} \left(u - \hat{U}_{il} \right) K_{h_n} \left(v - \hat{V}_{il} \right)$$

is the multiplicative bivariate kernel estimator. Here, $\hat{U}_{il}$ and $\hat{V}_{il}$ are as given in Definition 2.1, and $R(g)$ is as defined in (1.1). The MISE is then given by

$$MISE(\hat{c}(\cdot; h_n)) = E\left[R(\hat{c}(\cdot; h_n))\right] - 2E\left[\int_0^1 \int_0^1 \hat{c}(u, v; h_n) c(u, v) \, du \, dv\right] + E\left[\int_0^1 \int_0^1 c(u, v)^2 \, du \, dv\right].$$

Since the last term is independent of h_n, minimizing $MISE(\hat{c}(\cdot; h_n))$ is equivalent to minimizing

$$E\left[R(\hat{c}(\cdot; h_n))\right] - 2E\left[\int_0^1 \int_0^1 \hat{c}(u, v; h_n) c(u, v) \, du \, dv\right]. \tag{4.6}$$

This quantity is unknown but can be estimated unbiasedly as (Bowman (1984) [3] and Wand and Jones (1995) [31]):

$$LSCV(h_n) := R(\hat{c}(\cdot; h_n)) - \frac{2}{n} \sum_{i=1}^n \sum_{l=1}^9 \hat{c}_{-i}(U_{il}, V_{il}; h_n), \tag{4.7}$$

where

$$\hat{c}_{-i}(u, v; h_n) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{l=1}^9 K_{h_n}(u - U_{jl}) K_{h_n}(v - V_{jl})$$

is the leave-one-out KDE, computed by excluding (U_{il}, V_{il}) from the sample.

The first term in (4.7) is unbiased by design. The second term results from estimating $\int\int \hat{c}(u, v) c(u, v) \, du \, dv$ using a Monte Carlo approximation based on the sample $\{U_{il}, V_{il} \mid i = 1, \dots, n; \, l = 1, \dots, 9\}$, whose copula has density $c(u, v)$; specifically, this is achieved by replacing $c(u, v) \; du \, dv = dC(u, v)$ with $dC_n(u, v)$, where C_n denotes the empirical copula. This can be expressed as

$$\int \int \hat{c}(u, v; h_n) \, du \, dv \approx \frac{1}{n} \sum_{i=1}^n \hat{c}(U_{il}, V_{il}; h_n),$$

where, to reduce sample dependence, we substitute $\hat{c}(U_{il}, V_{il}; h_n)$ with $\hat{c}_{-i}(U_{il}, V_{il}; h_n)$. In this manner, we utilize the sample to estimate the integral involving $\hat{c}(\cdot; h_n)$, while for each (U_{il}, V_{il}), the kernel density estimate is computed from the remaining observations.

To demonstrate that (4.7) serves as an unbiased estimator of (4.6), observe that the expectations of (4.6) and $LSCV(h_n)$ correspond term by term, since

$$E\hat{c}_{-i}(u_i, v_i; h_n) = EK_{h_n}(u_i - U_j)K_{h_n}(v_i - V_j) = E\int_0^1 \int_0^1 K_{h_n}(u - U_j)K_{h_n}(v - V_j)c(u, v) \, du \, dv$$

$$= E\hat{c}(u, v; h_n) \, du \, dv.$$

We refer to (4.7) as an unbiased cross-validation criterion because its expectation satisfies

$$E(\text{LSCV}(h_n)) = \text{MISE}(\hat{c}(\cdot; h_n)) - \int_0^1 \int_0^1 c(u, v)^2 \, du \, dv.$$

This expression shows that the expected value of LSCV directly approximates the MISE, corrected by the integral term. In contrast, other theoretical expressions, such as AMISE, are only asymptotically unbiased, which means they approximate the MISE as $n \to \infty$ but are biased for finite sample sizes.

The LSCV selector is then defined by

$$\hat{h}_{\text{LSCV}} := \arg \min_{h_n > 0} \text{LSCV}(h_n).$$

To obtain $\hat{h}_{\text{LSCV}}$, numerical optimization is necessary. However, this process can be complicated by the LSCV function, which may contain multiple local minima, and its objective function can exhibit significant roughness, depending on n and c. Consequently, optimization algorithms may sometimes converge to incorrect solutions. To avoid this, one can verify the solution by plotting $\text{LSCV}(h_n)$ over a range of h_n values or by performing a search over a specified bandwidth grid. Hall (1983) [12] and Stone (1984) [28] demonstrated that the LSCV procedure yields a consistent sequence of smoothing parameters and is, in a specific sense, asymptotically optimal.

Analyzing the MISE in (4.2) requires understanding the kernel moments in (1.2), which is not immediately evident in (4.7). For large n, the bandwidth becomes small enough that the probability mass outside the unit square tends to zero. Using this, Gijbels and Mielniczuk (1990) [11] derived the bias of the mirror reflection estimator (Theorem 3.2). Incorporating this concept, (4.7) can be reformulated as:

$$\text{LSCV}(h) = \frac{R(K)^2}{nh^2} + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} \left[\frac{1}{n^2 h^4} \int_0^1 K\left(\frac{u - \hat{U}_i}{h_n}\right) K\left(\frac{u - \hat{U}_j}{h_n}\right) \, du \int_0^1 K\left(\frac{v - \hat{V}_i}{h_n}\right) K\left(\frac{v - \hat{V}_j}{h_n}\right) \, dv \right.$$

$$\left. - \frac{2}{n(n-1)h^2} K\left(\frac{\hat{U}_i - \hat{U}_j}{h_n}\right) K\left(\frac{\hat{V}_i - \hat{V}_j}{h_n}\right) \right]. \tag{4.8}$$

An instructive exercise shows that for p even, the expectation of (4.8) equals (4.3) minus the constant $R(c(u, v))$. Under the kernel properties in Assumptions 2.2, we define:

$$\gamma(\tau_1, \tau_2) = \int \int K(w_1)K(w_1 + \tau_1)K(w_2)K(w_2 + \tau_2) \, dw_1 \, dw_2 - 2K(\tau_1)K(\tau_2), \tag{4.9}$$

and let

$$\tau_{ij,1} = (\hat{U}_i - \hat{U}_j)/h_n \text{ and } \tau_{ij,2} = (\hat{V}_i - \hat{V}_j)/h_n, \tag{4.10}$$

then (4.8) becomes (replacing $n-1$ by n)

$$LSCV(h_n) = \frac{R(K)^2}{nh_n^2} + \frac{2}{n^2 h_n^2} \sum\sum_{i<j} \gamma(\tau_{ij,1}, \tau_{ij,2}). \tag{4.11}$$

The following theorem provides the mean of the function (4.11) for fixed h_n.

Theorem 4.1. *For the LSCV kernel criterion* (4.11)

$$E[LSCV(h_n)] = AMISE(h_n) - R(c) + O(n^{-1}). \tag{4.12}$$

Proof. Recall that K is a symmetric function with support on $[-1, 1]$. Using this property, we define $\gamma_+(\tau_1, \tau_2)$ and $\gamma_-(\tau_1, \tau_2)$ to represent $\gamma(\tau_1, \tau_2)$ as given in (4.9), over the intervals $[0, 2]^2$ and $[-2, 0]^2$, respectively. For $0 \leq \tau_1, \tau_2 \leq 2$, $\gamma_+(\tau_1, \tau_2)$ is given by:

$$\gamma_+(\tau_1, \tau_2) \equiv \int_{-1}^{1-\tau_1} \int_{-1}^{1-\tau_2} K(w_1)K(w_1 + \tau_1)K(w_2)K(w_2 + \tau_2) \, dw_1 \, dw_2 - 2K(\tau_1)K(\tau_2). \tag{4.13}$$

Similarly, for $-2 \leq \tau_1, \tau_2 \leq 0$, $\gamma_-(\tau_1, \tau_2)$ is defined analogously to (4.13), with the integration limits for both τ_1 and τ_2 replaced by $-1 - \tau$ and 1. Since K is symmetric, it follows that γ is also symmetric.

With these definitions in place, we can now evaluate:

$$E[\gamma(\tau_{ij,1}, \tau_{ij,2})] = \int_0^1 \int_0^1 c(u, v) \left[\int_{u-2h_n}^u \int_{v-2h_n}^v \gamma_+ \left(\frac{u - u_1}{h_n}, \frac{v - v_1}{h_n} \right) c(u_1, v_1) \, du_1 \, dv_1 + \right.$$

$$\left. \int_u^{u+2h_n} \int_v^{v+2h_n} \gamma_- \left(\frac{u - u_1}{h_n}, \frac{v - v_1}{h_n} \right) c(u_1, v_1) \, du_1 \, dv_1 \right] du \, dv$$

$$= h_n^2 \int_0^1 \int_0^1 c(u, v) \left[\int_0^2 \int_0^2 \gamma_+(\tau_1, \tau_2) \{ c(u - \tau_1 h_n, v - \tau_2 h_n) + \right.$$

$$c(u + \tau_1 h_n, v + \tau_2 h_n) \} \, d\tau_1 \, d\tau_2 \right] du \, dv$$

$$= 2h_n^2 \int_0^1 \int_0^1 c(u, v) \left[\int_0^2 \int_0^2 \gamma_+(\tau_1, \tau_2) \left(c(u, v) + \sum_{|\kappa|=2} \frac{h_n^{|\kappa|} \tau_1^{\kappa_1} \tau_2^{\kappa_2}}{\kappa_1! \kappa_2!} \frac{\partial^{|\kappa|} c(u, v)}{\partial u^{\kappa_1} \partial v^{\kappa_2}} + \right. \right.$$

$$\left. \left. \sum_{|\kappa|=4} \frac{h_n^{|\kappa|} \tau_1^{\kappa_1} \tau_2^{\kappa_2}}{\kappa_1! \kappa_2!} \frac{\partial^{|\kappa|} c(u, v)}{\partial u^{\kappa_1} \partial v^{\kappa_2}} \right) d\tau_1 \, d\tau_2 \right] du \, dv, \tag{4.14}$$

where $|\kappa| = \kappa_1 + \kappa_2$ is the total order of differentiation. Now, for κ_1 and κ_2 even

$$\int_0^2 \int_0^2 \tau_1^{\kappa_1} \tau_2^{\kappa_2} \gamma_+(\tau_1, \tau_2) \, d\tau_1 \, d\tau_2 = \frac{1}{2} \int_{-1}^1 K(w_1) \int_{-1}^1 (s_1 - w_1)^{\kappa_1} K(s_1) \, ds_1 \, dw_1 \times$$

$$\int_{-1}^1 K(w_2) \int_{-1}^1 (s_2 - w_2)^{\kappa_2} K(s_2) \, ds_2 \, dw_2 - \mu_{\kappa_1} \mu_{\kappa_2}, \tag{4.15}$$

where we have used the substitution $s_1 = w_1 + \tau_1$ and $s_2 = w_2 + \tau_2$. The expression (4.15) equals to $-1/2, 0, \mu_2^2$ and $3\mu_2^2$ for $\{\kappa_1, \kappa_2\} = \{0,0\}, \{0,2\}, \{2,2\}, \{0,4\}$, respectively. Hence

$$E[\gamma(\tau_{ij,1}, \tau_{ij,2})] = h_n^2 \int_0^1 \int_0^1 \left[-c(u,v)^2 + \frac{\mu_2^2 h_n^4}{4} \left(c(u,v) \frac{\partial^4 c(u,v)}{\partial u^4} + 2c(u,v) \frac{\partial^4 c(u,v)}{\partial u^2 \partial v^2} + c(u,v) \frac{\partial^4 c(u,v)}{\partial v^4} \right) \right] du \, dv$$

$$= h_n^2 \left[-R(c(u,v)) + \frac{\mu_2^2 h_n^4}{4} \int_0^1 \int_0^1 [c_{uu}(u,v) + c_{vv}(u,v)]^2 \right], \tag{4.16}$$

where we have used integration by parts. Thus (4.12) follows from (4.16), (4.11) and (4.2). □

The roughness of $\hat{c}(u,v)$, represented by $R(\hat{c}(\cdot; h_n))$ in (4.7), when using the Epanechnikov kernel can be described in terms of the convolution of the kernel K with itself. Before presenting this result, we first state the following lemma, which will be instrumental in the proof.

Lemma 4.2. *Let K be a symmetric kernel. Then*

$$\int K \left(\frac{u - U_i}{h} \right) K \left(\frac{u - U_j}{h} \right) du = h (K * K) \left(\frac{U_i - U_j}{h} \right),$$

where $K * K$ denotes the convolution of K with itself.

Proof. To prove Lemma 4.2, we start by applying the substitution $u = hw$, which implies $du = h \, dw$. Substituting this into the integral, we obtain:

$$\int K \left(\frac{u - U_i}{h} \right) K \left(\frac{u - U_j}{h} \right) du = h \int K \left(\frac{U_i}{h} - w \right) K \left(w - \frac{U_j}{h} \right) dw. \tag{4.17}$$

The expression on the right side of equation (4.17) can be interpreted as a product of two shifted kernels centered at U_i/h and U_j/h, respectively. For clarity, let us denote these shifted kernels as $K_{U_i}(w) = K(U_i/h - w)$ and $K_{U_j}(w) = K(w - U_j/h)$. Thus, we can rewrite equation (4.17) as:

$$h \int K \left(\frac{U_i}{h} - w \right) K \left(w - \frac{U_j}{h} \right) dw = h \int K_{U_i}(w) K_{U_j}(w) \, dw. \tag{4.18}$$

Now, to further simplify the right side of (4.18), let us make the substitution $w_1 = w - U_j/h$, so that $w = w_1 + U_j/h$ and $dw = dw_1$. Under this substitution $K_{U_j}(w) = K(w_1)$ and $K_{U_i}(w) = K((U_i - U_j)/h - w_1)$. Thus, we can rewrite (4.18) as:

$$h \int K \left(\frac{U_i - U_j}{h} - w_1 \right) K(w_1) \, dw_1 = h (K * K) \left(\frac{U_i - U_j}{h} \right),$$

as required. □

For the Epanechnikov kernel, the convolution $(K * K)(x) = \int K(t) K(x - t) \, dt$ implies that $-1 \le t \le 1$ and $-1 \le x - t \le 1$. These two inequalities imply that $x - 1 \le t \le x + 1$ and

$x \in [-2, 2]$. Thus the integral of the convolution of the Epanechnikov kernel by itself can be broken into three cases based on x as

$$(K * K)(x) = \begin{cases} \int_{-1}^{x+1} \frac{9}{16}(1 - t^2)(1 - (x - t)^2)dt, & -2 \leq x \leq -1, \\ \int_{-1}^{1} \frac{9}{16}(1 - t^2)(1 - (x - t)^2)dt, & -1 \leq x \leq 1, \\ \int_{x-1}^{1} \frac{9}{16}(1 - t^2)(1 - (x - t)^2)dt, & 1 \leq x \leq 2, \\ 0, & \text{otherwise.} \end{cases} \quad (4.19)$$

Proposition 4.3. *Let K be a univariate kernel. The roughness of the estimator $\hat{c}(u, v)$ given in* (3.1)*, when using the Epanechnikov kernel, is of the form*

$$R(\hat{c}) = \frac{81}{256n^2 h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{9} \sum_{m=1}^{9} (K * K)(\xi_{ijlm}) (K * K) (\zeta_{ijlm}), \quad (4.20)$$

where $\xi_{ijlm} = (U_{il} - U_{jm})/h_n$, $\zeta_{ijlm} = (V_{il} - V_{jm})/h_n$, and $(K * K)(x)$ is the convolution of the Epanechnikov kernel with itself.

Proof. Using Definition 2.1,

$$R(\hat{c}) = \frac{1}{n^2} \int \int \left(\sum_{i=1}^{n} \sum_{l=1}^{9} K_{h_n}(u - U_{il}) K_{h_n}(v - V_{il}) \right)^2 du \, dv$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{l=1}^{9} \sum_{j=1}^{n} \sum_{m=1}^{9} \int \int K_{h_n}(u - U_{il}) K_{h_n}(v - V_{il}) K_{h_n}(u - U_{jm}) K_{h_n}(v - V_{jm}) \, du \, dv$$

$$(4.21)$$

Using the Epanechnikov kernel, then (4.21) can be written as

$$R(\hat{c}) = \frac{81}{256n^2 h^4} \sum_{i=1}^{n} \sum_{l=1}^{9} \sum_{j=1}^{n} \sum_{m=1}^{9} \int \int \left\{ \left(1 - \left(\frac{u - U_{il}}{h_n} \right)^2 \right) \left(1 - \left(\frac{v - V_{il}}{h_n} \right)^2 \right) \left(1 - \left(\frac{u - U_{jm}}{h_n} \right)^2 \right) \times \right.$$

$$\left. \left(1 - \left(\frac{v - V_{jm}}{h_n} \right)^2 \right) \right\} du \, dv$$

$$= \frac{81}{256n^2 h_n^4} \sum_{i=1}^{n} \sum_{l=1}^{9} \sum_{j=1}^{n} \sum_{m=1}^{9} \int \left(1 - \left(\frac{u - U_{il}}{h_n} \right)^2 \right) \left(1 - \left(\frac{u - U_{jm}}{h_n} \right)^2 \right) du \times$$

$$\int \left(1 - \left(\frac{v - V_{il}}{h_n} \right)^2 \right) \left(1 - \left(\frac{v - V_{jm}}{h_n} \right)^2 \right) dv,$$

$$= \frac{81}{256n^2 h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{9} \sum_{m=1}^{9} (K * K)(\xi_{ijlm}) (K * K) (\zeta_{ijlm}), \quad (4.22)$$

where we have used Lemma 4.2 to arrive at the last line of (4.22), $(K * K)$ is the convolution of the Epanechnikov kernel by itself defined in (4.19), and $\xi_{ijlm} = (U_{il} - U_{jm})/h_n$, $\zeta_{ijlm} = (V_{il} - V_{jm})/h_n$. $\quad \square$

4.2.2. *Biased cross-validation (BCV).* Unlike the LSCV, which relies solely on the MISE, the Biased Cross-Validation (BCV) approach is a hybrid that combines both plug-in and cross-validation techniques. To introduce the bivariate case, we begin by examining the univariate BCV cases discussed in existing literature. The work by Scott and Terrell (1987) [24] proposed minimizing the AMISE:

$$\text{AMISE}(h) = \frac{R(K)}{nh} + \frac{h^4\mu_2(K)^2}{4}R\big(f''(x)\big), \tag{4.23}$$

where h represents the bandwidth for the univariate kernel K, $R(K)$ and $\mu_2(K)$ are as defined in (1.1) and (1.2), respectively, and $f''(x)$ is the second derivative of the underlying univariate density. For more details on univariate kernel density estimation, see Scott and Terrell (1987) [24] and Wand and Jones (1995) [31]. The only unknown term in the AMISE expression in (4.23) is the second derivative of the underlying density.

Scott and Terrel (1987) [24] proposed an estimate of $R(f''(x))$ of the form $\hat{R}^1(f'') = R(\hat{f}'') - R(K'')/nh^5$, where $\hat{f}''$ is the second derivative of the univariate KDE, K. They gave the corresponding AMISE estimate as

$$BCV^1(h) = \frac{R(K)}{nh} + \frac{h^4\mu_2(K)^2}{4}\hat{R}^1(f'').$$

Hall and Marron (1987) [13] investigated the precision of estimators for the quantity $R(f''(x))$ and derived an estimator using the identity

$$R(f'') = \int f''(x)^2\,dx = \int f^{(iv)}(x)f(x)\,dx = E[f''(x)],$$

while also discussing its implications for bandwidth selection. Utilizing this identity, they demonstrated that $\hat{R}^2(f'') = \frac{1}{n}\sum_{i=1}^{n}\hat{f}_{-i}^{(iv)}(x_i)$, leading to the criterion function

$$BCV^2(h) = \frac{R(K)}{nh} + \frac{h^4\mu_2(K)^2}{4}\hat{R}^2(f'').$$

In the univariate case, the estimator $\hat{R}^2$ has a higher variance than $\hat{R}^1$ (Sain *et al.*, 1994 [23]). However, this increased variance is accompanied by a reduction in bias. Importantly, $\hat{R}^2$ offers greater ease of extension to multivariate cases (Sain *et al.*, 1994 [23]), making it an appealing choice for our analysis due to its implementational simplicity and improved bias characteristics.

Consider the AMISE given in (4.3). We can rewrite this as

$$\text{AMISE}(h) = \frac{R(K)^2}{nh_n^2} + \frac{h_n^4\mu_2(K)^2}{4}\Bigg(\int_0^1\int_0^1 c_{uu}(u,v)^2\,du\,dv$$
$$+ \int_0^1\int_0^1 c_{vv}(u,v)^2\,du\,dv + 2\int_0^1\int_0^1 c_u(u,v)c_v(u,v)\,du\,dv \Bigg). \tag{4.24}$$

where K, $R(K)$ and $\mu_2(K)$ are as used in (4.3) and β in (4.3) has been expanded to obtain the factor in parenthesis in the second term.

Using the bivariate form of $\hat{R}^2$, we can estimate the roughness of the first second–order partial derivative in the AMISE expression as follows:

$$\int_0^1 \int_0^1 c_{uu}(u,v)^2 \, du \, dv = \frac{\partial^4 c(u,v)}{\partial u^4} \, c(u,v) du \, dv = E\left[\frac{\partial^4 c(u,v)}{\partial u^4}\right].$$

The second term in the AMISE expression can be dealt with similarly while the third is estimated as

$$\int_0^1 \int_0^1 c_{uu}(u,v) c_{vv}(u,v) \, du \, dv = \frac{\partial^4 c(u,v)}{\partial u^2 \, \partial v^2} \, c(u,v) du \, dv = E\left[\frac{\partial^4 c(u,v)}{\partial u^2 \, \partial v^2}\right]$$

These expressions can be estimated by

$$n^{-1}\sum_{i=1}^n \frac{\partial^4 \hat{c}_{-i}(U_i, V_i)}{\partial u^4} \tag{4.25}$$

and

$$n^{-1}\sum_{i=1}^n \frac{\partial^4 \hat{c}_{-i}(U_i, V_i)}{\partial u^2 \, \partial v^2}. \tag{4.26}$$

Substituting (4.25) and (4.26) into (4.24) we obtain the BCV function. Typically, a kernel with compact support, such as the Epanechnikov kernel, may not suffice for higher–order derivatives due to discontinuities at the boundaries. A Gaussian kernel or other infinitely differentiable kernels are more appropriate choices.

5. Simulation Study

One of the most insightful aspects of working with copulas is the ability to create exploratory plots that visually represent the dependence structure. In this section, we present visualizations of various aspects of the true density and the mirror–reflection estimator, based on simulated data ($n = 500$) generated from a Frank copula with parameter $\theta = 5$ (Kendall's $\tau \approx 0.46$). All computations are coded on R Studio software. Note that the Frank copula has a bounded density. The simple mirror–reflection estimator fails to capture the tail behavior of unbounded densities, such as those of the Gaussian copula or any copula exhibiting tail dependence.

Figure 1(a) below shows the scatterplot of the original data generated from the Frank copula model. In Figure 1(b), the data is transformed using the inverse of the Gaussian cumulative distribution function (CDF), which highlights lack of tail dependencies between the variables.

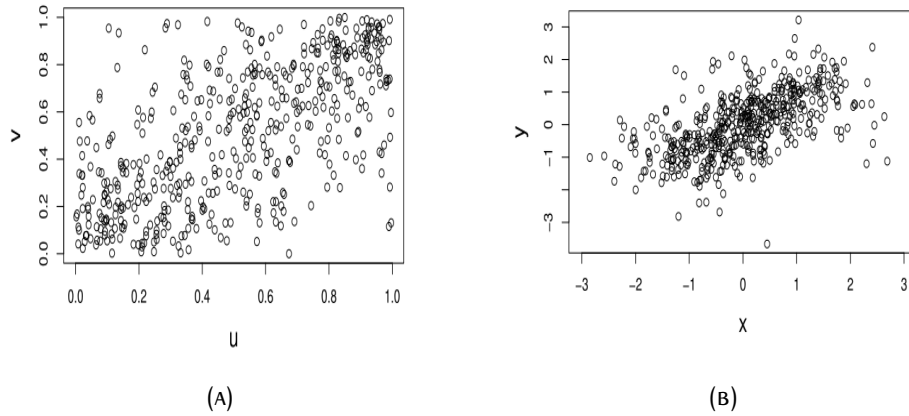(A)   (B)

FIGURE 1. (a) Scatterplot of original data from the Frank copula sample and (b) scatterplot of the transformed sample.

To address the bandwidth selection problem, we considered various sample sizes and calculated the bandwidths that optimize the AMISE and minimize the LSCV(h) criterion, respectively. Table 1 presents a comparison of the optimal bandwidths (bw), h_n^* and $\hat{h}_{LSCV}$, along with the ISE for different sample sizes. The rule–of–thumb selector has a convergence rate much faster than the LSCV selector.

	$n = 100$		$n = 200$		$n = 500$		$n = 1000$	
	bw	ISE	bw	ISE	bw	ISE	bw	ISE
h_n^*	0.239	1.138	0.213	1.030	0.183	0.917	0.163	0.652
h_{LSCV}	0.244	1.227	0.238	1.094	0.196	1.052	0.199	1.007

TABLE 1. Comparison of h_n^* and $\hat{h}_{LSCV}$ for different sample sizes (n) in terms of bandwidth (bw) and ISE.

Marginal normal contour plots of the true density and the mirror–reflection estimator are shown in Figure 2. Figure 2 (a) provides the contour plots of the true density, while Figure 2 (b) displays the mirror–reflection estimator based on the simulated data. Bandwidths for the kernel density estimates were selected on the bases of AMISE–optimality. Examining the contour plots in Figure 2, we conclude that the bandwidth selection rules are functioning appropriately for our finite samples. Using smaller bandwidths(resp., sample sizes) would result in wiggly estimates, which would make the visualizations a little unpleasant.
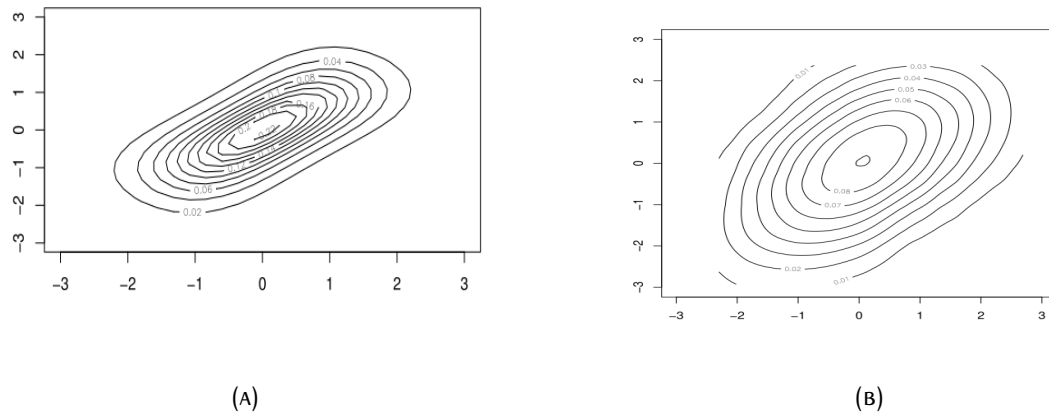
(A)

(B)

FIGURE 2. (a) Contour plot of the true density; (b) Contour plot of the mirror reflection estimates on simulated data (n=500) of a Frank copula. Bandwidths are selected based on the LSCV criterion.

Plot (a) in Figure 3 illustrates the mirror–reflection estimate of the density of the transformed data. The perspective plots in (b) shows the estimated density based on the simulated data of the Frank copula. Figure 3 (b) reveals some deviation of the estimated density from the true density (see the true Frank copula density plot in Nagler 2014 [19]), mainly due to the undersmoothing in the central region of the estimates. However, our bandwidth selection approach strikes a favorable balance between bias and variance, resulting in plots that closely align with the true density even with smaller samples.
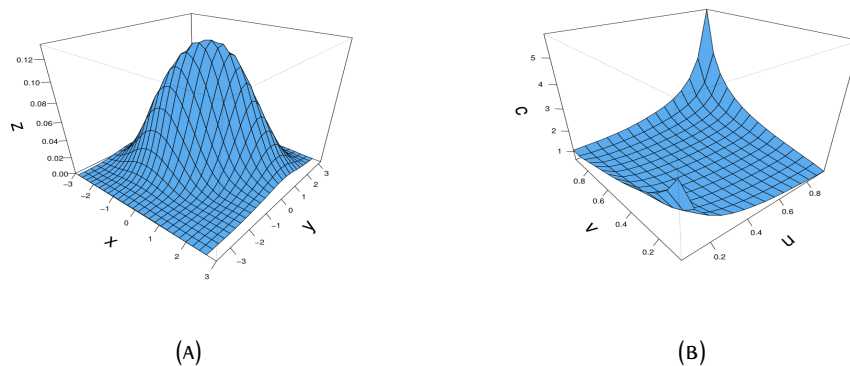

(A)

(B)

FIGURE 3. (a) Mirror reflection estimate for the Frank copula transformed sample; (b) perspective plot for the true density of the Frank copula, and (c) perspective plot of the mirror–reflection kernel density estimate.

6. Real-life data Application

In the practical application part of our study, we used the Wisconsin Breast Cancer Diagnostic Dataset (WBCDD) for analysis. This data set was derived from the examination of cell nuclei characteristics in 569 images obtained by Fine Needle Aspiration (FNA) of breast masses. Each sample in the dataset is categorized as either 'Benign' or 'Malignant', corresponding to noncancerous and cancerous cases, respectively.

The data set was compiled through the efforts of Dr. William Wolberg of Wisconsin Hospital, who provided the breast mass images, and William Nick Street from the Department of Computer Sciences, University of Wisconsin, who digitized the images in November 1995 (O. Mangasarian and W. Wolberg [17]). It contains 569 instances with 30 numeric features, in addition to an ID column and a class label. Among these samples, 357 are benign, while 212 are malignant. Notably, the data set does not have any missing values.

For this particular study, we chose two variables: the mean radius (column 2) and the mean concavity (column 8). The first step involved empirical estimation of the marginal distributions of these variables. This was achieved by computing their empirical cumulative distribution functions (ECDFs), $\hat{F}_1$ and $\hat{F}_2$, based on the observed data. Specifically, for an observation (X_i, Y_i), the transformed values were calculated as $(\hat{U}_i, \hat{V}_i)$, where $\hat{U}_i = \hat{F}_1(X_i), \quad \hat{V}_i = \hat{F}_2(Y_i)$. This transformation effectively maps the data to the unit square $[0, 1]^2$ while preserving the dependence structure between the variables. Once the data was transformed to uniform margins, a scatter plot was created to visualize the relationships between the variables in their new scale. This visualization offered insights into the structure of dependence present in the dataset, even before any further transformations.

Figure 4 (a) shows a scatter plot of the data for the two variables: mean radius (column 2) and mean concavity (column 8) from the Wisconsin Cancer dataset, after transformation using their marginal empirical distribution functions. Figure 4 (b) displays a scatter plot of the same data after transformation to the standard normal distribution. The scatter plot in (b) shows that the data have neither lower nor upper tail dependence.
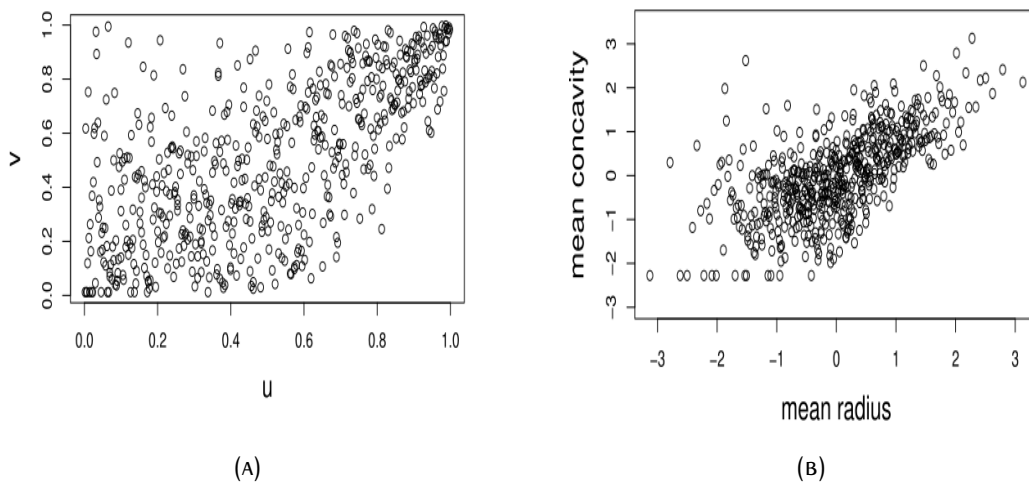
(A)

(B)

FIGURE 4. (a) Scatter plot of data after transformation by its marginal empirical distribution function; (b) scatter plot of the transformed data with standard normal margins.

The dependence structure of the WBCDD was then depicted using a contour plot. Figure 5 (a) shows contour plot of the copula density combined with standard normal margins, (b) the kernel density estimate for transformed sample (with standard normal margins), (c) a surface or perspective plot of the copula density. The bandwidths were selected on the basis of the optimality of the LSCV criterion.
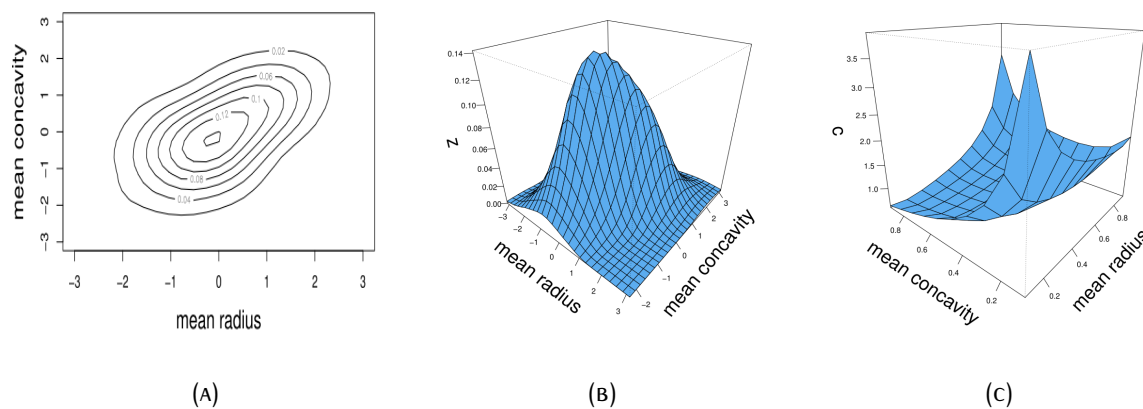

(A)

(B)

(C)

FIGURE 5. Exploratory visualizations of the WBCDD data and copula density: (a) Contour plot of the copula density combined with standard normal margins; (b) kernel density estimate for transformed sample (with standard normal margins); and (c) surface plot of the copula density.
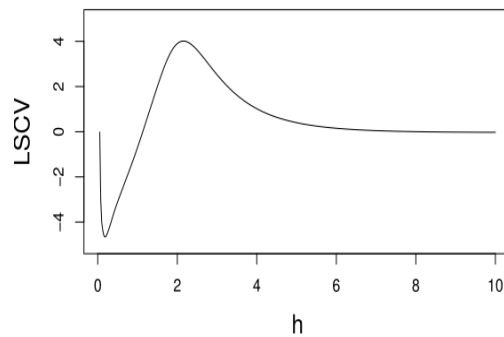
FIGURE 6. LSCV curve for the mirror–reflection estimator applied to the WBCDD, showing the optimal smoothing parameter $h_{\mathrm{LSCV}} = 0.031$.

In Figure 6, we provided a plot of the LSCV curve for our mirror–reflection estimator for the WBCDD. The optimal LSCV smoothing parameter was found to be $h_{LSCV} = 0.031$.

7. CONCLUSION

Methods for copula density estimation have been explored in the literature but remain insufficiently investigated, primarily due to the challenge of unbounded copula densities and their derivatives at the boundaries of $[0, 1]^2$.

In this study, we focused on estimating copula densities that are bounded within the unit square. This assumption facilitated a straightforward derivation of the asymptotic bias using Taylor expansion techniques. We introduced two kernel smoothing methods and demonstrated that the rule–of–thumb approach provides superior bandwidth selection, particularly in cases where the margins have unbounded support.

We also proved a theorem establishing that the LSCV kernel criterion has an asymptotic expectation equal to the AMISE minus the roughness of the copula density. Simulation results showed that AMISE-optimal bandwidth selection is preferable in scenarios where the margins have unbounded support, even when the kernel used for estimation is bounded. Finally, for any copula to be applied to real–world data, it is essential to ensure that the data's distribution aligns with the theoretical properties of the chosen copula.

Competing interests: The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

[1] K. Behnen, M. Hušková, G. Neuhaus, Rank Estimators of Scores for Testing Independence, Stat. Risk Model. 3 (1985), 239–262. https://doi.org/10.1524/strm.1985.3.34.239.

[2] A. Bowman, P. Hall, T. Prvan, Bandwidth Selection for the Smoothing of Distribution Functions, Biometrika, 85 (1998), 799–808. https://www.jstor.org/stable/2337484.

[3] A.W. Bowman, An Alternative Method of Cross-Validation for the Smoothing of Density Estimates, Biometrika 71 (1984), 353–360. https://doi.org/10.1093/biomet/71.2.353.

[4] A. Charpentier, J.D. Fermanian, O. Scaillet, The Estimation of Copulas: Theory and Practice, in: Copulas: From the Theory to Aplications in Finance, (2007).

[5] X. Chen, Y. Fan, Estimation and Model Selection of Semiparametric Copula-Based Multivariate Dynamic Models under Copula Misspecification, J. Econometrics 135 (2006), 125–154. https://doi.org/10.1016/j.jeconom.2005.07.027.

[6] S.X. Chen, T.M. Huang, Nonparametric Estimation of Copula Functions for Dependence Modelling, Can. J. Stat. 35 (2007), 265-285. https://www.jstor.org/stable/20445253.

[7] B. Choroś, R. Ibragimov, E. Permiakova, Copula Estimation, in: P. Jaworski, F. Durante, W.K. Härdle, T. Rychlik (Eds.), Copula Theory and Its Applications, Springer, Berlin, Heidelberg, 2010: pp. 77–91. https://doi.org/10.1007/978-3-642-12465-5_3.

[8] W.F. Darsow, B. Nguyen, E.T. Olsen, Copulas and Markov Processes, Illinois J. Math. 36 (1992), 600-642.

[9] P. Hominal, P. Deheuvels, Estimation Non Paramétrique de la Densité Compte-tenu d'Informations sur le Support, Rev. Stat. Appl. 27 (1979), 47-68.

[10] F. Durante, C. Sempi, Principles of Copula Theory, CRC Press, (2015).

[11] I. Gijbels, J. Mielniczuk, Estimating the Density of a Copula Function, Commun. Stat. - Theory Methods 19 (1990), 445–464. https://doi.org/10.1080/03610929008830212.

[12] P. Hall, Large Sample Optimality of Least Squares Cross-Validation in Density Estimation, Ann. Stat. 11 (1983), 1156-1174. https://doi.org/10.1214/aos/1176346329.

[13] P. Hall, J.S. Marron, Estimation of Integrated Squared Density Derivatives, Stat. Prob. Lett. 6 (1987), 109–115. https://doi.org/10.1016/0167-7152(87)90083-6.

[14] S.G. Iyengar, P.K. Varshney, T. Damarla, A Parametric Copula-Based Framework for Hypothesis Testing Using Heterogeneous Data, IEEE Trans. Signal Process. 59 (2011), 2308–2319. https://doi.org/10.1109/TSP.2011.2105483.

[15] H. Joe, Multivariate Models and Multivariate Dependence Concepts, Chapman and Hall/CRC, 1997. https://doi.org/10.1201/9780367803896.

[16] M. Longla, F.D. Ndikwa, M.N. Muia, P.S. Takam, Perturbations of Copulas and Mixing Properties, J. Korean Stat. Soc. 51 (2022), 149–171. https://doi.org/10.1007/s42952-021-00133-5.

[17] L. Mangasarian, W. Wolberg, Cancer Diagnosis via Linear Programming, Computer Sciences Technical Report #958, University of Wisconsin-Madison, (1990).

[18] M.M. Muia, Dependence and Mixing for Perturbations of Copula-Based Markov Chains, Dissertation, The University of Mississippi, (2024). https://egrove.olemiss.edu/etd/2847.

[19] T. Nagler, Kernel Methods for Vine Copula Estimation, Master's Thesis, Technische Universität München, (2014).

[20] R.B. Nelsen, An Introduction to Copulas, Springer, New York, 2006. https://doi.org/10.1007/0-387-28678-0.

[21] M. Omelka, I. Gijbels, N. Veraverbeke, Improved Kernel Estimation of Copulas: Weak Convergence and Goodness-of-Fit Testing, Ann. Stat. 37 (2009), 3023-3058. https://doi.org/10.1214/08-AOS666.

[22] M. Rudemo, Empirical Choice of Histograms and Kernel Density Estimators, Scandinavian J. Stat. 9 (1982), 65-78. https://www.jstor.org/stable/4615859.

[23] S.R. Sain, K.A. Baggerly, D.W. Scott, Cross-Validation of Multivariate Densities, J. Amer. Stat. Assoc. 89 (1994), 807–817. https://doi.org/10.1080/01621459.1994.10476814.

[24] D.W. Scott, G.R. Terrell, Biased and Unbiased Cross-Validation in Density Estimation, J. Amer. Stat. Assoc. 82 (1987), 1131–1146. https://doi.org/10.1080/01621459.1987.10478550.

[25] E.F. Schuster, Incorporating Support Constraints into Nonparametric Estimators of Densities, Commun. Stat. – Theory Methods 14 (1985), 1123–1136. https://doi.org/10.1080/03610928508828965.

[26] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Routledge, 2018. https://doi.org/10.1201/9781315140919.

[27] M. Sklar, Fonctions de Répartition à N Dimensions et Leurs Marges, Ann. l'ISUP, 8 (1959), 229-231. https://hal.science/hal-04094463.

[28] C.J. Stone, An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates, Ann. Stat. 12 (1984), 1285-1297. https://www.jstor.org/stable/2241002.

[29] G.R. Terrell, D.W. Scott, Oversmoothed Nonparametric Density Estimates, J. Amer. Stat. Assoc. 80 (1985), 209–214. https://doi.org/10.1080/01621459.1985.10477163.

[30] P.K. Trivedi, D.M. Zimmer, Copula Modeling: An Introduction for Practitioners, Found. Trends Econometrics 1 (2006), 1–111. https://doi.org/10.1561/0800000005.

[31] M.P. Wand, M.C. Jones, Kernel Smoothing, Chapman and Hall/CRC, 1995. https://doi.org/10.1201/b14876.

[32] Y. Wang, H. Ni, S. Wang, Nonparametric Bivariate Copula Estimation Based on Shape-Restricted Support Vector Regression, Knowl.-Based Syst. 35 (2012), 235–244. https://doi.org/10.1016/j.knosys.2012.05.004.