Transformation-Kernel Estimation of Copula Densities

Kuangyu Wen[*], Ximing Wu[†]

Abstract

Standard kernel estimator of copula densities suffers from boundary biases and inconsistency due to unbounded densities. Transforming the domain of estimation into an unbounded one remedies both problems, but also introduces an unbounded multiplier that may produce erratic boundary behaviors in the final density estimate. We propose an improved transformation-kernel estimator that employs a smooth directional tapering device to counter the undesirable influence of the multiplier. We establish the theoretical properties of the new estimator and its automatic higher order improvement under Gaussian copulas. We present two practical methods of optimal smoothing parameter selection. Extensive Monte Carlo simulations demonstrate the competence of the proposed estimator in terms of global and tail performance. Two real-world examples are provided.

JEL Classification: C01; C14

Keywords: *kernel density estimation; copula density; boundary bias; transformation*

[*]International School of Economics and Management, Capital University of Economics and Business, Beijing 100070, P.R. China; Email: kweneco@gmail.com.

[†]Corresponding author. Department of Agricultural Economics, Texas A&M University, College Station, TX 77843, U.S.A.; Email: xwu@tamu.edu.

1 Introduction

The past two decades have seen increasing use of copulas in multivariate analyses. Given a bivariate random vector $(X, Y)^\top$, denote its joint cumulative distribution function (cdf) by F and corresponding marginal distributions by F_X and F_Y. According to Sklar (1959), we can rewrite the joint distribution

$$F(x, y) = C\left(F_X(x), F_Y(y)\right),$$

where C is termed the copula function. If F_X and F_Y are continuous, C is unique. The copula approach facilitates multivariate analyses by allowing separate modeling of the marginal distributions and copula, which completely characterizes the dependence between X and Y. As a result of these advantages, copula has found widespread applications in many areas of economic and statistical analyses, especially modern quantitative finance and risk management. For book-length treatments of copula, see Joe (1997) and Nelsen (2006).

A bivariate copula function C is a cumulative distribution function of random vector $(U, V)^\top$ defined on the unit square $\mathcal{I} = [0, 1]^2$, with uniform marginal distributions as $U = F_X(X)$ and $V = F_Y(Y)$. If C is absolutely continuous, it admits a probability density function (pdf) of the form

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

where c is called the copula density. Compared with the copula distribution, copula density is more readily interpretable in many aspects as suggested by Geenens et al. (2014). Moreover, Fermanian (2005, 2012), Scaillet (2007) and Lin and Wu (2015) demonstrate the advantages of density-based tests of copula specification.

This study concerns the estimation of copula densities. There exist two general approaches. Parametric copula estimations, if correctly specified, are efficient (Nelsen, 2006; Genest et al., 1995; Chen and Fan, 2006a,b; Chen et al., 2006; Patton, 2006; Lee and Long, 2009; Prokhorov and Schmidt, 2009; Chen et al., 2010; Okhrin et al., 2013; Fan and Patton, 2014). However, they may suffer from specification errors, particularly because parametric copulas are often parametrized by one or two parameters and therefore rather restrictive. For instance, the commonly used Gaussian copulas do not allow asymmetric correlation or non-zero tail dependence, and therefore are not suitable for the modeling of financial returns, which often exhibit asymmetric dependence and tend to move together under extreme market conditions. Nonparametric copula estimations offer a flexible alternative. Various nonparametric techniques have been adapted to copula density estimation, including the splines (Shen et al., 2008; Kauermann et al., 2013), wavelets (Hall and Neumeyer, 2006; Genest et al., 2009; Autin et al., 2010), Bernstein polynomials (Sancetta and Satchell, 2004; Sancetta, 2007; Bouezmarni et al., 2010, 2012, 2013; Janssen et al., 2014; Taamouti et al., 2014), and maximum penalized likelihood (Qu and Yin, 2012; Gao et al., 2015). Fermanian and Scaillet (2003), Chen and Huang (2007) and Omelka et al. (2009) study kernel estimation of copula distribution functions.

We focus on kernel type estimators of copula densities in this study. Despite being one of the most commonly used nonparametric methods, kernel estimation has found but scant applications in copula densities (Gijbels and Mielniczuk, 1990; Charpentier et al., 2006; Geenens et al., 2014). There are at least two reasons for this. First, standard kernel estimators are known to suffer from boundary biases while copula densities are defined on bounded supports. Second, the consistency of kernel density estimators requires that the underlying densities are bounded on their supports. However, many copula densities are unbounded at the boundaries. This unboundedness violates a key assumption of the kernel density estimation and renders it inconsistent. Gijbels and Mielniczuk (1990) consider a mirror reflection kernel estimator to correct boundary biases. This approach works best when the underlying densities have null first derivatives at the boundaries and therefore may not be suitable to copula densities that tend to infinity at the boundaries.

Transformation of variables provides an alternative solution to boundary bias. This idea dates back to Wand et al. (1991) and Marron and Ruppert (1994). Specifically, one first transforms a variable with a bounded support to that with an unbounded support. The density of the transformed variable can then be estimated, free of boundary bias, using the standard kernel estimator. Lastly the estimated density of the transformed variable is 'back-transformed' to obtain an estimated density of the original variable. This approach is explored by Charpentier et al. (2006) in kernel estimation of copula densities. Because the back-transformation introduces to the final estimate a possibly unbounded multiplicative factor near the boundaries, it also solves the inconsistency issue that plagues standard kernel estimators. This unbounded multiplier, however, proves to be a double-edged sword as it can lead to erratic tail behaviors in the final estimate. Geenens et al. (2014) propose a local likelihood modification to the standard transformation-kernel estimator to tackle this problem.

Given that copulas are widely used in financial analyses and risk management, wherein reliable tail estimates are of crucial importance, great care should be exercised in the estimation of copula densities especially their tails. This study therefore aims to provide an improved transformation-kernel estimator of the copula density that is free of boundary biases, consistent under unbounded densities and possesses satisfactory tail properties. Our solution employs a smooth infinitesimal tapering device to mitigate the aforementioned undesirable influence of the unbounded multiplier. Moreover, it incorporates an interaction parameter to further allow directional tapering since copula densities are often elongated, typically along either diagonal of the unit square. We derive the statistical properties of the proposed estimator and demonstrate that it dominates the conventional transformation-kernel estimator asymptotically. Based on our theoretical analyses, we propose two practical methods of selecting optimal smoothing parameters, which are computationally simple.

The modified transformation-kernel estimator has some practical advantages. First, it produces a *bona fide* density. Second, it retains the simplicity of the conventional transformation-kernel estimator with a fixed transformation and a single global bandwidth, in contrast to data driven transformation or locally varying bandwidths. Third, we show that for Gaussian copulas, our estimator obtains a faster convergence rate. Consequently, it yields outstanding performance when the underlying copulas are Gaussian or near Gaussian—a common occurrence in the analyses of financial data. The second and third features distinguish our estimator from the local log-likeliood estimation of transformation kernel estimator by Geenens et al. (2014). Numerical experiments and empirical applications demonstrate the competence of the proposed method in terms of global and tail performance.

The rest of the article is structured as follows. In Section 2, we briefly describe the transformation-kernel estimator and then introduce the modified transformation-kernel estimator under simple diagonal bandwidth matrix. In Section 3, we derive the asymptotic properties of the new estimator, followed by two practical methods to select the optimal smoothing parameters in Section 4. We present improved convergence results under Gaussian copulas in Section 5. Extensions to non-diagonal bandwidth matrix are considered in Section 6. We report simulation results in Section 7 and apply the proposed method to two real world datasets in Section 8. The last section concludes. Some technical details and proofs of theorems are gathered in the appendices.

2 Estimator

2.1 Preliminaries

Consider an i.i.d. sample $(U_i = F_X(X_i), V_i = F_Y(Y_i)) \in (0,1)^2, i = 1, \ldots, n$, from an absolutely continuous distribution with a copula density c. The standard kernel estimator (KDE) of c is defined as

$$\hat{c}(u,v) = \frac{1}{n|\boldsymbol{H}|^{1/2}} \sum_{i=1}^{n} \boldsymbol{K}\left(\boldsymbol{H}^{-1/2}\begin{pmatrix} u-U_i \\ v-V_i \end{pmatrix}\right), \ (u,v) \in (0,1)^2, \tag{2.1}$$

where $\boldsymbol{K}$ is a bivariate kernel function and $\boldsymbol{H}$ is a symmetric positive-definite bandwidth matrix. In this section we adopt the simplification that $\boldsymbol{H} = h^2\boldsymbol{I}$ for some $h > 0$, where $\boldsymbol{I}$ is a two-dimensional identity matrix. This bandwidth specification eases exposition and subsequent theoretical analysis with little loss of generality as it is qualitatively no different from that of a general $\boldsymbol{H}$. We shall return to the case of a general $\boldsymbol{H}$ in Section 6. Under the simplified bandwidth, we can write the KDE as

$$\hat{c}(u,v) = \frac{1}{n}\sum_{i=1}^{n} K_h(u-U_i)K_h(v-V_i),$$

where $K_h(\cdot) = K(\cdot/h)/h$ and K is a univariate kernel function.

Since copula density is defined on the bounded support $\mathcal{I} = [0,1]^2$, the KDE $\hat{c}$ suffers boundary biases. Marron and Ruppert (1994) propose a transformation approach to remedy boundary bias of the KDE. Their method is employed by Charpentier et al. (2006) and Geenens et al. (2014) in kernel estimation of copula densities. In particular, define the Probit transformation

$$S = \Phi^{-1}(U) \text{ and } T = \Phi^{-1}(V),$$

where Φ is the standard Gaussian cumulative distribution function (CDF) and Φ^{-1} is its quantile function. Denote the density of (S,T) by g. It follows that

$$g(s,t) = c(\Phi(s),\Phi(t))\phi(s)\phi(t), \quad \forall (s,t) \in \mathcal{R}^2,$$

where ϕ is the PDF of the standard Gaussian distribution. Geenens et al. (2014) show that g and its partial derivatives up to the second order are uniformly bounded on $\mathcal{R}^2$, even though c may be unbounded on $\mathcal{I}$. It follows that g can be estimated, free of boundary biases, by the following standard KDE

$$\hat{g}(s,t) = \frac{1}{n}\sum_{i=1}^{n} K_h(s-S_i)K_h(t-T_i),$$

where $S_i = \Phi^{-1}(U_i)$ and $T_i = \Phi^{-1}(V_i), i = 1,\ldots,n$. The transformation-kernel estimator (TKE) of the copula density is then obtained via a back-transformation to its original coordinate:

$$\hat{c}_t(u,v) = \frac{\hat{g}(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}, \quad \forall (u,v) \in (0,1)^2. \tag{2.2}$$

Geenens et al. (2014) show that $\hat{c}_t$ is free of boundary biases and retains many desirable properties of $\hat{g}$.

Since in practice the true marginal distributions are usually not observed, it is customary in copula estimations to replace (U_i, V_i) with the so-called "pesudo-data"

$$\hat{U}_i = \frac{n}{n+1}\hat{F}_{n,X}(X_i) \quad \text{and} \quad \hat{V}_i = \frac{n}{n+1}\hat{F}_{n,Y}(Y_i), \tag{2.3}$$

where $\hat{F}_{n,X}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \le x\}}$ is the empirical distribution of X and $\hat{F}_{n,Y}$ is similarly defined.

Remark 1. *The influence of using the pseudo-data in place of the true CDF is asymptotically negligible in kernel estimation of copula densities, as demonstrated by Genest and Segers (2010) and Geenens et al. (2014). Intuitively this is because the empirical distribution function has a* $\sqrt{n}$ *convergence rate, faster than that of the kernel density estimator. In addition, there exist some advantages of using the pseudo-data in copula density estimation. First, since the pseudo-data are more "uniform", subsequent estimates may have smaller variance; for details, see Charpentier et al. (2006) and Genest and Segers (2010). Second, oftentimes* U *and* V *are given by some statistical models* $U(\beta)$ *and* $V(\beta)$*; for instance, they may be residuals from a GARCH type model parametrized by a finite dimensional parameter* β*. Let* $\hat{\beta}$ *be a* $\sqrt{n}$*-consistent estimate of* β *and* $(U_i(\hat{\beta}), V_i(\hat{\beta}))$ *be estimated marginal CDF's associated with* $\hat{\beta}$*. Naturally subsequent copula estimations are influenced by the estimation of* β*. Nonetheless Chen and Fan (2006a) show that the asymptotic distributions of copula estimates based on the pseudo-data* $(\hat{U}_i(\hat{\beta}), \hat{V}_i(\hat{\beta}))$ *are invariant to the estimation of* β *under mild regularity conditions. This is generally not true if* $(U_i(\hat{\beta}), V_i(\hat{\beta}))$ *are used in copula estimations.*

2.2 Modified transformation-kernel estimator

The Probit transformation aptly remedies the boundary biases of kernel density estimation. Although not intended by its original motivation (Marron and Ruppert, 1994), it also resolves the inconsistency of the KDE due to unbounded densities. To see this, note that the back-transformation introduces to $\hat{c}_t$ in (2.2) a "multiplier", $\{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))\}^{-1}$, which grows unboundedly as u and/or v tend to zero or one. As a result, the TKE of copula density admits unbounded densities at the boundaries. This appealing feature, however, can be compromising at the same time as the unbounded multiplier near the boundaries may produce erratic boundary behavior of the TKE. Slight biases of $\hat{g}$ at the tails might be magnified substantially by the multiplier, resulting in large biases of $\hat{c}_t$ near the boundaries, especially at the corners of $\mathcal{I}$.

Wen and Wu (2015) introduce an improved transformation-kernel estimator for densities on the unit interval, employing a tapering device to mitigate the multiplier of the TKE. In this study we propose a modified transformation-kernel estimator (MTK) for copula densities in a similar spirit. Specifically, we replace the multiplier of the TKE with

$$\{\phi_{1+\theta_1}(\Phi^{-1}(u))\phi_{1+\theta_1}(\Phi^{-1}(v))\}^{-1},$$

where $\theta_1 > 0$ and $\phi_{1+\theta_1}(\cdot)$ is the Gaussian PDF with mean 0 and standard deviation $1 + \theta_1$ (for simplicity the subscript is omitted when $\theta_1 = 0$). This simple tapering device reduces the multiplier considerably near the boundaries since $\{\phi_{1+\theta_1}(\Phi^{-1}(u))\}^{-1}$ increases much slower than $\{\phi(\Phi^{-1}(u))\}^{-1}$ as u approaches 0 or 1.

We shall hereafter rewrite the modified multiplier, up to a normalization constant, as follows

$$\{\phi_{1+\theta_1}(\Phi^{-1}(u))\phi_{1+\theta_1}(\Phi^{-1}(v))\}^{-1} \sim \frac{\exp\left(-\theta_1[\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2]\right)}{\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)}$$

for it facilitates subsequent analysis. Since copula densities are often elongated along one of the diagonals of $\mathcal{I}$ in the presence of dependence between U and V, it is desirable that the degree of tapering adapts to the orientation of copula densities. This motivates us to further introduce an 'interaction' term that allows for directional tapering. The modified multiplier thus takes the more general form

$$\frac{\exp\left(-\theta_1[\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2] - \theta_2\Phi^{-1}(u)\Phi^{-1}(v)\right)}{\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)},$$

where the coefficient θ_2 controls the magnitude of the interaction term. The modified transformation-kernel estimator then takes the form, for any $(u, v) \in (0, 1)^2$,

$$\begin{aligned}\hat{c}_m(u, v) =& \frac{\exp\left(-\theta_1[\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2] - \theta_2\Phi^{-1}(u)\Phi^{-1}(v)\right)}{n\eta\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)} \\ & \sum_{i=1}^{n} K_h(\Phi^{-1}(u) - \hat{S}_i)K_h(\Phi^{-1}(v) - \hat{T}_i),\end{aligned} \tag{2.4}$$

where η is a normalization factor such that $\hat{c}_m$ integrates to unity.

Remark 2. *For consistency, it is necessary that* $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top \to \mathbf{0}$ *as* $n \to \infty$ *such that the adjustment is asymptotically negligible. In practice, small values of* $\boldsymbol{\theta}$ *ensure that this modification effectively tames the possibly erratic boundary behaviors of the TKE and at the same time has little effect on the density estimate in the interior of* $\mathcal{I}$*. The tuning parameters* $\boldsymbol{\theta}$ *smoothly control the degree of tapering and need to be chosen carefully for good performance.*

Remark 3. *It is well known that the choice of kernel functions has little effect on the performance of kernel estimation. The Gaussian kernel is a popular choice and enjoys some appealing properties (see, e.g., Chaudhuri and Marron (1999)). We note that for the proposed MTK estimator with a Probit transformation, the Gaussian kernel admits an analytical normalization factor given by*

$$\eta = \frac{1}{n\delta} \sum_{i=1}^{n} \exp\left\{ -\frac{(4h^2\theta_1^2 - h^2\theta_2^2 + 2\theta_1)(\hat{S}_i^2 + \hat{T}_i^2) + 2\theta_2 \hat{S}_i \hat{T}_i}{2\delta^2} \right\},$$

where

$$\delta = \sqrt{h^4(4\theta_1^2 - \theta_2^2) + 4h^2\theta_1 + 1}.$$

Thus combining the Gaussian kernel with the proposed Probit transformation-kernel estimator offers the practical advantage of an analytical normalization factor that can be easily calculated. We therefore restrict our attention to Gaussian kernel functions for the rest of this study.

Remark 4. *By definition, the marginal distributions of a copula are uniform. Most nonparametric copula density estimators, including the present one, only satisfy this property asymptotically. Nonetheless given a nonparametric copula density estimator* $\hat{c}$*, a simple mitigation to this problem is to divide the estimate by the product of its marginal densities, yielding*

$$\frac{\hat{c}(u, v)}{\int_0^1 \hat{c}(u, z)dz \int_0^1 \hat{c}(z, v)dz}.$$

This renormalized estimator is asymptotically equivalent to the original estimator under mild conditions and usually provides more accurate estimates (Nagler, 2016).

3 Asymptotic properties

The theoretical properties of TKE for copula densities have been studied by Charpentier et al. (2006) and Geenens et al. (2014). Given a function $f(x, y)$, denote its partial derivative by $f^{(r_1,r_2)}(x, y) = \partial^{(r_1+r_2)} f(x, y)/\partial x^{r_1} \partial y^{r_2}$ if it exists. They establish that under Assumptions 1-3 and 5 in Appendix A, in the ST-domain,

$$\sqrt{nh^2}\left(\hat{g}(s,t)-g(s,t)-b_g(s,t)\right)\xrightarrow{d}\mathcal{N}\left(0,\sigma_g^2(s,t)\right),\quad \forall(s,t)\in\mathcal{R}^2, \tag{3.1}$$

where $b_g(s,t)=\frac{h^2\left(g^{(2,0)}+g^{(0,2)}\right)(s,t)}{2}$ and $\sigma_g^2(u,v)=\frac{g(s,t)}{4\pi}$. It follows that in the UV-domain,

$$\sqrt{nh^2}\left(\hat{c}_t(u,v)-c(u,v)-b_t(u,v)\right)\xrightarrow{d}\mathcal{N}\left(0,\sigma_t^2(u,v)\right),\quad \forall(u,v)\in(0,1)^2, \tag{3.2}$$

where $b_t(u,v)=\frac{h^2\left(g^{(2,0)}+g^{(0,2)}\right)\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)}{2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$ and $\sigma_t^2(u,v)=\frac{c(u,v)}{4\pi\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$. In particular, Geenens et al. (2014) show that the TKE based on pseudo-data shares the same asymptotic distribution with the (infeasible) estimator based on true marginal distributions.

Define

$$J(u,v;h,\boldsymbol{\theta})=\frac{1}{\eta}\exp\left(-\theta_1[\{\Phi^{-1}(u)\}^2+\{\Phi^{-1}(v)\}^2]-\theta_2\Phi^{-1}(u)\Phi^{-1}(v)\right). \tag{3.3}$$

The MTK can be rewritten as

$$\hat{c}_m(u,v)=J(u,v;h,\boldsymbol{\theta})\hat{c}_t(u,v), \tag{3.4}$$

suggesting that the MTK introduces a multiplicative adjustment to the TKE. The adjustment $J(u,v;h,\boldsymbol{\theta})$ is controlled by the tuning parameters $\boldsymbol{\theta}$. When $\boldsymbol{\theta}=\mathbf{0}$, $J(u,v;h,\boldsymbol{\theta})=1$ and the MTK coincides with the TKE.

Given a fixed point $(u,v)\in(0,1)^2$, a Taylor expansion of $J(u,v;h,\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ at zero yields

$$J(u,v;h,\boldsymbol{\theta})=1+\boldsymbol{\theta}^\top\boldsymbol{B}\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)+o\left(\boldsymbol{\theta}\right), \tag{3.5}$$

where

$$\boldsymbol{B}(s,t)=\begin{pmatrix}2-s^2-t^2\\ \mathbb{E}[S_iT_i]-st\end{pmatrix}.$$

Combining the known asymptotic properties of $\hat{c}_t$ and (3.5) yields the following asymptotic bias and variance of MTK:

$$\begin{aligned}\text{abias}\left\{\hat{c}_m(u,v)\right\} =& b_m(u,v) \\ \equiv& \frac{h^2\left(g^{(2,0)}+g^{(0,2)}\right)\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)}{2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \\ &+\boldsymbol{\theta}^\top\boldsymbol{B}\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)c(u,v),\end{aligned} \tag{3.6}$$

and

$$\text{avar}\left\{\hat{c}_m(u,v)\right\} = \frac{\sigma_m^2(u,v)}{nh^2} \equiv \frac{c(u,v)}{4\pi nh^2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}. \tag{3.7}$$

We establish the following asymptotic properties of the MTK.

Theorem 1. *Under Assumptions 1-5 in Appendix A, the MTK estimator $\hat{c}_m$ given by (2.4), with Gaussian kernels, is such that for any $(u,v)\in(0,1)^2$,*

$$\sqrt{nh^2}\left(\hat{c}_m(u,v)-c(u,v)-b_m(u,v)\right)\xrightarrow{d}\mathcal{N}\left(0,\sigma_m^2(u,v)\right),$$

where $b_m(u,v)$ and $\sigma_m^2(u,v)$ are given in (3.6) and (3.7) respectively.

Remark 5. *Theorem 1 indicates that the MTK $\hat{c}_m$ introduces a bias correction to the TKE $\hat{c}_t$. On the other hand, they share the same asymptotic variance. Interestingly, this asymptotic variance is shared by several other boundary-bias-corrected copula density estimators, such as the kernel estimator $\tilde{c}^{(\tau,1)}$ in Geenens et al. (2014) as well as the Beta kernel estimator and the Bernstein estimators considered in Janssen et al. (2014).*

We next explore the global properties of the MTK. Let $w:\mathcal{I}\rightarrow\{0\cup\mathcal{R}^+\}$ be some non-negative weight function. The weighted mean integrated squared error (MISE) for a generic estimator $\hat{c}$ is defined as

$$\text{mise}\left\{\hat{c}\right\} = \mathbb{E}\left\{\int_{\mathcal{I}}\left(\hat{c}(u,v)-c(u,v)\right)^2 w(u,v)\mathrm{d}u\mathrm{d}v\right\}. \tag{3.8}$$

We set $w(u,v)=\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))$ to ensure the integrability of the weighted MISE. In fact, the weighted MISE in the UV-domain with this particular weight function amounts to the unweighted MISE in the ST-domain. Wand et al. (1991) note that the good performance of density estimation in the transformed domain is usually retained upon its back-transformation to the original domain. This observation has been confirmed by many numerical experiments, including our simulations reported below.

Since the MISE equals integrated squared bias plus integrated variance, we can write the asymptotic MISE as

$$\text{amise}\{\hat{c}_m\} = \frac{h^4}{4}\Gamma_3 + h^2\boldsymbol{\theta}^\top\boldsymbol{\Gamma}_2 + \boldsymbol{\theta}^\top\boldsymbol{\Gamma}_1\boldsymbol{\theta} + \frac{1}{4\pi n h^2}, \tag{3.9}$$

where

$$\begin{aligned}
\boldsymbol{\Gamma}_1 &= \int_{\mathcal{R}^2} \boldsymbol{B}(s,t)\boldsymbol{B}(s,t)^\top g^2(s,t)\mathrm{d}s\mathrm{d}t \\
\boldsymbol{\Gamma}_2 &= \int_{\mathcal{R}^2} \boldsymbol{B}(s,t)g(s,t)\left(g^{(2,0)}(s,t) + g^{(0,2)}(s,t)\right)\mathrm{d}s\mathrm{d}t \\
\Gamma_3 &= \int_{\mathcal{R}^2} \left(g^{(2,0)}(s,t) + g^{(0,2)}(s,t)\right)^2 \mathrm{d}s\mathrm{d}t.
\end{aligned} \tag{3.10}$$

Suppose for now that $\Gamma_3 - \boldsymbol{\Gamma}_2^\top\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2 > 0$. The optimal smoothing parameters that minimize (3.9) are then given by

$$h_{0,m} = \left[\frac{1}{2\pi(\Gamma_3 - \boldsymbol{\Gamma}_2^\top\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2)}\right]^{1/6} n^{-1/6}, \tag{3.11}$$

and

$$\boldsymbol{\theta}_{0,m} = -\frac{h_{0,m}^2}{2}\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2. \tag{3.12}$$

It follows that the optimal asymptotic MISE is obtained at

$$\text{amise}_0\{\hat{c}_m\} = \frac{1}{4}(2\pi)^{-2/3}(\Gamma_3 - \boldsymbol{\Gamma}_2^\top\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2)^{1/3} n^{-2/3}. \tag{3.13}$$

We also establish the follows:

Proposition 1. *Let $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$ and Γ_3 be given as in* (3.10)*, then $\Gamma_3 - \boldsymbol{\Gamma}_2^\top\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2 \geq 0$.*

In Section 5 we shall show that the equality holds when the underlying copula is Gaussian, implying higher order improvement. Furthermore, inspection of the optimal smoothing parameters (3.11) and (3.12) indicates that they satisfy Assumptions 4 and 5 for Theorem 1. As a result, the optimal MISE (3.13) is obtainable by plugging proper estimates of $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$ and Γ_3 into (3.11) and (3.12).

The corresponding results for the TKE—a special case of the MTK with $\boldsymbol{\theta} = \mathbf{0}$—are readily obtained as

$$\begin{aligned}
h_{0,t} &= \left[\frac{1}{2\pi\Gamma_3}\right]^{1/6} n^{-1/6}, \\
\text{amise}_0\{\hat{c}_t\} &= \frac{1}{4}(2\pi)^{-2/3}\Gamma_3^{1/3} n^{-2/3}.
\end{aligned} \tag{3.14}$$

Remark 6. *Since $\boldsymbol{\Gamma}_1$, by construction, is positive-semidefinite, $\boldsymbol{\Gamma}_2^\top\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2 \geq 0$. Therefore, compar-*

ison of (3.13) *and* (3.14) *suggests that the MTK dominates the TKE in terms of the asymptotic MISE* (3.9). *Since the two estimators share the same asymptotic variance, it is understood that the reduction in the asymptotic MISE is due to the bias correction of the MTK.*

4 Smoothing parameter selection

The MTK requires the selection of the usual bandwidth h plus the tuning parameters $\boldsymbol{\theta}$. It is well known that the selection of smoothing parameters plays a critical role in kernel density estimation. In this section, we present two methods of selecting optimal smoothing parameters.

4.1 Plug in method

We have derived in the previous section the optimal $h_{0,m}$ and $\boldsymbol{\theta}_{0,m}$ that minimize the asymptotic MISE (3.9). A logical way to proceed is to adopt a plug in bandwidth selector that replaces the unknown quantities in the optimal theoretical smoothing parameters with their sample analogs. This requires estimating $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$ and Γ_3 given in (3.10).

The estimation of $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ is relatively straightforward. Note that

$$
\begin{aligned}
\boldsymbol{\Gamma}_1 &= \mathbb{E}\left(\boldsymbol{B}(S_i,T_i)\boldsymbol{B}(S_i,T_i)^\top g(S_i,T_i)\right), \\
\boldsymbol{\Gamma}_2 &= \mathbb{E}\left(\boldsymbol{B}(S_i,T_i)(g^{(2,0)}(S_i,T_i)+g^{(0,2)}(S_i,T_i))\right).
\end{aligned}
$$

It follows that they can be readily estimated by

$$
\hat{\boldsymbol{\Gamma}}_1 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{B}\left(\hat{S}_i,\hat{T}_i\right)\boldsymbol{B}^\top\left(\hat{S}_i,\hat{T}_i\right)K_b\left(\hat{S}_i-\hat{S}_j\right)K_b\left(\hat{T}_i-\hat{T}_j\right) \tag{4.1}
$$

and

$$
\begin{aligned}
\hat{\boldsymbol{\Gamma}}_2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{B}\left(\hat{S}_i,\hat{T}_i\right)&\left\{K_b^{(2)}\left(\hat{S}_i-\hat{S}_j\right)K_b\left(\hat{T}_i-\hat{T}_j\right)\right. \\
&\left.+K_b\left(\hat{S}_i-\hat{S}_j\right)K_b^{(2)}\left(\hat{T}_i-\hat{T}_j\right)\right\},
\end{aligned} \tag{4.2}
$$

where b is a preliminary bandwidth, $K_b(x)=K(x/b)/b$ and $K_b^{(r)}(x)=\mathrm{d}^rK_b(x)/\mathrm{d}x^r$.

The estimation of quantities like Γ_3 has been studied in Wand and Jones (1995) and Duong and Hazelton (2003). Define $\psi_{r_1,r_2}=\int_{\mathcal{R}^2}g^{(r_1,r_2)}(s,t)g(s,t)\mathrm{d}s\mathrm{d}t$. We can write Γ_3 as

$$
\Gamma_3=\psi_{4,0}+\psi_{0,4}+2\psi_{2,2}.
$$

This decomposition is based on the fact that

$$
\begin{aligned}
&\int_{\mathcal{R}^2} g^{(r_1,r_2)}(s,t) g^{(r_1',r_2')}(s,t) \mathrm{d}s\mathrm{d}t \\
&= \begin{cases} (-1)^{r_1+r_2}\psi_{r_1+r_1',r_2+r_2'}, & \text{if } \sum_{i=1,2} r_i + r_i' \text{ is even} \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
$$

given that the density g is sufficiently smooth (see, e.g., Chapter 4 of Wand and Jones (1995) and references therein). Since ψ_{r_1,r_2} admits the expectation form $\mathbb{E}\left(g^{(r_1,r_2)}(S_i,T_i)\right)$, we can estimate it nonparametrically by

$$
\hat{\psi}_{r_1,r_2} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_b^{(r_1)}\left(\hat{S}_i - \hat{S}_j\right) K_b^{(r_2)}\left(\hat{T}_i - \hat{T}_j\right).
$$

Gathering the above individual components then yields the following estimator of Γ_3:

$$
\begin{aligned}
\hat{\Gamma}_3 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \Big\{ & K_b^{(4)}\left(\hat{S}_i - \hat{S}_j\right) K_b\left(\hat{T}_i - \hat{T}_j\right) \\
& +2K_b^{(2)}\left(\hat{S}_i - \hat{S}_j\right) K_b^{(2)}\left(\hat{T}_i - \hat{T}_j\right) + K_b\left(\hat{S}_i - \hat{S}_j\right) K_b^{(4)}\left(\hat{T}_i - \hat{T}_j\right) \Big\}.
\end{aligned} \tag{4.3}
$$

Equipped with the estimated $\hat{\mathbf{\Gamma}}_1$, $\hat{\mathbf{\Gamma}}_2$ and $\hat{\Gamma}_3$, we calculate the plug in smoothing parameters as

$$
\hat{h}_{0,m} = \left[\frac{1}{2\pi(\hat{\Gamma}_3 - \hat{\mathbf{\Gamma}}_2^\top \hat{\mathbf{\Gamma}}_1^{-1}\hat{\mathbf{\Gamma}}_2)}\right]^{1/6} n^{-1/6} \quad \text{and} \quad \hat{\boldsymbol{\theta}}_{0,m} = -\frac{\hat{h}_{0,m}^2}{2}\hat{\mathbf{\Gamma}}_1^{-1}\hat{\mathbf{\Gamma}}_2. \tag{4.4}
$$

We use the Gaussian product kernel in these estimations. In principle, we can derive separate optimal bandwidths for the estimation of $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$ and Γ_3. We note, however, using multiple bandwidths in this estimation may result in a negative $\hat{\Gamma}_3 - \hat{\mathbf{\Gamma}}_2^\top \hat{\mathbf{\Gamma}}_1^{-1}\hat{\mathbf{\Gamma}}_2$ (a similar observation is made by Duong and Hazelton (2003)). We therefore elect to use a single bandwidth to estimate $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$ and Γ_3. Since our experience indicates that Γ_3 is the most difficult to estimate among the three, we use a bandwidth that is optimal to the estimation of Γ_3 in our calculations. The construction of optimal bandwidth for quantities like Γ_3 can be found in Duong and Hazelton (2003).

4.2 Profile cross validation

We next present an alternative method to select the smoothing parameters for the MTK based on the principle of cross validation. The least square cross validation is commonly used in kernel density estimations. Given a non-negative weight function w, the weighted Cross Validation (CV) criterion for the MTK is defined as

$$\mathrm{CV} = \int_{\mathcal{I}} (c(u,v) - \hat{c}_m(u,v))^2 w(u,v)\mathrm{d}u\mathrm{d}v.$$

It can be shown that the CV criterion, net of a constant that is not affected by the estimation, can be consistently estimated by the following criterion:

$$\mathrm{CV}(h,\boldsymbol{\theta}) = \int_{\mathcal{I}} \left(\hat{c}_m(u,v)\right)^2 w(u,v)\mathrm{d}u\mathrm{d}v - \frac{2}{n}\sum_{i=1}^{n} \hat{c}_m^{(-i)}\left(\hat{U}_i,\hat{V}_i\right) w\left(\hat{U}_i,\hat{V}_i\right), \tag{4.5}$$

where $\hat{c}_m^{(-i)}\left(\hat{U}_i,\hat{V}_i\right)$ is the "leave-one-out" version of $\hat{c}_m$, calculated based on all observations except the ith one and evaluated at $\left(\hat{U}_i,\hat{V}_i\right)$. As discussed earlier, setting $w(u,v)=1$ leads to the unweighted cross validation in the UV-domain, while setting $w(u,v)=\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))$ amounts to the unweighted cross validation in the ST domain. Our numerical experiments indicate that the latter performs considerably better, underscoring the merit of smoothing parameters selection in the ST-domain. We therefore focus on the case of $w(u,v)=\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))$.

One advantage the cross validation has over the plug in method is that it does not require the estimation of complicated unknown population quantities. On the other hand, the optimization of the cross validation criterion can be numerically demanding, especially in the presence of multi-dimensional smoothing parameters. For the MTK, a direct implementation of the cross validation entails a three-dimensional optimization of a highly nonlinear objective function. To avoid potential numerical difficulty, we propose an alternative profile cross validation approach. Recall that the optimal tuning parameters are given by $\boldsymbol{\theta}_{0,m} = -\frac{h_{0,m}^2}{2}\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2$. Our strategy is to first calculate $\hat{\boldsymbol{\Gamma}}_1$ and $\hat{\boldsymbol{\Gamma}}_2$ and treat $\boldsymbol{\theta}_{0,m}$ as a function of h, taking $\hat{\boldsymbol{\Gamma}}_1$ and $\hat{\boldsymbol{\Gamma}}_2$ as given. Define the 'profiled' tuning parameters as $\boldsymbol{\theta}(h;\hat{\boldsymbol{\Gamma}}_1,\hat{\boldsymbol{\Gamma}}_2) = -\frac{h^2}{2}\hat{\boldsymbol{\Gamma}}_1^{-1}\hat{\boldsymbol{\Gamma}}_2$. We then plug $\boldsymbol{\theta}(h;\hat{\boldsymbol{\Gamma}}_1,\hat{\boldsymbol{\Gamma}}_2)$ into (4.5) and conduct the cross validation with respect to a single smoothing parameter h.

We note that the profile cross validation procedure effectively combines the strengths of the plug in method and cross validation. It reduces the dimension of numerical optimization required by the cross validation to one, at the same time it avoids the difficult task of estimating Γ_3, as is required in the plug in method.

5 Higher order improvement for Gaussian copulas

When the underlying copula is Gaussian, the transformed density g in the ST-domain becomes the bivariate Gaussian density. Denote by ρ the correlation coefficient between S and T. Suppose that $|\rho| < 1$. Setting $\boldsymbol{\theta} = -\frac{h^2}{2}\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2$ as a function of h, according to (3.12), yields

$$\theta_1 = \frac{1+\rho^2}{(1-\rho^2)^2}\frac{h^2}{2} \quad \text{and} \quad \theta_2 = -\frac{4\rho}{(1-\rho^2)^2}\frac{h^2}{2}. \tag{5.1}$$

We can then show that the first order asymptotic bias $b_m(u,v)$, given in (3.6), vanishes for any $(u,v) \in (0,1)^2$ regardless of the value of h. Consequently under Gaussian copulas, the MTK provides higher order bias reduction. To see this, first note that the asymptotic bias of the TKE, up to the second order as given in Charpentier et al. (2006), has the form

$$\begin{aligned} \text{abias}\left\{\hat{c}_t(u,v)\right\} = b_t^{(G)}(u,v) \equiv & \frac{h^2\left(g^{(2,0)}+g^{(0,2)}\right)\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)}{2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \\ & + \frac{h^4\left(g^{(4,0)}+g^{(0,4)}+2g^{(2,2)}\right)\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)}{8\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}. \end{aligned} \tag{5.2}$$

Similarly, a second order Taylor expansion of $J(u,v;h,\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ at $\mathbf{0}$ yields

$$J(u,v;h,\boldsymbol{\theta}) \approx 1 + \boldsymbol{\theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u),\Phi^{-1}(v)\right) + h^2\boldsymbol{\theta}^\top\boldsymbol{A}_1 + \frac{1}{2}\boldsymbol{\theta}^\top\boldsymbol{A}_2\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)\boldsymbol{\theta}, \tag{5.3}$$

where

$$\boldsymbol{A}_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

and

$$\boldsymbol{A}_2(s,t) = \begin{pmatrix} (s^2+t^2-2)^2 - 4(1+\rho^2) & (st-\rho)(s^2+t^2-2) - 4\rho \\ (st-\rho)(s^2+t^2-2) - 4\rho & (st-\rho)^2 - (\rho^2+1) \end{pmatrix}.$$

Using that $\hat{c}_m(u,v) = J(u,v;h,\boldsymbol{\theta})\hat{c}_t(u,v)$ and plugging (5.1) into the above second order expansions, we obtain, after tedious yet straightforward algebraic manipulations, that under Gaussian copulas

$$\text{abias}\left\{\hat{c}_m(u,v)\right\} = b_m^{(G)}(u,v) \equiv h^4 R\left(\Phi^{-1}(u),\Phi^{-1}(v);\rho\right), \tag{5.4}$$

where

$$R(s,t;\rho) = c(\Phi(s),\Phi(t)) \left\{ \frac{-(1+3\rho^2)(s^2+t^2) + 2\rho(\rho^2+3)st + 2(1-\rho^4)}{2(1-\rho^2)^3} \right\}.$$

Thus the MTK reduces the asymptotic bias from order $O(h^2)$ to $O(h^4)$ under Gaussian copulas. Its asymptotic variance, which is shown to be identical to that of the TKE in Section 3, remains the same. The asymptotic properties of the MTK under Gaussian copulas is given formally as follows.

Theorem 2. *Suppose that the underlying copula is Gaussian with* $|\rho| < 1$ *and* $\boldsymbol{\theta}$ *is specified according to* (5.1)*, under Assumptions 1, 4 and 6 in Appendix A, the MTK estimator* $\hat{c}_m$ *is such that for any* $(u,v) \in (0,1)^2$*,*

$$\sqrt{nh^2}\left(\hat{c}_m(u,v) - c(u,v) - b_m^{(G)}(u,v)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_m^2(u,v)\right),$$

where $b_m^{(G)}(u,v)$ *and* $\sigma_m^2(u,v)$ *are given in* (5.4) *and* (3.7) *respectively.*

Note that Assumptions 2 and 3 are not required for Theorem 2 since they are satisfied by Gaussian copulas automatically; see, e.g., Equation (9) of Omelka et al. (2009).

We next examine the global properties of the MTK under Gaussian copulas, in the same token as that under general copulas in Section 3. To save space, the derivation details are not reported; they are available from the authors upon request. We establish that the asymptotic MISE, as defined in (3.8), is given by

$$\text{amise}^{(G)}\{\hat{c}_m\} = h^8\Lambda + \frac{1}{4\pi nh^2}, \tag{5.5}$$

where $\Lambda = \int_{\mathcal{R}^2} R^2(s,t;\rho)\phi^2(s)\phi^2(t)\mathrm{d}s\mathrm{d}t$. The optimal bandwidth that minimizes (5.5) is given by

$$h_{0,m}^{(G)} = (16\pi\Lambda)^{-1/10}\, n^{-1/10}, \tag{5.6}$$

which obviously satisfies the assumptions of Theorem 2. Plugging (5.1) and (5.6) into (5.5) then yields

$$\text{amise}_0^{(G)}\{\hat{c}_m\} = \frac{5}{16\pi}(16\pi\Lambda)^{1/5}\, n^{-4/5}.$$

Thus under Gaussian copulas, the MISE convergence rate of $\hat{c}_m$ is improved to $O(n^{-4/5})$, faster than the usual $O(n^{-2/3})$ rate for bivariate densities.

The two methods of smoothing parameter selection proposed in the previous section can be applied to the present case of Gaussian copulas. There exist, however, some noteworthy differences compared with the general case. The plug in method is based on the first order asymptotic approximation and therefore ceases to be optimal, because the optimal MTK converges at a faster rate under Gaussian copulas. Nevertheless, this method remains a viable choice in practice as is suggested by our numerical experiments. In principle, we can derive optimal smoothing parameters based on higher order asymptotic expansions for the Gaussian case. This approach, however, entails nonparametric estimations of unknown population moments and derivatives up to the fourth order and is practically not quite appealing.

Unlike the plug in method, the cross validation approach obtains its optimality regardless whether the underlying copula is Gaussian. This is because the CV criterion, rather than relying on explicit asymptotic approximations, consistently estimates the MISE and therefore automatically adapts to the unknown convergence rate of an estimator (see, e.g., Stone (1984) for the general optimality of the cross validation under rather mild conditions). Furthermore since it is derived based on the relationship $\boldsymbol{\theta} = -\frac{h^2}{2}\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2$ for an arbitrary h (rather than a bandwidth with a particular order), the profile CV method presented in the previous section is as adaptive as the full CV method. Therefore the profile CV method is preferred to the plug in method when the underlying copulas are Gaussian. This conjecture is confirmed by our numerical experiments with Gaussian and near-Gaussian copulas.

Remark 7. *The improved convergence rate of the MTK under Gaussian copulas has important practical ramifications. Gaussian copulas have been widely used in financial risk management and portfolio optimization. This approach, however, has been roundly criticized for its restrictiveness such as symmetry, thin tails and lack of tail dependence. An extensive body of literature has established that many financial data, although unlikely exactly Gaussian, are near Gaussian with idiosyncrasies such as asymmetry, fat tails and non-zero tail dependence, etc (Ang and Chen, 2002; Chen et al., 2004; Hong et al., 2007; Longin and Solnik, 2001; Zimmer, 2012). Clearly any single parametric copula is inadequate to capture these multitudinous aberrants from Gaussianity. Such a situation calls for flexible nonparametric copula estimators. The MTK is free of the severe boundary biases that plague conventional kernel estimators and provides improved convergence rate for Gaussian copulas. By virtue of contiguity, we expect it to be a competent estimator for near-Gaussian copulas as well. Given the similarity between the proposed estimator and a kernel estimator with an initial Gaussian copula estimate, we note that the improved convergence can also be interpreted using the framework of nonparametric estimation with a parametric start by Hjort and Glad (1995).*

6 Generalization to bivariate kernels

We have so far confined our discussion to the case of product kernels. It is sometimes desirable to entertain bivariate kernels with a non-diagonal bandwidth matrix $\boldsymbol{H}$, whose off-diagonal elements control the direction towards which the smoothing is placed (Duong and Hazelton, 2005; Geenens et al., 2014). In this section we consider the more general case of bivariate Gaussian kernels with a bandwidth matrix

$$\boldsymbol{H} = h^2 \begin{pmatrix} 1 & \lambda \\ \lambda & 1 \end{pmatrix}, \tag{6.1}$$

where $|\lambda| < 1$. The MTK under this bandwidth matrix takes the form

$$\begin{aligned} \tilde{c}_m(u,v) =& \frac{\exp\left(-\theta_1[\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2] - \theta_2\Phi^{-1}(u)\Phi^{-1}(v)\right)}{n\eta h^2 \phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)} \\ & \sum_{i=1}^n \phi_{(\lambda)}\left(\frac{\Phi^{-1}(u) - \hat{S}_i}{h}, \frac{\Phi^{-1}(v) - \hat{T}_i}{h}\right), \end{aligned} \tag{6.2}$$

where $\phi_{(\lambda)}$ is a bivariate Gaussian density defined as

$$\phi_{(\lambda)}(s,t) = \frac{1}{2\pi\sqrt{1-\lambda^2}} \exp\left(-\frac{s^2 + t^2 - 2\lambda st}{2(1-\lambda^2)}\right).$$

The corresponding normalization term η is then given by

$$\eta = \frac{1}{n\delta}\sum_{i=1}^n \exp\left\{-\frac{(4h^2\theta_1^2 - h^2\theta_2^2 + 2\theta_1)(\hat{S}_i^2 + \hat{T}_i^2) + (2\lambda h^2\theta_2^2 - 8\lambda h^2\theta_1^2 + 2\theta_2)\hat{S}_i\hat{T}_i}{2\delta^2}\right\}$$

with

$$\delta = \sqrt{h^4(1-\lambda^2)(4\theta_1^2 - \theta_2^2) + 2h^2(2\theta_1 + \lambda\theta_2) + 1}.$$

Similarly to those under simplified bandwidth configuration in Section 2, the asymptotic bias and variance of the $\tilde{c}_m$ for any $(u,v) \in (0,1)^2$ are derived as follows:

$$\begin{aligned} \text{abias}\left\{\tilde{c}_m(u,v)\right\} =& \frac{h^2\left(g^{(2,0)} + g^{(0,2)} + 2\lambda g^{(1,1)}\right)\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)}{2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \\ & + \boldsymbol{\theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) c(u,v) \end{aligned} \tag{6.3}$$

and

$$\operatorname{avar}\left\{\tilde{c}_m(u,v)\right\} = \frac{c(u,v)}{4\pi n h^2\sqrt{1-\lambda^2}\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}. \tag{6.4}$$

Remark 8. *Denote the TKE under the bivariate Gaussian kernel with bandwidth matrix* (6.1) *by* $\tilde{c}_t$*. As in the case of product kernels, the first term of* (6.3) *is the same as the asymptotic bias of* $\tilde{c}_t$ *and the second term attempts a bias correction to the first term; furthermore,* $\tilde{c}_t$ *and* $\tilde{c}_m$ *share the same asymptotic variance* (6.4).

As is defined in (3.8), the asymptotic weighted MISE of $\tilde{c}_m$ is given by

$$\operatorname{amise}\left\{\tilde{c}_m\right\} = \frac{h^4}{4}\Gamma_3(\lambda) + h^2\boldsymbol{\theta}^\top\boldsymbol{\Gamma}_2(\lambda) + \boldsymbol{\theta}^\top\boldsymbol{\Gamma}_1\boldsymbol{\theta} + \frac{1}{4\pi n h^2\sqrt{1-\lambda^2}}, \tag{6.5}$$

where $\boldsymbol{\Gamma}_1$ is again given by (3.10) and

$$\begin{aligned}
\boldsymbol{\Gamma}_2(\lambda) &= \int_{\mathcal{R}^2}\boldsymbol{B}(s,t)g(s,t)\left(g^{(2,0)}(s,t)+g^{(0,2)}(s,t)+2\lambda g^{(1,1)}(s,t)\right)\mathrm{d}s\mathrm{d}t\\
\Gamma_3(\lambda) &= \int_{\mathcal{R}^2}\left(g^{(2,0)}(s,t)+g^{(0,2)}(s,t)+2\lambda g^{(1,1)}(s,t)\right)^2\mathrm{d}s\mathrm{d}t.
\end{aligned}$$

In parallel to (3.11) and (3.12), the optimal smoothing parameters that minimize the asymptotic MISE (6.5) are given by, for a given value of λ,

$$\begin{aligned}
h_{0,m}(\lambda) &= \left[\frac{1}{2\pi\sqrt{1-\lambda^2}\left(\Gamma_3(\lambda)-\boldsymbol{\Gamma}_2(\lambda)^\top\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2(\lambda)\right)}\right]^{1/6} n^{-1/6}\\
\boldsymbol{\theta}_{0,m}(\lambda) &= -\frac{h_{0,m}^2(\lambda)}{2}\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2(\lambda),
\end{aligned}$$

provided that $\Gamma_3(\lambda)-\boldsymbol{\Gamma}_2(\lambda)^\top\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2(\lambda)>0$. The first order condition with respect to λ is complicated but can be reduced to the following minimization problem

$$\begin{aligned}
&\lambda_0 = \arg\min_{\lambda}\frac{\Gamma_3(\lambda)-\boldsymbol{\Gamma}_2(\lambda)^\top\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2(\lambda)}{1-\lambda^2}\\
&\text{subject to } |\lambda|<1.
\end{aligned} \tag{6.6}$$

It follows that the optimal asymptotic MISE of $\tilde{c}_m$ takes the form

$$\operatorname{amise}_0\left\{\tilde{c}_m\right\} = \frac{1}{4}(2\pi)^{-2/3}\left(\frac{\Gamma_3(\lambda_0)-\boldsymbol{\Gamma}_2(\lambda_0)^\top\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2(\lambda_0)}{1-\lambda_0^2}\right)^{1/3} n^{-2/3}.$$

Since $\text{amise}_0 \{\hat{c}_m\}$ specified in (3.13) is a special case of $\text{amise}_0 \{\tilde{c}_m\}$ under the restriction $\lambda = 0$, $\text{amise}_0 \{\hat{c}_m\} \geq \text{amise}_0 \{\tilde{c}_m\}$, implying that $\tilde{c}_m$ dominates $\hat{c}_m$ asymptotically.

Remark 9. *Similar to the case of simplified bandwidth discussed in the previous sections, when the underlying copula is Gaussian, setting* $\boldsymbol{\theta} = -\frac{h^2}{2}\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2(\lambda)$, *for any* $\lambda \in (-1,1)$, *reduces the asymptotic bias of* $\tilde{c}_m$ *from order* $O(h^2)$ *to* $O(h^4)$. *Thus under Gaussian copulas, the optimal bandwidth* $h_{0,m}^{(G)}(\lambda) \sim n^{-1/10}$ *and the resultant optimal asymptotic MISE,* $\text{amise}_0^{(G)} \{\tilde{c}_m\}$, *is improved to* $O(n^{-4/5})$.

Introducing the additional parameter λ complicates the selection of smoothing parameters. First of all, we need to estimate $\boldsymbol{\Gamma}_2(\lambda)$ and $\Gamma_3(\lambda)$ for a given λ (note that $\boldsymbol{\Gamma}_1$ does not depend on λ). Generalizing (4.2), we can estimate $\boldsymbol{\Gamma}_2(\lambda)$ by

$$\hat{\boldsymbol{\Gamma}}_2(\lambda) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{B}\left(\hat{S}_i, \hat{T}_i\right)\left\{K_b^{(2)}\left(\hat{S}_i - \hat{S}_j\right)K_b\left(\hat{T}_i - \hat{T}_j\right)\right.$$
$$\left. + K_b\left(\hat{S}_i - \hat{S}_j\right)K_b^{(2)}\left(\hat{T}_i - \hat{T}_j\right) + 2\lambda K_b^{(1)}\left(\hat{S}_i - \hat{S}_j\right)K_b^{(1)}\left(\hat{T}_i - \hat{T}_j\right)\right\}. \tag{6.7}$$

Also note that $\Gamma_3(\lambda)$ can be decomposed to

$$\Gamma_3(\lambda) = \psi_{4,0} + \psi_{0,4} + (4\lambda^2 + 2)\psi_{2,2} + 4\lambda\psi_{3,1} + 4\lambda\psi_{1,3}.$$

It follows that this quantity can be estimated by

$$\hat{\Gamma}_3(\lambda) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{K_b^{(4)}\left(\hat{S}_i - \hat{S}_j\right)K_b\left(\hat{T}_i - \hat{T}_j\right)\right.$$
$$+ (4\lambda^2 + 2)K_b^{(2)}\left(\hat{S}_i - \hat{S}_j\right)K_b^{(2)}\left(\hat{T}_i - \hat{T}_j\right) + K_b\left(\hat{S}_i - \hat{S}_j\right)K_b^{(4)}\left(\hat{T}_i - \hat{T}_j\right) \tag{6.8}$$
$$\left. + 4\lambda K_b^{(3)}\left(\hat{S}_i - \hat{S}_j\right)K_b^{(1)}\left(\hat{T}_i - \hat{T}_j\right) + 4\lambda K_b^{(1)}\left(\hat{S}_i - \hat{S}_j\right)K_b^{(3)}\left(\hat{T}_i - \hat{T}_j\right)\right\}.$$

In keeping with our first estimator, we use the Gaussian kernel and employ the same preliminary bandwidth b suggested by Duong and Hazelton (2003) in the estimation of $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2(\lambda)$ and $\Gamma_3(\lambda)$.

We next describe how to implement the two methods of smoothing parameter selection for the more general estimator $\tilde{c}_m$. Our strategy is to first solve for the optimal λ according to (6.6). Denote the solution by $\tilde{\lambda}_0$. The plug-in smoothing parameters are then calculated as

$$\tilde{h}_{0,m} = \left\{2\pi\sqrt{1 - \tilde{\lambda}_0^2}\left(\hat{\Gamma}_3(\tilde{\lambda}_0) - \hat{\boldsymbol{\Gamma}}_2(\tilde{\lambda}_0)^\top\hat{\boldsymbol{\Gamma}}_1^{-1}\hat{\boldsymbol{\Gamma}}_2(\tilde{\lambda}_0)\right)\right\}^{-1/6} n^{-1/6},$$

$$\tilde{\boldsymbol{\theta}}_{0,m} = -\frac{1}{2}\tilde{h}_{0,m2}^2\hat{\boldsymbol{\Gamma}}_1^{-1}\hat{\boldsymbol{\Gamma}}_2(\tilde{\lambda}_0).$$

The profile cross validation approach can be similarly adapted. Equipped with an estimated $\tilde{\lambda}_0$, we set $\boldsymbol{\theta}_{0,m}(h) = -\frac{1}{2}h^2\hat{\boldsymbol{\Gamma}}_1^{-1}\hat{\boldsymbol{\Gamma}}_2(\tilde{\lambda}_0)$. We then plug this quantity into (4.5) and conduct the cross validation with respect to h. To facilitate the implementation of this procedure, we provide in Appendix C an analytical form of the first integral term in (4.5) for the MTK $\tilde{c}_m$ with a non-diagonal bandwidth matrix.

7 Monte Carlo simulation

We use simulations to examine the finite sample performance of the proposed estimators and to compare them with a number of alternative estimators. In particular, we include in our numerical investigations the following estimators:

- The proposed MTK $\hat{c}_m$ with a product kernel and $\tilde{c}_m$ with a bivariate kernel. Either estimator is calculated using the plug-in and profile cross validation methods of smoothing parameter selection.

- The conventional TKE $\hat{c}_t$ with a product kernel and $\tilde{c}_t$ with a bivariate kernel. Either estimator is calculated using the plug-in and profile cross validation methods of smoothing parameter selection as in the case of MTK and under the restriction $\boldsymbol{\theta} = \mathbf{0}$.

- Gijbels and Mielniczuk (1990)'s mirror reflection estimator $\hat{c}_r$. An initial bandwidth is obtained via the rule of thumb based on the augmented dataset, i.e. $9n$ observations mirror-reflected from the original data. The final bandwidth is obtained by multiplying the former by $(1/9)^{2/3} \approx 0.23$ to adjust for the effective sample size and range.

- The Beta kernel estimator studied in Charpentier et al. (2006), which adopts the bias correction suggested by Chen (1999). Since method of optimal smoothing parameter is not available, we follow Geenens et al. (2014) and consider two bandwidths: $h = 0.02$ and $h = 0.05$. The corresponding estimators are denoted by $\hat{c}_{b1}$ and $\hat{c}_{b2}$ respectively.

- The penalized hierarchical B-splines estimator of Kauermann et al. (2013). As in their study, the parameters d and D are set to 4 and 8 and the vector of penalty coefficients are set at three different levels: $\lambda = (10, 10), (100, 100)$, and $(1000, 1000)$. The corresponding estimators are denoted by $\hat{c}_{p1}$, $\hat{c}_{p2}$ and $\hat{c}_{p3}$ respectively.

We include in our simulations a wide range of copulas, which are separated into two groups. The first group includes some commonly used parametric copulas and the second group focuses on Gaussian and near-Gaussian copulas. For each copula distribution, we consider two levels of dependence with Kendall's τ being 0.3 and 0.6 respectively. Their specifics are as follows:

Group One

(A) The Student t-copula with 5 degrees of freedom, with parameters $\rho = 0.454$ and $\rho = 0.809$.

(B) The Frank copula, with parameters $\theta = 2.92$ and $\theta = 7.93$.

(C) The Gumbel copula, with parameters $\theta = 10/7$ and $\theta = 2.5$.

(D) The Clayton copula, with parameters $\theta = 6/7$ and $\theta = 3$.

Group Two

(E) The Gaussian copula, with parameters $\rho = 0.454$ and $\rho = 0.809$.

(F) A mixture of 85% Gaussian copula and 15% Clayton copula with two pairs of parameters $\big(\rho = 0.454, \theta = 6/7\big)$ and $(\rho = 0.809, \theta = 3)$. These mixed copulas are asymmetric with higher dependence in the lower tail than in the upper tail.

(G) The Student t-copula with 15 degrees of freedom, with parameters $\rho = 0.454$ and $\rho = 0.809$. These copulas are near Gaussian with fat tails.

(H) A mixture of 85% Student t-copula with 15 degrees of freedom and 15% Clayton copula with two pairs of parameters $\big(\rho = 0.454, \theta = 6/7\big)$ and $(\rho = 0.809, \theta = 3)$. These mixed copulas are near Gaussian but asymmetric and having fat tails.

We consider two sample sizes $n = 500$ and $n = 2000$ and repeat each experiment 1,000 times. Pseudo-data calculated from the simulated data are used in all estimations. For a generic estimator $\hat{c}$, we evaluate its global performance by the integrated squared error (ISE) and the integrated absolute error (IAE) evaluated on an equally spaced grid on $[0, 1]^2$ as follows

$$\begin{aligned}
\mathrm{ISE}(\hat{c}, c) &\approx \frac{1}{(\#\mathcal{G})^2} \sum_{k \in \mathcal{G}} \sum_{l \in \mathcal{G}} \big(\hat{c}(k, l) - c(k, l)\big)^2 \\
\mathrm{IAE}(\hat{c}, c) &\approx \frac{1}{(\#\mathcal{G})^2} \sum_{k \in \mathcal{G}} \sum_{l \in \mathcal{G}} \big|\hat{c}(k, l) - c(k, l)\big| ,
\end{aligned} \tag{7.1}$$

where $\mathcal{G} = \{0.01, 0.02, \cdots, 0.99\}$ and $\#\mathcal{G}$ denotes the cardinality of $\mathcal{G}$. We report the average ISE and the average IAE across 1,000 repetitions for each experiment in Tables 1 and 2 for $n = 500$ and $n = 2000$ respectively. We further examine the tail performance of the estimators. This is of practical importance since copulas are commonly used in the finance and insurance industries, which emphasize tail distributions and extreme events. We retain the two measures in (7.1) but confine the evaluation points $\mathcal{G}$ to the four corners of the unit square. Specifically, we set $\mathcal{G} = \{0.0005, 0.0015, 0.0025, \cdots, 0.0485, 0.0495\} \cup \{0.9505, 0.9515, 0.9525, \cdots, 0.9985, 0.9995\}$ and focus on the case of $n = 2000$ in our evaluation of tail performance. We report the average ISE and the average IAE across 1,000 repetitions in Table 3. To facilitate comparison, we use the boldface font and the underline font to indicate the minimum and the second minimum ISE's (or IAE's) respectively. Except for a few exceptions, the MTK estimators generally outperform the other estimators, oftentimes by considerable margins. Some observations on individual comparisons are in order.

- Overall, the MTK improves on the TKE substantially, corroborating its asymptotic dominance over the TKE as derived in the previous sections. Improvements in both global and tail performance are observed.

- Both the plug-in and the profile cross validation methods of smoothing parameter selection provide satisfactory performance.

- The MTK with a bivariate kernel generally outperforms that with a product kernel, especially for copulas with high dependence (those with Kendall's $\tau = 0.6$).

- For the Gaussian and near-Gaussian copulas in Group Two, the MTK clearly dominates its competitors. This is consistent with our theoretical analysis that the MTK provides higher order improvement. As is discussed above, the CV method is rate-adaptive and remains optimal under Gaussian copulas but not the plug-in method, which is derived based on the first order asymptotic approximation. This is also confirmed by our simulations in Group Two, wherein the CV method consistently outperforms the plug in method. (On the other hand, the plug in method seems to perform slightly better in Group One for general copulas, possibly because the convergence rate of the plug in method is usually faster than that of the CV; see, e.g., Wand and Jones (1995) for a general treatment of this subject.)

Table 1: Average ISE and IAE for $n = 500$

Copulas	$\hat{c}_m$		$\tilde{c}_m$		$\hat{c}_t$		$\tilde{c}_t$		$\hat{c}_r$	$\hat{c}_{b1}$	$\hat{c}_{b2}$	$\hat{c}_{p1}$	$\hat{c}_{p2}$	$\hat{c}_{p3}$
	plug in	cv	plug in	cv	plug in	cv	plug in	cv						
Panel 1: ISE														
A1	0.0264	0.0214	0.0245	**0.0208**	0.0453	0.0434	0.0423	0.0420	0.0634	0.0571	0.0411	0.0490	0.0658	0.0983
A2	0.0735	0.0654	0.0664	**0.0608**	0.1056	0.1055	0.0859	0.0777	0.2546	0.1190	0.2939	0.3867	0.5650	0.8142
B1	0.0195	0.0114	0.0193	0.0120	0.0384	0.0388	0.0390	0.0409	0.0451	0.0492	0.0160	0.0137	**0.0072**	0.0170
B2	0.0552	0.0430	0.0512	**0.0383**	0.0845	0.0818	0.0750	0.0707	0.0466	0.0488	0.0557	0.0448	0.1421	0.3488
C1	0.0313	0.0332	**0.0295**	0.0332	0.0443	0.0459	0.0423	0.0425	0.0779	0.0586	0.0554	0.0685	0.0875	0.1167
C2	0.1304	0.1657	**0.1078**	0.1441	0.1507	0.1644	0.1135	0.1216	0.4139	0.2403	0.4618	0.5686	0.7450	0.9877
D1	0.0367	0.0471	**0.0344**	0.0481	0.0452	0.0485	0.0424	0.0445	0.0921	0.0607	0.0723	0.0940	0.1241	0.1687
D2	0.3956	0.4404	0.3084	0.3531	0.3977	0.3986	**0.2681**	0.2871	1.0053	0.7100	1.0992	1.2693	1.4918	1.7463
E1	0.0189	0.0087	0.0181	**0.0083**	0.0367	0.0361	0.0366	0.0363	0.0492	0.0501	0.0208	0.0214	0.0208	0.0314
E2	0.0524	0.0280	0.0499	**0.0255**	0.0799	0.0707	0.0720	0.0582	0.1210	0.0570	0.1347	0.1733	0.3106	0.5336
F1	0.0198	**0.0095**	0.0188	0.0098	0.0388	0.0363	0.0372	0.0366	0.0509	0.0511	0.0235	0.0249	0.0261	0.0382
F2	0.0613	0.0440	0.0588	**0.0421**	0.0902	0.0850	0.0776	0.0690	0.1601	0.0827	0.1763	0.2212	0.3611	0.5837
G1	0.0200	0.0109	0.0187	**0.0104**	0.0397	0.0385	0.0389	0.0376	0.0514	0.0519	0.0252	0.0272	0.0311	0.0471
G2	0.0546	0.0351	0.0518	**0.0327**	0.0858	0.0790	0.0753	0.0644	0.1540	0.0700	0.1769	0.2319	0.3830	0.6132
H1	0.0210	**0.0120**	0.0205	0.0124	0.0395	0.0380	0.0395	0.0375	0.0538	0.0523	0.0276	0.0308	0.0359	0.0528
H2	0.0698	0.0555	0.0634	**0.0504**	0.0980	0.0963	0.0826	0.0758	0.1944	0.0985	0.2188	0.2777	0.4299	0.6587
Panel 2: IAE														
A1	0.0968	0.0839	0.0939	**0.0817**	0.1379	0.1433	0.1282	0.1361	0.1707	0.1612	0.1055	0.1137	0.1115	0.1637
A2	0.1270	0.1182	0.1218	**0.1154**	0.1744	0.1875	0.1519	0.1602	0.1913	0.1485	0.2038	0.2192	0.3551	0.5318
B1	0.0895	0.0721	0.0886	0.0724	0.1304	0.1381	0.1235	0.1304	0.1650	0.1593	0.0910	0.0886	**0.0626**	0.1041
B2	0.1339	0.1346	0.1264	**0.1243**	0.1740	0.1789	0.1514	0.1556	0.1449	0.1412	0.1657	0.1427	0.2845	0.4795
C1	0.0955	0.0823	0.0937	**0.0818**	0.1326	0.1422	0.1268	0.1343	0.1723	0.1583	0.1027	0.1117	0.1018	0.1318
C2	0.1397	0.1417	**0.1306**	0.1320	0.1781	0.1891	0.1541	0.1644	0.1954	0.1593	0.2073	0.2159	0.3428	0.5194
D1	0.1010	0.0983	0.0993	**0.0977**	0.1321	0.1450	0.1274	0.1380	0.1734	0.1564	0.1046	0.1158	0.1067	0.1444
D2	0.1918	0.2005	**0.1677**	0.1745	0.2139	0.2146	0.1789	0.1823	0.2315	0.1978	0.2564	0.2594	0.3768	0.5502
E1	0.0845	0.0584	0.0838	**0.0577**	0.1246	0.1343	0.1190	0.1264	0.1670	0.1563	0.0926	0.0976	0.0755	0.0884
E2	0.1085	**0.0792**	0.1078	0.0795	0.1519	0.1667	0.1376	0.1484	0.1625	0.1277	0.1542	0.1434	0.2749	0.4679
F1	0.0858	**0.0579**	0.0856	0.0596	0.1260	0.1335	0.1211	0.1282	0.1664	0.1575	0.0942	0.0989	0.0778	0.0926
F2	0.1122	**0.0868**	0.1112	0.0882	0.1596	0.1716	0.1401	0.1512	0.1684	0.1340	0.1640	0.1547	0.2846	0.4728
G1	0.0866	0.0629	0.0849	**0.0617**	0.1291	0.1376	0.1229	0.1299	0.1668	0.1590	0.0963	0.1018	0.0842	0.1086
G2	0.1119	**0.0861**	0.1090	0.0862	0.1581	0.1740	0.1426	0.1529	0.1696	0.1328	0.1687	0.1676	0.3012	0.4881
H1	0.0875	**0.0626**	0.0884	0.0645	0.1282	0.1369	0.1226	0.1296	0.1676	0.1584	0.0958	0.1031	0.0854	0.1103
H2	0.1188	0.0973	0.1137	**0.0947**	0.1643	0.1764	0.1462	0.1549	0.1755	0.1395	0.1770	0.1766	0.3079	0.4909

Note: A1 and A2 refer to copula specification A with Kendall's τ being 0.3 and 0.6 respectively; other copulas are similarly denoted.

Table 2: Average ISE and IAE for $n = 2000$

Copulas	$\hat{c}_m$		$\tilde{c}_m$		$\hat{c}_t$		$\tilde{c}_t$		$\hat{c}_r$	$\hat{c}_{b1}$	$\hat{c}_{b2}$	$\hat{c}_{p1}$	$\hat{c}_{p2}$	$\hat{c}_{p3}$
	plug in	cv	plug in	cv	plug in	cv	plug in	cv						
Panel 1: ISE														
A1	0.0121	<u>0.0111</u>	0.0114	**0.0102**	0.0197	0.0191	0.0184	0.0177	0.0305	0.0171	0.0294	0.0306	0.0460	0.0781
A2	0.0343	0.0320	<u>0.0305</u>	**0.0278**	0.0506	0.0500	0.0392	0.0357	0.1767	0.0912	0.2863	0.6752	0.7056	0.6572
B1	0.0087	0.0057	0.0086	0.0064	0.0167	0.0166	0.0171	0.0175	0.0188	0.0123	<u>0.0052</u>	0.0075	**0.0029**	0.0084
B2	0.0271	0.0222	0.0234	<u>0.0195</u>	0.0393	0.0385	0.0333	0.0313	0.0205	**0.0187**	0.0481	0.4246	0.2170	0.2129
C1	0.0151	0.0177	**0.0140**	<u>0.0174</u>	0.0204	0.0205	0.0187	0.0184	0.0421	0.0220	0.0445	0.0486	0.0681	0.0986
C2	0.0689	0.0880	<u>0.0558</u>	0.0717	0.0774	0.0808	**0.0535**	0.0568	0.3243	0.2111	0.4528	0.8612	0.7912	0.8343
D1	<u>0.0178</u>	0.0245	**0.0172**	0.0238	0.0210	0.0219	0.0189	0.0193	0.0525	0.0251	0.0621	0.0685	0.0978	0.1399
D2	0.2175	0.2207	0.1531	0.1596	0.2207	0.2050	**0.1289**	<u>0.1346</u>	0.8682	0.6781	1.0895	1.5953	1.4621	1.5885
E1	0.0080	<u>0.0037</u>	0.0078	**0.0034**	0.0162	0.0155	0.0158	0.0154	0.0212	0.0135	0.0104	0.0142	0.0131	0.0231
E2	0.0236	<u>0.0129</u>	0.0224	**0.0105**	0.0374	0.0334	0.0320	0.0267	0.0699	0.0299	0.1280	0.3275	0.2653	0.3889
F1	0.0083	**0.0042**	0.0080	**0.0042**	0.0164	0.0159	0.0159	0.0155	0.0226	0.0139	0.0128	0.0166	0.0174	0.0291
F2	0.0298	<u>0.0245</u>	0.0281	**0.0219**	0.0441	0.0427	0.0366	0.0324	0.1029	0.0540	0.1690	0.4451	0.3291	0.4396
G1	0.0087	<u>0.0047</u>	0.0082	**0.0045**	0.0172	0.0164	0.0167	0.0159	0.0230	0.0141	0.0142	0.0156	0.0204	0.0361
G2	0.0258	<u>0.0156</u>	0.0240	**0.0138**	0.0404	0.0384	0.0342	0.0289	0.0964	0.0425	0.1687	0.4787	0.4176	0.4647
H1	0.0090	**0.0056**	0.0085	<u>0.0057</u>	0.0172	0.0166	0.0166	0.0158	0.0245	0.0147	0.0169	0.0187	0.0246	0.0414
H2	0.0336	<u>0.0288</u>	0.0299	**0.0260**	0.0479	0.0474	0.0380	0.0343	0.1314	0.0697	0.2110	0.5573	0.4704	0.5113
Panel 2: IAE														
A1	0.0646	<u>0.0604</u>	0.0625	**0.0564**	0.0903	0.0931	0.0849	0.0869	0.1122	0.0826	0.0699	0.0814	0.0774	0.1290
A2	0.0841	0.0825	<u>0.0801</u>	**0.0763**	0.1172	0.1213	0.0999	0.1028	0.1352	0.1054	0.1978	0.3539	0.4059	0.4252
B1	0.0599	0.0501	0.0584	<u>0.0493</u>	0.0874	0.0897	0.0823	0.0843	0.1064	0.0797	0.0512	0.0639	**0.0391**	0.0719
B2	0.0919	0.0916	**0.0836**	<u>0.0840</u>	0.1174	0.1186	0.0987	0.0996	0.0955	0.0927	0.1584	0.4260	0.3270	0.3617
C1	0.0635	<u>0.0565</u>	0.0614	**0.0549**	0.0888	0.0921	0.0839	0.0870	0.1136	0.0828	0.0670	0.0827	0.0762	0.1095
C2	0.0959	0.0998	**0.0874**	<u>0.0876</u>	0.1206	0.1236	0.1019	0.1045	0.1393	0.1149	0.1986	0.3582	0.3443	0.4125
D1	0.0682	0.0691	**0.0665**	<u>0.0671</u>	0.0896	0.0938	0.0844	0.0882	0.1147	0.0817	0.0686	0.0867	0.0823	0.1123
D2	0.1322	0.1334	<u>0.1106</u>	**0.1101**	0.1440	0.1407	0.1155	0.1170	0.1718	0.1570	0.2455	0.4123	0.3432	0.4447
E1	0.0547	<u>0.0372</u>	0.0548	**0.0361**	0.0834	0.0873	0.0794	0.0826	0.1089	0.0796	0.0558	0.0744	0.0568	0.0733
E2	0.0713	<u>0.0501</u>	0.0688	**0.0463**	0.1023	0.1086	0.0907	0.0951	0.1111	0.0806	0.1488	0.2336	0.2119	0.3515
F1	0.0547	**0.0368**	0.0542	<u>0.0374</u>	0.0837	0.0875	0.0794	0.0825	0.1090	0.0800	0.0563	0.0752	0.0587	0.0760
F2	0.0732	<u>0.0599</u>	0.0724	**0.0554**	0.1059	0.1116	0.0933	0.0963	0.1151	0.0872	0.1575	0.2869	0.2335	0.3595
G1	0.0566	<u>0.0404</u>	0.0554	**0.0400**	0.0856	0.0884	0.0814	0.0832	0.1095	0.0809	0.0587	0.0746	0.0616	0.0876
G2	0.0731	<u>0.0546</u>	0.0716	**0.0520**	0.1067	0.1130	0.0934	0.0967	0.1183	0.0862	0.1627	0.3021	0.2952	0.3749
H1	0.0567	**0.0408**	0.0551	<u>0.0424</u>	0.0859	0.0886	0.0815	0.0832	0.1099	0.0806	0.0592	0.0758	0.0629	0.0891
H2	0.0784	<u>0.0649</u>	0.0755	**0.0632**	0.1101	0.1155	0.0958	0.0982	0.1227	0.0928	0.1706	0.3287	0.3067	0.3804

Note: A1 and A2 refer to copula specification A with Kendall's τ being 0.3 and 0.6 respectively; other copulas are similarly denoted.

Table 3: Average ISE and IAE for tail performance

Copulas	$\hat{c}_m$		$\tilde{c}_m$		$\hat{c}_t$		$\tilde{c}_t$		$\hat{c}_r$	$\hat{c}_{b1}$	$\hat{c}_{b2}$	$\hat{c}_{p1}$	$\hat{c}_{p2}$	$\hat{c}_{p3}$
	plug in	cv	plug in	cv	plug in	cv	plug in	cv						
Panel 1: ISE														
A1	6.14	3.94	5.30	**3.80**	15.09	12.23	14.75	12.06	14.80	11.87	16.57	16.12	18.68	21.07
A2	31.97	16.08	27.79	**13.76**	56.51	41.98	51.62	31.94	103.84	89.50	112.75	136.01	140.40	134.39
B1	0.90	0.55	0.92	0.68	1.81	1.71	2.16	2.10	0.09	0.15	0.10	0.05	**0.04**	0.14
B2	4.89	3.26	4.18	2.66	7.24	7.05	7.29	6.62	0.54	**0.35**	1.21	7.98	5.27	3.56
C1	6.52	8.18	**6.40**	8.52	10.00	8.50	10.04	8.33	27.15	23.20	29.30	29.04	32.06	34.69
C2	38.53	41.49	**31.97**	36.48	47.69	40.18	38.69	32.57	170.40	156.05	179.42	202.37	200.88	200.86
D1	7.10	9.64	**6.81**	9.98	8.42	7.72	8.91	7.38	38.95	33.28	41.92	41.71	45.83	49.57
D2	84.41	91.05	77.21	83.18	81.17	78.10	**56.06**	57.18	385.05	367.36	395.88	420.00	416.58	420.38
E1	2.37	0.77	2.06	**0.62**	5.80	4.75	6.32	5.39	1.92	1.34	2.48	2.16	2.94	3.86
E2	24.38	7.43	23.23	**6.12**	44.82	30.41	43.58	26.31	36.99	27.93	43.32	52.92	53.99	59.45
F1	2.89	1.23	2.65	**1.08**	6.99	5.47	7.41	6.26	3.45	2.58	4.15	3.84	4.83	5.90
F2	27.07	12.54	24.34	**10.67**	46.25	34.06	42.08	26.02	50.28	41.10	56.63	69.40	68.10	72.61
G1	3.55	1.40	2.98	**1.14**	9.44	7.51	9.12	7.75	4.46	3.25	5.36	4.92	6.28	7.60
G2	27.70	9.11	25.38	**7.39**	51.51	35.32	49.01	28.89	54.60	43.65	61.67	78.84	79.02	79.73
H1	3.89	1.86	3.49	**1.68**	9.73	7.99	9.81	7.96	6.20	4.78	7.24	6.81	8.30	9.73
H2	29.45	14.57	27.09	**12.54**	50.79	36.62	45.77	29.67	68.14	57.36	75.21	93.63	92.47	92.72
Panel 2: IAE														
A1	0.71	0.69	0.68	**0.67**	0.91	0.85	0.90	0.84	0.97	0.84	1.27	1.20	1.58	1.96
A2	1.08	0.97	1.02	**0.89**	1.29	1.22	1.19	1.05	2.59	2.07	3.09	4.79	4.80	4.18
B1	0.41	0.36	0.42	0.42	0.54	0.53	0.61	0.64	0.20	0.27	0.22	**0.15**	0.16	0.30
B2	0.76	0.61	0.71	0.59	0.89	0.86	0.89	0.86	0.43	**0.32**	0.72	2.31	1.76	1.31
C1	0.65	0.68	**0.62**	0.68	0.75	0.72	0.74	0.71	0.94	0.80	1.19	1.12	1.43	1.68
C2	1.35	1.43	**1.21**	1.28	1.43	1.42	1.24	1.24	2.69	2.34	3.11	4.59	4.20	4.00
D1	0.64	0.77	**0.62**	0.77	0.68	0.69	0.67	0.68	1.02	0.87	1.27	1.24	1.50	1.73
D2	1.86	1.86	1.54	1.56	1.83	1.77	**1.46**	1.49	3.33	3.28	3.44	4.29	3.78	3.75
E1	0.45	0.29	0.43	**0.27**	0.63	0.57	0.65	0.62	0.53	0.45	0.72	0.60	0.83	1.05
E2	0.98	0.67	0.95	**0.61**	1.19	1.01	1.14	0.92	1.67	1.18	2.29	3.19	3.16	3.49
F1	0.47	0.33	0.46	**0.32**	0.64	0.59	0.66	0.63	0.57	0.47	0.78	0.67	0.92	1.14
F2	1.04	0.79	1.00	**0.73**	1.24	1.12	1.17	1.02	1.79	1.36	2.33	3.49	3.23	3.47
G1	0.52	0.38	0.50	**0.36**	0.72	0.66	0.73	0.69	0.66	0.55	0.88	0.78	1.09	1.34
G2	1.00	0.71	0.96	**0.66**	1.21	1.07	1.15	0.95	1.96	1.42	2.54	4.04	3.84	3.72
H1	0.53	0.41	0.51	**0.40**	0.72	0.66	0.73	0.69	0.69	0.58	0.93	0.83	1.14	1.39
H2	1.07	0.84	1.02	**0.78**	1.26	1.17	1.18	1.03	2.06	1.60	2.57	4.14	3.82	3.67

Note: A1 and A2 refer to copula specification A with Kendall's τ being 0.3 and 0.6 respectively; other copulas are similarly denoted.

8 Empirical applications

In this section, we apply the proposed MTK copula density estimator to two real world datasets. For comparison, we also estimate copula densities with suitable parametric copulas and the conventional TKE. The parametric copulas are estimated by the method of maximum likelihood. For the MTK and TKE, we use the bivariate kernel and select their smoothing parameters using the method of profile cross validation. All estimations are based on the pseudo-data of their corresponding data.

8.1 Loss and ALAE data

We first consider the classic loss and ALAE data, which contain the indemnity payment and allocated loss adjustment expense from 1,500 insurance claims. Copulas are employed to model the dependence between these two variables. This dataset has been used in the literature to illustrate copula fitting and goodness-of-fit testing, see, e.g., Frees and Valdez (1998),Klugman and Parsa (1999), and Lin and Wu (2015) among others. It is generally agreed that the Gumbel copula provides an adequate fit to these data. We include only uncensored observations (1,466 in total) in our estimation. The estimation results are reported in Figure 1 with 3-D density plots in the top panel and their corresponding contour plots in the bottom panel. It is seen that the MTK estimator $\tilde{c}_m$ produces a smooth and visually pleasant density that closely resembles the parametric Gumbel estimate. In contrast, the TKE $\tilde{c}_t$ clearly exhibits rather undesirable boundary behaviors, especially near the two diagonal corners. The contour plot of the MTK hints a slight asymmetry about the 45-degree diagonal; this is not permitted by the Gumbel copula, which is exchangeable between its two arguments.

8.2 FTSE 100 and Hang Seng indexes

Next we use copula to estimate the dependence structure between two financial series. Specifically, we examine the weekly log returns of FTSE 100 (London Stock Exchange) and Hang Seng (Hong Kong Stock Market) indexes from January 2010 through December 2013. In total, we have 729 observations.

Apparently stock returns are not i.i.d., which entails some initial modeling to remove features such as inter-temporal dependence and heteroskedasticity. Following a common practice in the literature, for each series $r_t : t = 1, \cdots, n$, we employ a GARCH(1,1) model, i.e. $r_t = \mu + h_t$, $h_t = \sigma_t \epsilon_t$ and $\sigma_t^2 = \kappa + \alpha h_{t-1}^2 + \beta \sigma_{t-1}^2$ where the standardized residuals $\epsilon_t : t = 1, \cdots, n$ are assumed to be i.i.d. and follow the Student t-distribution with zero mean and unity variance.

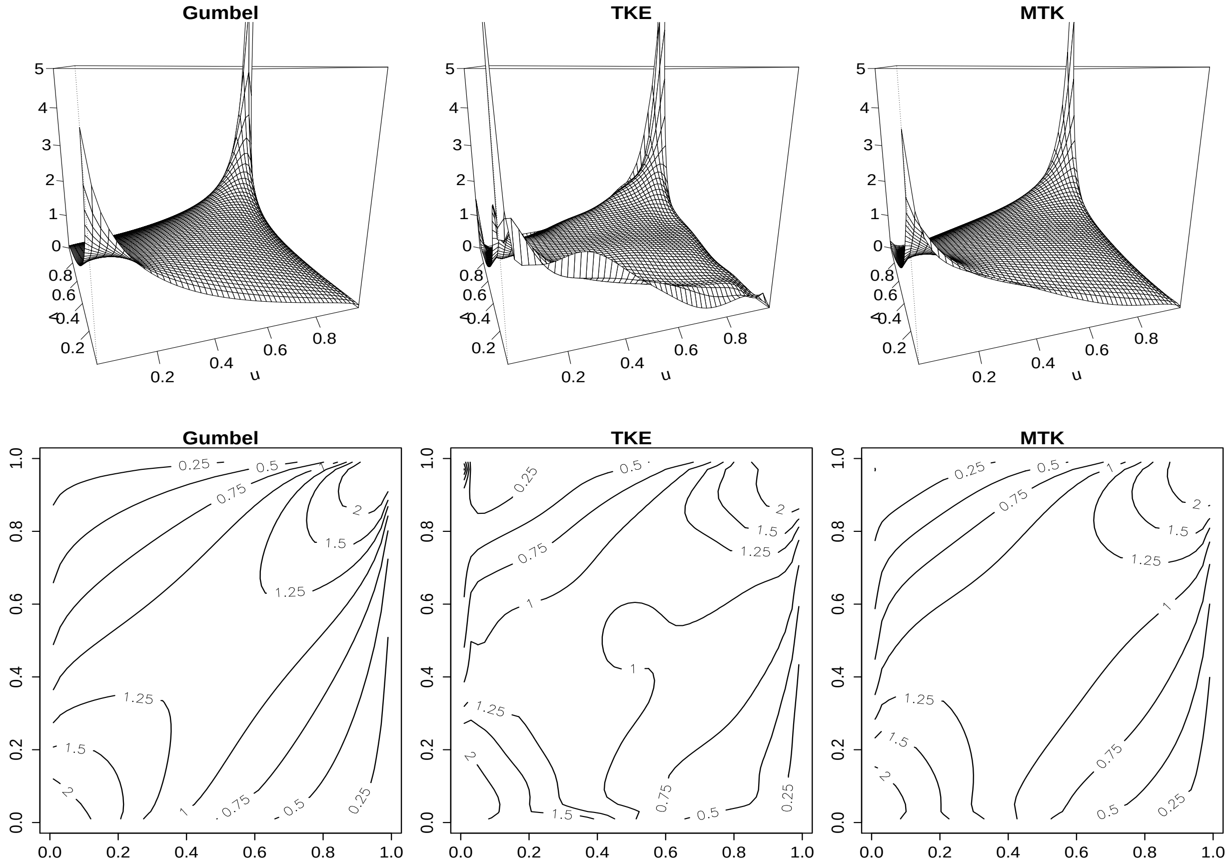

Figure 1: Copula density estimates of Loss and ALAE data

The scatterplot of the standardized residuals, which is not reported to save space, shows that the bottom left quarter of $\mathcal{I}$ contains considerably more observations than dose the top right quarter, indicating stronger lower tail dependence than upper tail dependence between the two markets. This is consistent with a well-documented feature of stock markets: bear markets tend to move together more likely than bull markets do. The primary task here is to model the dependence structure of the standardized residuals obtained from the two return series. For comparison, we also estimate a Gaussian copula, which is routinely used to model the dependence structure among standardized residuals of financial time series models. The estimated copula densities are presented in Figure 2. It is seen that the MTK successfully captures the asymmetric tail dependence between the two stock series: while the lower end density of the MTK is comparable to that of the Gaussian copula, the upper end density appears to be smaller than its Gaussian counterpart. The TKE estimate again exhibits some undesirable boundary irregularities.

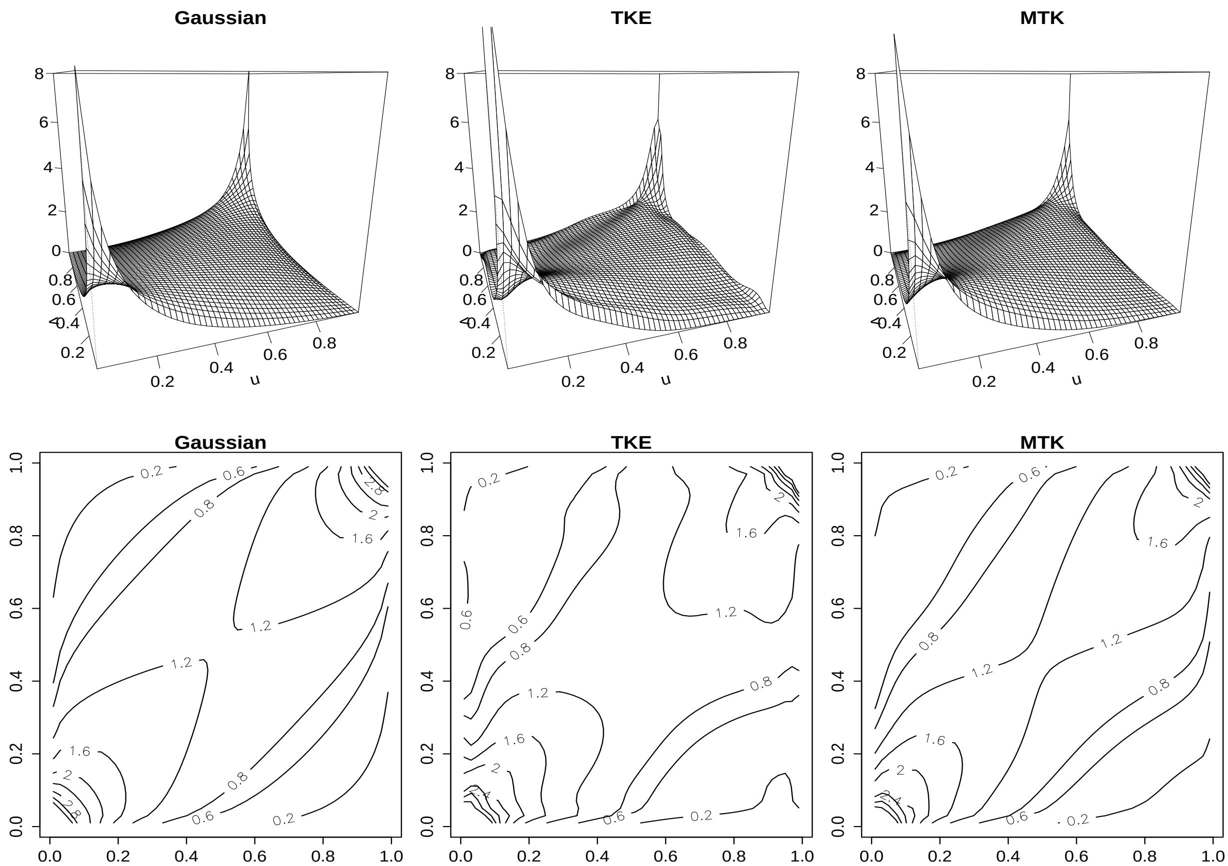

Figure 2: Copula density estimate of FTSE 100 and Hang Seng Indexes

9 Concluding remarks

Despite its popularity in density estimation, kernel type estimator is not widely used to estimate copula densities because the standard kernel estimator suffers severe boundary biases. The transformation-kernel estimator provides a natural solution to boundary bias correction but may lead to erratic boundary behaviors because of an unbounded multiplier associated with the back-transformation. We propose a modified transformation-kernel estimator that employs a tapering method to mitigate the influence of the multiplier while maintaining the simplicity of fixed transformation and a single global bandwidth. We establish the theoretical properties of the proposed estimator and show that it asymptotically dominates the naive transformation-kernel estimator. We further show that the proposed estimator enjoys higher order convergence rate under Gaussian copulas. Therefore, our estimator should provide outstanding performance for Gaussian copulas and near Gaussian copulas, which are practically important in financial data analyses. Extensions to non-diagonal bandwidth matrix produce further improvement. We propose two practically simple methods to select the optimal smoothing parameters. Our simulation results demonstrate the superior finite sample performance of the proposed estimator. We consider only i.i.d. samples in this study. Extensions of the proposed methods to copula density estimation based on time series data may be of interest for future work. Generalization to accommodate censored/truncated data is another possibility.

References

Ang, A. and Chen, J. (2002), “Asymmetric correlations of equity portfolios,” *Journal of Financial Economics*, 63, 443–494.

Autin, F., Pennec, E. L., and Tribouley, K. (2010), “Thresholding methods to estimate copula density,” *Journal of Multivariate Analysis*, 101, 200 – 222.

Bouezmarni, T., El Ghouch, A., and Taamouti, A. (2013), “Bernstein estimator for unbounded copula densities,” *Statistics & Risk Modeling*, 30, 343–360.

Bouezmarni, T., Rombouts, J. V., and Taamouti, A. (2010), “Asymptotic properties of the Bernstein density copula estimator for α-mixing data,” *Journal of Multivariate Analysis*, 101, 1–10.

— (2012), “Nonparametric copula-based test for conditional independence with applications to granger causality,” *Journal of Business & Economic Statistics*, 30, 275–287.

Boyd, S. and Vandenberghe, L. (2004), *Convex Optimization*, Cambridge University Press.

Charpentier, A., Fermanian, J.-D., and Scaillet, O. (2006), “The Estimation of Copulas: Theory and Practice,” in *Copulas, from theory to application in Finance*, ed. Rank, J., Risk Books.

Chaudhuri, P. and Marron, J. S. (1999), “SiZer for Exploration of Structure in Curves,” *Journal of the American Statistical Association*, 94, 807–823.

Chen, S. X. (1999), “Beta kernel estimators for density functions,” *Computational Statistics and Data Analysis*, 31, 131–145.

Chen, S. X. and Huang, T.-M. (2007), “Nonparametric estimation of copula functions for dependence modelling,” *The Canadian Journal of Statistics*, 35, 265–282.

Chen, X. and Fan, Y. (2006a), "Estimation and Model Selection of Semiparametric Copula-based Multivariate Dynamic Models under Copula Misspecification," *Journal of Econometrics*, 135, 125–154.

— (2006b), "Estimation of copula-based semiparametric time series models," *Journal of Econometrics*, 130, 307–335.

Chen, X., Fan, Y., and Patton, A. J. (2004), "Simple Tests for Models of Dependence between Multiple Financial Time Series, with Applications to US Equity Returns and Exchange Rates," *London Economics Financial Markets Group Working Paper*.

Chen, X., Fan, Y., Pouzo, D., and Ying, Z. (2010), "Estimation and model selection of semiparametric multivariate survival functions under general censorship," *Journal of Econometrics*, 157, 129 – 142.

Chen, X., Fan, Y., and Tsyrennikov, V. (2006), "Efficient Estimation of Semiparametric Multivariate Copula Models," *Journal of the American Statistical Association*, 101, 1228–1240.

Duong, T. and Hazelton, M. (2003), "Plug-in bandwidth matrices for bivariate kernel density estimation," *Journal of Nonparametric Statistics*, 15, 17–30.

Duong, T. and Hazelton, M. L. (2005), "Convergence rates for unconstrained bandwidth matrix selectors in multivariate kernel density estimation," *Journal of Multivariate Analysis*, 93, 417 – 433.

Fan, Y. and Patton, A. J. (2014), "Copulas in Econometrics," *Annual Review of Economics*, 6, 179–200.

Fermanian, J.-D. (2005), "Goodness-of-fit tests for copulas," *Journal of Multivariate Analysis*, 95, 119–152.

— (2012), "An overview of the goodness-of-fit test problem for copulas," in *Copulae in Mathematical and Quantitative Finance*, eds. Jaworski, P., Durant, F., and Härdle, W., Springer.

Fermanian, J.-D. and Scaillet, O. (2003), "Nonparametric Estimation of Copulas for Time Series," *Journal of Risk*, 5, 25–54.

Frees, E. W. and Valdez, E. A. (1998), "Understanding Relationships Using Copulas," *North American Actuarial Journal*, 2, 1–25.

Gao, Y., Zhang, Y. Y., and Wu, X. (2015), "Penalized Exponential Series Estimation of Copula Densities with an Application to Intergenerational Dependence of Body Mass Index," *Empirical Economics*, 48, 61–81.

Geenens, G., Charpentier, A., and Paindaveine, D. (2014), "Probit transformation for nonparametric kernel estimation of the copula density," *arXiv preprint arXiv:1404.4414*.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995), "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrika*, 82, 543–552.

Genest, C., Masiello, E., and Tribouley, K. (2009), "Estimating copula densities through wavelets," *Insurance: Mathematics and Economics*, 44, 170 – 181.

Genest, C. and Segers, J. (2010), "On the covariance of the asymptotic empirical copula process," *Journal of Multivariate Analysis*, 101, 1837–1845.

Gijbels, I. and Mielniczuk, J. (1990), "Estimating the density of a copula function," *Communications in Statistics. Theory and Methods*, 19, 445–464.

Hall, P. and Neumeyer, N. (2006), "Estimating a bivariate density when there are extra data on one or both components," *Biometrika*, 93, 439–450.

Hjort, N. L. and Glad, I. K. (1995), "Nonparametric density estimation with a parametric start," *The Annals of Statistics*, 882–904.

Hong, Y., Tu, J., and Zhou, G. (2007), "Asymmetries in Stock Returns: Statistical Tests and Economic Evaluation," *Review of Financial Studies*, 20, 1547–1581.

Janssen, P., Swanepoel, J., and Veraverbeke, N. (2014), "A note on the asymptotic behavior of the Bernstein estimator of the copula density," *Journal of Multivariate Analysis*, 124, 480 – 487.

Joe, H. (1997), *Multivariate models and dependence concepts*, Chapman and Hall.

Kauermann, G., Schellhase, C., and Ruppert, D. (2013), "Flexible Copula Density Estimation with Penalized Hierarchical B-splines," *Scandinavian Journal of Statistics*, 40.

Klugman, S. A. and Parsa, R. (1999), "Fitting bivariate loss distributions with copulas," *Insurance: Mathematics and Economics*, 24, 139 – 148.

Lee, T.-H. and Long, X. (2009), "Copula-based multivariate GARCH model with uncorrelated dependent errors," *Journal of Econometrics*, 150, 207–218.

Lin, J. and Wu, X. (2015), "Smooth Tests of Copula Specifications," *Journal of Business and Economic Statistics*, 33, 128–143.

Longin, F. and Solnik, B. (2001), "Extreme Correlation of International Equity Markets," *Journal of Finance*, 649–676.

Marron, J. S. and Ruppert, D. (1994), "Transformations to Reduce Boundary Bias in Kernel Density Estimation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 56, 653–671.

Nagler, T. (2016), "kdecopula: An R Package for the Kernel Estimation of Copula Densities," *arXiv preprint arXiv:1603.04229*.

Nelsen, R. B. (2006), *An introduction to copulas*, Springer.

Okhrin, O., Okhrin, Y., and Schmid, W. (2013), "On the structure and estimation of hierarchical Archimedean copulas," *Journal of Econometrics*, 173, 189–204.

Omelka, M., Gijbels, I., and Veraverbeke, N. (2009), "Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing," *The Annals of Statistics*, 37, 3023–3058.

Patton, A. J. (2006), "Modelling asymmetric exchange rate dependence*," *International economic review*, 47, 527–556.

Prokhorov, A. and Schmidt, P. (2009), "Likelihood-based estimation in a panel setting: robustness, redundancy and validity of copulas," *Journal of Econometrics*, 153, 93–104.

Qu, L. and Yin, W. (2012), "Copula density estimation by total variation penalized likelihood with linear equality constraints," *Computational Statistics & Data Analysis*, 56, 384 – 398.

Sancetta, A. (2007), "Nonparametric estimation of distributions with given marginals via Bernstein–Kantorovich polynomials: L1 and pointwise convergence theory," *Journal of Multivariate Analysis*, 98, 1376–1390.

Sancetta, A. and Satchell, S. (2004), "The Bernstein copula and its applications to modeling and approximations of multivariate distributions," *Econometric Theory*, 20, 535–562.

Scaillet, O. (2007), "Kernel-based goodness-of-fit tests for copulas with fixed smoothing parameters," *Journal of Multivariate Analysis*, 98, 533–543.

Segers, J. (2012), "Asymptotics of empirical copula processes under non-restrictive smoothness assumptions," *Bernoulli*, 18, 764–782.

Shen, X., Zhu, Y., and Song, L. (2008), "Linear B-spline copulas with applications to nonparametric estimation of copulas," *Computational Statistics & Data Analysis*, 52, 3806 – 3819.

Sklar, A. (1959), "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, 8, 229–231.

Stone, C. J. (1984), "An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates," *Annals of Statistics*, 12, 1285–1297.

Taamouti, A., Bouezmarni, T., and El Ghouch, A. (2014), "Nonparametric estimation and inference for conditional density based Granger causality measures," *Journal of Econometrics*, 180, 251–264.

Wand, M. and Jones, M. (1995), *Kernel Smoothing*, Chapman and Hall.

Wand, M. P., Marron, J. S., and Ruppert, D. (1991), "Transformations in Density Estimation," *Journal of the American Statistical Association*, 86, 343–353.

Wen, K. and Wu, X. (2015), "An Improved Transformation-based Kernel Estimator of Densities on the Unit Interval," *Journal of the American Statistical Association*, 110, 773–783.

Zimmer, D. M. (2012), "The Role of Copulas in the Housing Crisis," *Review of Economics and Statistics*, 94, 607–620.

Appendices

A Assumptions

Assumption 1. $\left\{(X_i, Y_i)^\top, i = 1, \cdots, n\right\}$ *is an i.i.d. sample from a joint distribution* F *that is absolutely continuous. The associated marginal distributions* F_X *and* F_Y *are strictly increasing on their support.*

Assumption 2. *The copula C of F is such that $(\partial C/\partial u)(u,v)$ and $(\partial^2 C/\partial u^2)(u,v)$ exist and are continuous on $\{(u,v): u \in (0,1), v \in [0,1]\}$, and $(\partial C/\partial v)(u,v)$ and $(\partial^2 C/\partial v^2)(u,v)$ exist and are continuous on*
$\{(u,v): u \in [0,1], v \in (0,1)\}$. In addition, there are constants K_1 and K_2 such that

$$\left|\frac{\partial^2 C}{\partial u^2}(u,v)\right| \leq \frac{K_1}{u(1-u)} \quad \forall (u,v) \in (0,1) \times [0,1]$$

and

$$\left|\frac{\partial^2 C}{\partial v^2}(u,v)\right| \leq \frac{K_2}{v(1-v)} \quad \forall (u,v) \in [0,1] \times (0,1).$$

Assumption 3. *The copula density c exists, is positive and admits continuous second order partial derivatives on the interior of $\mathcal{I}$. In addition, there is a constant K_{00} such that*

$$c(u,v) \leq K_{00} \min\left(\frac{1}{u(1-u)}, \frac{1}{v(1-v)}\right) \quad \forall (u,v) \in (0,1)^2.$$

Assumption 4. *$\boldsymbol{\theta} \sim h^2$; as $n \to \infty$, $\boldsymbol{\theta} \to 0$.*

Assumption 5. *Under the diagonal bandwidth matrix $\boldsymbol{H} = h^2 \boldsymbol{I}$, $h \sim n^{-a}$ where $a \in \left[\frac{1}{6}, \frac{1}{4}\right)$.*

Assumption 6. *Under the diagonal bandwidth matrix $\boldsymbol{H} = h^2 \boldsymbol{I}$, $h \sim n^{-a}$ where $a \in \left[\frac{1}{10}, \frac{1}{4}\right)$.*

Assumption 1 guarantees the copula C associated with the distribution F is unique. Assumptions 2 and 3 are imposed in Geenens et al. (2014) to establish the asymptotic normality (3.2) of the transformation-kernel estimator $\hat{c}_t$, which eases our proof considerably. These two assumptions mostly reduce to some conditions required in empirical copula process literature, e.g. Segers (2012). Moreover, they are non-restrictive and hold for many common copula families, such as Gaussian, Student, Clayton, Gumbel copulas, among others (Omelka et al., 2009; Segers, 2012). In particular, Assumption 3 allows the copula density c to grow unboundedly as long as in a suitable manner. Assumption 4 provides the optimal order for the tuning parameters introduced in the tapering device. Intuitively, in order for bias reduction, the decaying rate of $\boldsymbol{\theta}$ should match the leading bias term from the transformation-kernel estimator, which is of order $O(h^2)$. Lastly, Assumptions 5 and 6 restrict the order of bandwidth to ensure the validity of Theorem 1 and 2.

B Proofs

B.1 Proof of Theorem 1

Proof. We first consider the properties of the 'ideal' estimators. Define $\hat{c}_m^*$, $\hat{c}_t^*$, and $J^*(u, v; h, \boldsymbol{\theta})$ analogously to $\hat{c}_m$, $\hat{c}_t$, and $J(u, v; h, \boldsymbol{\theta})$ respectively but use the true marginal distributions $\{(S_i, T_i), i = 1, \cdots, n\}$. Similarly to (3.4), we have the relation $\hat{c}_m^*(u, v) = J^*(u, v; h, \boldsymbol{\theta})\hat{c}_t^*(u, v)$. A Taylor expansion of $J^*(u, v; h, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ at $\mathbf{0}$ yields

$$J^*(u, v; h, \boldsymbol{\theta}) = 1 + \boldsymbol{\theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) + o(\boldsymbol{\theta}),$$

which coincides with (3.4) since J and J^* only differ in their normalization factors, both of which are reduced to one when $\boldsymbol{\theta} = 0$. The asymptotic bias and variance of $\hat{c}_t^*$ are given in Charpentier et al. (2006). We then have, for a fixed point $(u, v) \in (0, 1)^2$,

$$\begin{aligned}
\text{abias}\left\{\hat{c}_m^*(u, v)\right\} =& b_m(u, v) \\
\equiv& \frac{h^2\left(g^{(2,0)} + g^{(0,2)}\right)\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)}{2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \\
&+ \boldsymbol{\theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) c(u, v),
\end{aligned}$$

and

$$\text{avar}\left\{\hat{c}_m^*(u, v)\right\} = \frac{\sigma_m^2(u, v)}{nh^2} \equiv \frac{c(u, v)}{4\pi nh^2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}.$$

Next we shall show that the above asymptotic properties of $\hat{c}_m^*$ are inherited by the feasible $\hat{c}_m$. First note that

$$\begin{aligned}
\hat{c}_m(u, v) &= J(u, v; h, \boldsymbol{\theta})\hat{c}_t(u, v) \\
&= \left\{J(u, v; h, \boldsymbol{\theta}) - J^*(u, v; h, \boldsymbol{\theta})\right\}\hat{c}_t(u, v) + J^*(u, v; h, \boldsymbol{\theta})\hat{c}_t(u, v),
\end{aligned}$$

and

$$b_m(u, v) = b_t(u, v) + \boldsymbol{\theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) c(u, v).$$

It follows that

$$\begin{aligned}
&\sqrt{nh^2}\left(\hat{c}_m(u,v)-c(u,v)-b_m(u,v)\right)\\
=&\left(J(u,v;h,\boldsymbol{\theta})-J^*(u,v;h,\boldsymbol{\theta})\right)\sqrt{nh^2}\hat{c}_t(u,v)\\
+&\sqrt{nh^2}\{J^*(u,v;h,\boldsymbol{\theta})\hat{c}_t(u,v)-c(u,v)-b_t(u,v)\\
-&\boldsymbol{\theta}^\top\boldsymbol{B}\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)c(u,v)\}\\
\equiv& I_1+I_2.
\end{aligned}$$

The first term above can be re-written as

$$\begin{aligned}
I_1=&\left(J(u,v;h,\boldsymbol{\theta})-J^*(u,v;h,\boldsymbol{\theta})\right)\sqrt{nh^2}\left(\hat{c}_t(u,v)-c(u,v)-b_t(u,v)\right)\\
&+\left(J(u,v;h,\boldsymbol{\theta})-J^*(u,v;h,\boldsymbol{\theta})\right)\sqrt{nh^2}\left(c(u,v)+b_t(u,v)\right).
\end{aligned}$$

Both $J(u,v;h,\boldsymbol{\theta})$ and $J^*(u,v;h,\boldsymbol{\theta})$ are in the form of sample average. It can then be shown that, by a simple Taylor expansion, $J(u,v;h,\boldsymbol{\theta})-J^*(u,v;h,\boldsymbol{\theta})=o_p\left(n^{-1/2}\right)=o_p\left((nh^2)^{-1/2}\right)$. From (3.2), we have
$\sqrt{nh^2}\left(\hat{c}_t(u,v)-c(u,v)-b_t(u,v)\right)=O_p(1)$; together with the above result, the first term in I_1 is $o_p(1)$. Since $J(u,v;h,\boldsymbol{\theta})-J^*(u,v;h,\boldsymbol{\theta})=o_p\left((nh^2)^{-1/2}\right)$, the second term in I_1 is $o_p(1)$ as well given that $c(u,v)$ is bounded for interior (u,v) and $b_t(u,v)$ is $o_p(1)$. Therefore, we have $I_1=o_p(1)$.

Nextly plugging (3.5) into I_2 yields

$$\begin{aligned}
I_2=&\sqrt{nh^2}\left(\hat{c}_t(u,v)-c(u,v)-b_t(u,v)\right)\\
&+\sqrt{nh^2}\boldsymbol{\theta}^\top\boldsymbol{B}\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)\left(\hat{c}_t(u,v)-c(u,v)\right)\\
&+\sqrt{nh^2}\hat{c}_t(u,v)o\left(\boldsymbol{\theta}\right)\\
&\equiv I_{21}+I_{22}+I_{23}.
\end{aligned}$$

According to (3.2), we have

$$I_{21}\xrightarrow{d}\mathcal{N}\left(0,\sigma_m^2(u,v)\right),$$

since $\sigma_m^2(u,v)=\sigma_t^2(u,v)$. For I_{22}, we have

$$\begin{aligned}
I_{22}=&\sqrt{nh^2}\boldsymbol{\theta}^\top\boldsymbol{B}\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)\left(\hat{c}_t(u,v)-c(u,v)-b_t(u,v)\right)\\
&+\sqrt{nh^2}\boldsymbol{\theta}^\top\boldsymbol{B}\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)b_t(u,v).
\end{aligned}$$

The first term in I_{22} is apparently of order $o(1)\cdot O_p(1)=o_p(1)$. Since $b_t(u,v)=O\left(h^2\right)$ and by

Assumption 4 that $\boldsymbol{\theta} = O\left(h^2\right)$, we have the second term in I_{22} being $O\left(nh^{10}\right) = o(1)$ by the fact that $h \propto n^{-a}$ where $a \in [1/6, 1/4)$. Therefore, $I_{22} = o_p(1)$. For I_{23}, we have

$$I_{23} = \sqrt{nh^2}\left(\hat{c}_t(u,v) - c(u,v) - b_t(u,v)\right) o\left(\boldsymbol{\theta}\right) + \sqrt{nh^2}\left(c(u,v) + b_t(u,v)\right) o\left(\boldsymbol{\theta}\right).$$

The first term in I_{23} is again $o(1) \cdot O_p(1) = o_p(1)$. For the second term in I_{23}, we have

$$\sqrt{nh^2} c(u,v) o\left(\boldsymbol{\theta}\right) = \sqrt{nh^6} c(u,v) o(1) = o(1)$$

since $\sqrt{nh^6} = O(1)$; similarly we also have $\sqrt{nh^2} b_t(u,v) o\left(\boldsymbol{\theta}\right) = o(1)$. Thus, we have $I_{23} = o_p(1)$.

The results of Theorem 1 is then obtained by combining the above results. $\square$

B.2 Proof of Theorem 2

Proof. We first slightly extend the results stated in (3.2). Under the assumptions in Theorem 2, we have, for any $(u,v) \in (0,1)^2$,

$$\sqrt{nh^2}\left(\hat{c}_t(u,v) - c(u,v) - b_t^{(G)}(u,v)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_m^2(u,v)\right),$$

where $b_t^{(G)}(u,v)$ is defined in (5.2). Since $b_t^{(G)}(u,v)$ includes the higher order bias term that is associated with h^4, we can relax the condition to $h \sim n^{-a}$ where $a \in [1/10, 1/4)$.

Note that $b_m^{(G)}(u,v) = b_m(u,v) + h^4 R\left(\Phi^{-1}(u), \Phi^{-1}(v); \rho\right)$ using that $b_m(u,v) = 0$ in this case and the higher order Taylor expansion of $J^*(u,v;h,\boldsymbol{\theta})$ given in (5.3). Then the proof is analogous to that of Theorem 1 above, thus is omitted here. $\square$

B.3 Proof of Proposition 1

Proof. Define a 3-dimensional vector
$q(s,t) = \left(\boldsymbol{B}(s,t)^\top g(s,t), g^{(2,0)}(s,t) + g^{(0,2)}(s,t)\right)^\top$ and a matrix

$$Q = \int_{\mathcal{R}^2} q(s,t) q(s,t)^\top \mathrm{d}s\mathrm{d}t.$$

By construction, we have $q(s,t)q(s,t)^\top$ is positive-semidefinite and so is Q. Here Q can also be written in the block matrix form, namely

$$Q = \begin{pmatrix} \mathbf{\Gamma}_1 & \mathbf{\Gamma}_2 \\ \mathbf{\Gamma}_2^\top & \Gamma_3 \end{pmatrix}.$$

Note that $\Gamma_3 - \mathbf{\Gamma}_2^\top \mathbf{\Gamma}_1^{-1} \mathbf{\Gamma}_2$ is the Schur complement of Γ_3. According to Schur complement lemma, see Boyd and Vandenberghe (2004), the Schur complement of Γ_3 in Q is positive-semidefinite if and only if Q is positive-semidefinite. Therefore, we have $\Gamma_3 - \mathbf{\Gamma}_2^\top \mathbf{\Gamma}_1^{-1} \mathbf{\Gamma}_2 \geq 0$. It is straightforward to check that when the underlying copula is Gaussian, i.e. g is the pdf of a bivariate Gaussian distribution, we have $\Gamma_3 - \mathbf{\Gamma}_2^\top \mathbf{\Gamma}_1^{-1} \mathbf{\Gamma}_2 = 0$. □

C Exact formula of $\int_{\mathcal{I}} \left(\tilde{c}_m(u,v)\right)^2 \phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))\mathrm{d}u\mathrm{d}v$

To facilitate the implementation of the profile cross validation, we provide here an analytical form of the first integral term in (4.5) for the MTK with a non-diagonal bandwidth matrix (6.1):

$$\begin{aligned}
&\int_{\mathcal{I}} \left(\tilde{c}_m(u,v)\right)^2 \phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))\mathrm{d}u\mathrm{d}v = \frac{1}{4\pi n^2 \eta^2 h^2 \delta \sqrt{1-\lambda^2}} \\
&\sum_{i=1}^n \sum_{j=1}^n \exp\left\{ \frac{\alpha_1 \left(\hat{S}_i^2 + \hat{S}_j^2 + \hat{T}_i^2 + \hat{T}_j^2\right) + \alpha_2 \left(\hat{S}_i\hat{T}_i + \hat{S}_j\hat{T}_j\right)}{4h^2(1-\lambda^2)\delta^2} \right. \\
&\left. + \frac{\alpha_3 \left(\hat{S}_i\hat{T}_j + \hat{S}_j\hat{T}_i\right) + \alpha_4 \left(\hat{S}_i\hat{S}_j + \hat{T}_i\hat{T}_j\right)}{4h^2(1-\lambda^2)\delta^2} \right\},
\end{aligned}$$

where

$$\begin{aligned}
\alpha_1 &= -2h^4(1-\lambda^2)(4\theta_1^2 - \theta_2^2) + 2h^2((\lambda^2-3)\theta_1 - \lambda\theta_2) - 1 \\
\alpha_2 &= 4h^4(1-\lambda^2)(4\theta_1^2 - \theta_2^2) + 2h^2(4\lambda\theta_1 + (3\lambda^2-1)\theta_2) + 2\lambda \\
\alpha_3 &= -2h^2(4\lambda\theta_1 + (1+\lambda^2)\theta_2) - 2\lambda \\
\alpha_4 &= 4h^2((1+\lambda^2)\theta_1 + \lambda\theta_2) + 2.
\end{aligned}$$

Note that the MTK $\hat{c}_m$ with a diagonal bandwidth matrix can be obtained as a special case by setting $\lambda = 0$.