Estimating copula densities through wavelets

Christian Genest, Esterina Masiello, Karine Tribouley

HAL Id: hal-00257425

https://hal.science/hal-00257425v1

Preprint submitted on 19 Feb 2008 (v1), last revised 26 Feb 2008 (v2)

Estimating copula densities through wavelets

Christian Genest

Département de mathématiques et de statistique, Université Laval
1045, avenue de la Médecine, Québec (Québec), Canada G1V 0A6

Esterina Masiello

Modal'X, Université Paris X, 200, avenue de la République
92000 Nanterre, France

Karine Tribouley

Laboratoire de probabilités et modèles aléatoires
175, rue du Chevaleret, 75013 Paris
Modal'X, Université Paris X, 200, avenue de la République
92001 Nanterre, France

Abstract

Wavelet analysis is used to construct a rank-based estimator of a copula density. The procedure, which can be easily implemented with ready-to-use wavelet packages, is based on an algorithm that handles boundary effects automatically. The resulting estimator provides a nonparametric benchmark for the selection of a parametric copula family. From a theoretical point of view, the estimation procedure is shown to be optimal in the minimax sense on a large functional class of regular copula densities. The approach is illustrated with actuarial and financial data.

Key words: Copulas; Nonparametric Estimation; Ranks; Wavelets.

1 Introduction

Copulas are quickly gaining in popularity as a modeling tool for multivariate data. The books by Joe (1997) and Nelsen (2006) describe the mathematical and statistical foundations of the subject. A host of applications in insurance and risk management can be found in the more recent texts by Cherubini et al. (2004) and McNeil et al. (2005), among others.

Suppose that the relation between variables X and Y is of interest and assume for simplicity that both of them are real-valued. Let $F(x) = \Pr(X \leq x)$ and $G(y) = \Pr(Y \leq y)$ be their cumulative distribution functions. Following Sklar (1959), the joint distribution function of the pair (X, Y) may be expressed in the form

$$H(x, y) = C(F(x), G(y)), \tag{1}$$

for some distribution function C whose margins are uniform on the interval $(0, 1)$. When F and G are continuous, C is unique and coincides with the distribution function of the pair $(U, V) = (F(X), G(Y))$. In practice, H is unknown. A copula model for the pair (X, Y) can then be constructed by assuming that F, G and C belong to specific classes of distributions. An advantage of this approach is that the copula C, which characterizes the dependence between X and Y, can be chosen separately from the marginal models.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from the unknown distribution H. Denote by F_n and G_n the empirical distributions associated with F and G. A first step in selecting an appropriate class of copulas consists of plotting the pairs

$$\left(\frac{R_i}{n}, \frac{S_i}{n} \right) = (F_n(X_i), G_n(Y_i)), \quad i \in \{1, \ldots, n\}. \tag{2}$$

Here, R_i is the rank of X_i among $X_1, \ldots, X_n$ and S_i is the rank of Y_i among $Y_1, \ldots, Y_n$. The motivation behind this graphical approach is that the pseudo observations $(R_i/n, S_i/n)$ are close substitutes to the unobservable pairs $(U_i, V_i) = (F(X_i), G(Y_i))$ forming a random sample from C.

To illustrate this point, Figure 1 displays four scatter plots generated from the same random sample of size $n = 2000$ from a Gumbel copula with Kendall's tau $\tau = 1/2$. Panel a) shows the original sample; panels b) and c) show what happens when G is exponential with mean $1/3$ and F is either $\mathcal{N}(0, 1)$ or $\mathcal{B}(0.5, 0.7)$, respectively; panel d) shows the pairs of normalized ranks. As one can easily see, the effect of the marginals, which is dominant in panels b) and c), is completely suppressed in plot d). This rank-rank plot enhances the underlying characteristics of the dependence structure, such as the tendency of extreme values of both variables to occur together.
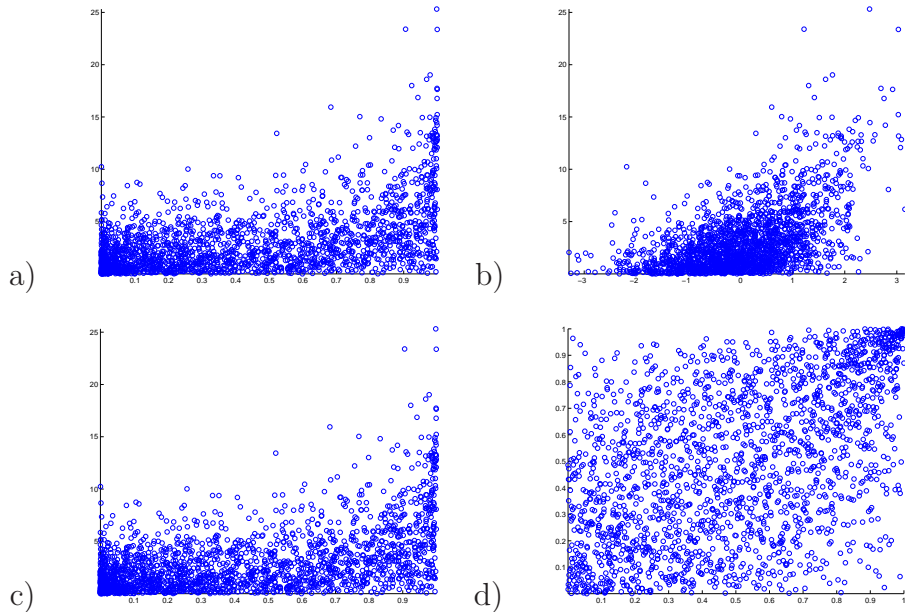
Fig. 1. Scatter and rank-rank plots for a sample of size 2000 from the Gumbel copula with $\tau = 1/2$, once the margins have been transformed to F and G. Panel a): $F = \mathcal{B}(0.5, 0.7)$, $G = \mathcal{E}(3)$; panel b): $F = \mathcal{N}(0, 1)$, $G = \mathcal{E}(3)$; panel c): $F = G = \mathcal{U}(0, 1)$; panel d): the rank-rank plot associated with the three transformed samples coincide.

A clear limitation of the rank-rank plot is that it becomes quickly unreadable as the number of observations increases. For a copula with small to moderate dependence, a random sample of size 10,000 or more can easily fill the square, and all features of the distribution are lost. As an alternative, one might consider plotting the empirical distribution C_n of the pairs $(R_i/n, S_i/n)$. Although the convergence of C_n to C was established by Deheuvels (1979), the graph is not of much help.

Consider for example random samples of size 2000 from the Gumbel and Clayton copulas with $\tau = 1/2$. Panels a) and b) of Figure 2 show that while these two Archimedean copulas induce the same degree of dependence, they have very different behavior, particularly in the tails. However, none of these differences is apparent in panels c) and d), which display the corresponding empirical copulas.

A more promising avenue consists of plotting 3D-histograms of the relative frequencies of the pseudo observations $(R_i/n, S_i/n)$, as measured on an arbitrary partition of the unit square. This is done in Figure 3 for the same data as in Figure 2. The distinctive features of the Gumbel and Clayton copulas are now much more apparent, but this is at the cost of a rather erratic picture.

The purpose of this paper is to present a smoothed version of the 3D-histogram that practitioners could use as a graphical tool to spot the key features of a copula
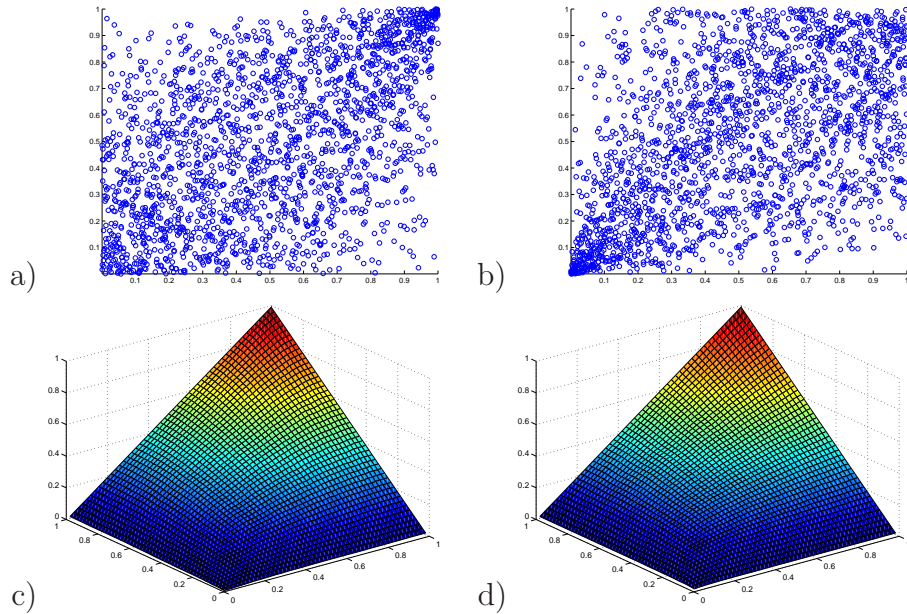
Fig. 2. Rank-rank plots and corresponding empirical copulas of random samples of size $n = 2000$ from two copulas with $\tau = 1/2$. Panels a) and c): Gumbel copula with parameter 1.5; panels b) and d): Clayton copula with parameter 1. In all cases, the margins are $F = \mathcal{B}(0.5, 0.7)$ and $G = \mathcal{E}(3)$.

dependence structure such as symmetry, skewness or heavy-tail behavior. In more technical terms, what will be proposed is a nonparametric (rank-based) estimator of the copula density

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v), \quad u, v \in (0, 1).$$

This density is assumed to exist and to be square-integrable in the sequel.
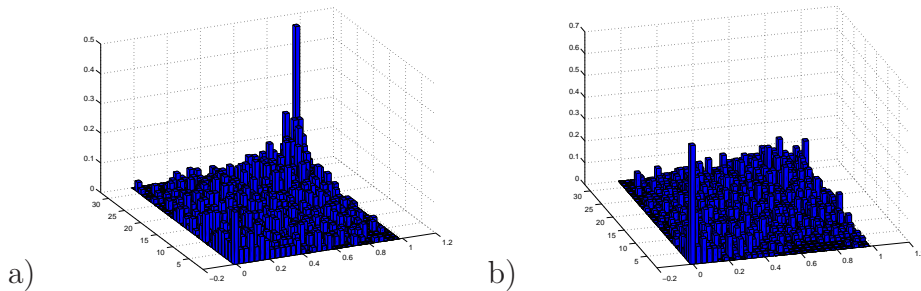

Fig. 3. 3D-histograms showing the relative frequency of $n = 2000$ pairs $(R_i/n, S_i/n)$ in a 32×32 regular partition of the unit square for a copula with $\tau = 1/2$. Panel a): Gumbel; panel b): Clayton.

3

The approach described here is based on the wavelet decomposition of the copula density. A brief introduction to this methodology is given in Section 2, where multi-resolution techniques are shown to be helpful in capturing the main features of a known copula c. Section 3 considers the case where c is unknown and must be estimated from ranks. A mathematical framework for the study of this procedure is then described in Section 4, and optimality results are stated in Section 5. Examples of applications to actuarial science and finance are discussed in Section 6. All mathematical derivations are deferred to the Appendix.

This paper is not the first to propose a nonparametric, rank-based estimator of the copula density. An early contribution along those lines was made by Gijbels and Mielniczuk (1990), who used kernel methods. They proved the uniform strong consistency and asymptotic normality of their estimator, but no rate of convergence was given. Fermanian (2005) also proposed a kernel-type estimator and studied the pointwise error, but he did not provide optimality results and imposed bandwidth conditions that are too restrictive in the copula context. Finally, Biau and Wegkamp (2005) proposed to estimate the copula density through a minimum distance criterion. Their estimator enjoys good properties but its computation entails non-trivial implementation issues that are left unaddressed.

Wavelets offer several advantages over alternative approaches in the present context:

(1) Wavelet decompositions can be computed easily and efficiently using the "Fast Wavelet Transform" (FWT) algorithm of Mallat (1989); ready-to-use wavelet packages are available in R, MatLab, etc.
(2) Wavelets handle automatically the boundary effects due to the fact that copula densities are often large along the borders of the unit square, yet zero outside this region; this is due to the good localization properties of the wavelet basis.
(3) In the specific case of copula estimation, the wavelet solution is natural and easy to understand; as will be seen, it provides automatically a collection of smoothings of the 3D-histogram, one for each level of resolution considered.

2   Multiresolution analysis in a nutshell

Wavelets have been applied successfully in different fields, such as signal and image processing, numerical analysis and geophysics, as well as statistics. Roughly speaking, the wavelet analysis of a two-place function $h(x, y)$ is a procedure by which this mapping can be decomposed simultaneously at an infinite number of resolution levels $j = 0, 1, \ldots$ The decomposition at arbitrary level $j_0 \in \mathbb{N}$ is given by

$$h(x, y) = h_{j_0}(x, y) + D_{j_0}h(x, y), \quad x, y \in \mathbb{R} \tag{3}$$

where

$$h_{j_0}(x, y) = \sum_{k \in \mathbb{Z}^2} \alpha_{j_0 k} \phi_{j_0 k}(x, y)$$

is a trend (or approximation) and

$$D_{j_0}h(x, y) = \sum_{j=j_0}^{\infty} \left(\sum_{k \in \mathbb{Z}^2} \beta_{jk}^{(1)} \psi_{jk}^{(1)}(x, y) + \sum_{k \in \mathbb{Z}^2} \beta_{jk}^{(2)} \psi_{jk}^{(2)}(x, y) + \sum_{k \in \mathbb{Z}^2} \beta_{jk}^{(3)} \psi_{jk}^{(3)}(x, y) \right)$$

is a sum of details of three types: vertical (1), horizontal (2), and oblique (3).

In this representation, the coefficients $\alpha_{j_0 k}$ and $\beta_{jk}^{(1)}$, $\beta_{jk}^{(2)}$ and $\beta_{jk}^{(3)}$ with $j \geq j_0$ are unique for every choice of $j_0 \in \mathbb{N}$. As for the functions $\phi_{j_0 k}$ and $\psi_{jk}^{(1)}$, $\psi_{jk}^{(2)}$ and $\psi_{jk}^{(3)}$, they are defined as follows

$$\phi_{jk_1 k_2}(x, y) = \phi_{jk_1}(x)\phi_{jk_2}(y), \quad \psi_{jk_1 k_2}^{(1)}(x, y) = \phi_{jk_1}(x)\psi_{jk_2}(y),$$

$$\tag{4}$$

$$\psi_{jk_1 k_2}^{(2)}(x, y) = \psi_{jk_1}(x)\phi_{jk_2}(y), \quad \psi_{jk_1 k_2}^{(3)}(x, y) = \psi_{jk_1}(x)\psi_{jk_2}(y),$$

in terms of a specific scaling function ϕ, and associated wavelet ψ and their location-scale transforms given by

$$\phi_{jk_3}(t) = 2^{j/2}\phi(2^j t - k_3) \quad \text{and} \quad \psi_{jk_3}(t) = 2^{j/2}\psi(2^j t - k_3)$$

for all $t \in \mathbb{R}$, $j \in \mathbb{N}$ and $k_3 \in \mathbb{Z}$. The functions ϕ and ψ must satisfy a number of technical conditions which ensure that the location-scale families they generate form an orthonormal system of L^2, the collection of square-integrable functions.

Each choice of pair (ϕ, ψ) leads to a different multiresolution analysis whose degree of regularity is as large as desired. Classical examples include the Haar, coiflet, Meyer and symlet families of wavelets. In this paper, ϕ and ψ are assumed to have a compact support $[0, L]$ as in the very common Daubechies wavelets (Daubechies, 1992). See Meyer (1992) for a general introduce to this theory.

The key feature of a wavelet representation is that the trend at level $j_0 + 1$ coincides with the trend at level j_0, supplemented with the horizontal, vertical and oblique details corresponding to level j_0. In other words,

$$h_{j_0+1} = h_{j_0} + \left(\sum_{k \in \mathbb{Z}^2} \beta_{j_0 k}^{(1)} \psi_{j_0 k}^{(1)} + \sum_{k \in \mathbb{Z}^2} \beta_{j_0 k}^{(2)} \psi_{j_0 k}^{(2)} + \sum_{k \in \mathbb{Z}^2} \beta_{j_0 k}^{(3)} \psi_{j_0 k}^{(3)} \right).$$

To illustrate the properties of the wavelet decomposition, consider for example the density of a Cauchy copula with Spearman correlation 0.6. This density is displayed in Figure 4, along with the wavelet approximations $h_4, \ldots, h_0$ corresponding to resolution levels $j_0 = 4, \ldots, 0$. To construct this series of approximations, the true copula c was identified with h_5 by setting $\alpha_{5k_1k_2} = c(k_1/2^5, k_2/2^5)$ for all $k_1, k_2 \in \{1, \ldots, 2^5\}$. The successive approximations $h_4, \ldots, h_0$ were then obtained by filtering h_5 with MatLab's wavelet toolbox, using the Daubechies(4) wavelet whose support is $[0, 3]$.

It is clear from Figure 4 that the approximation deteriorates as the resolution goes down. The characteristics of the underlying copula are gradually smoothed out. This is due to the fact that although it is regular in a theoretical sense, this copula density is heavy-tailed: it exhibits steep peaks at $(0, 0)$ and $(1, 1)$. If the filtering procedure had been applied to a Gaussian or Frank copula with an average level of dependence, the successive approximations would have been much closer to the original density.
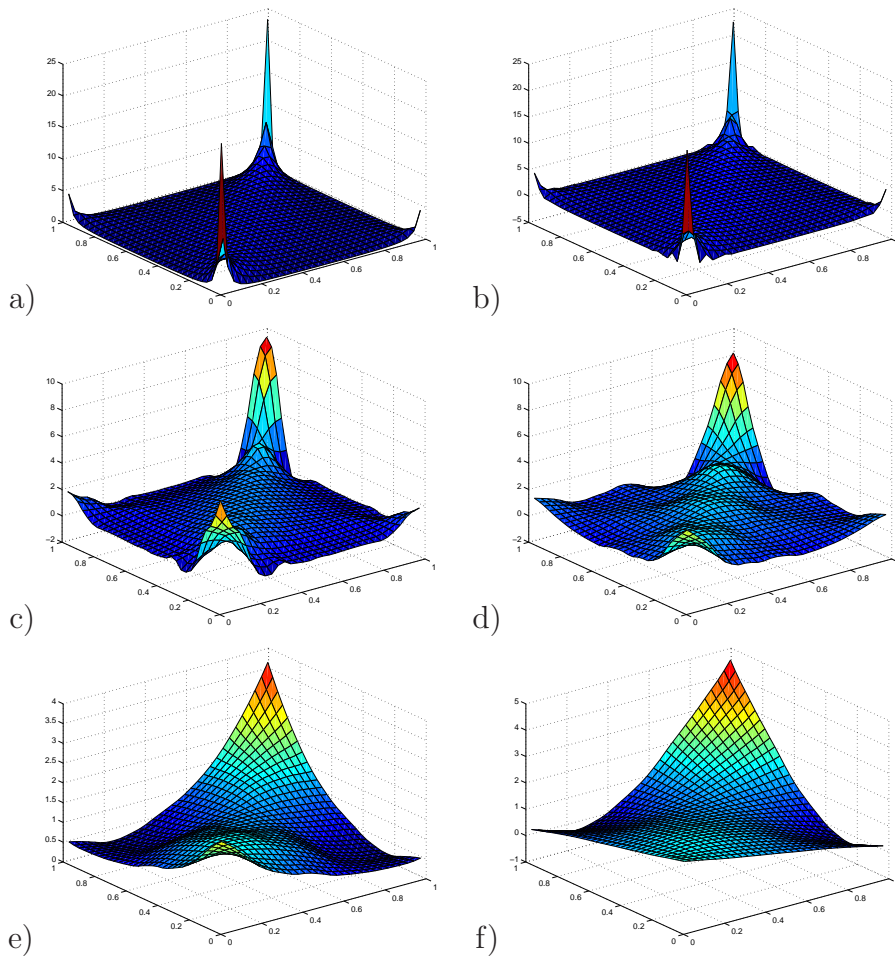

Fig. 4. Cauchy copula with Spearman correlation 0.6 and its Daubechies(4) wavelet approximations at levels $j = 5, \ldots, 1$ on a 32×32 grid.

6

Intuitively speaking, the more regular h is, the smaller are the details represented by $D_{j_0}h$, and hence the closer h_{j_0} is to h. In such cases, h can be approximated accurately by a trend h_{j_0} of small resolution j_0.

3 Estimation procedure

The wavelet decomposition leads naturally to an estimation procedure for the density of a copula. The idea is simple: starting from a jagged 3D-histogram of the pairs $(R_i/n, S_i/n)$ of normalized ranks, the graph is gradually made smoother through successive applications of the wavelet filter.

To be specific, the numerical procedure is as follows:

(1) Find the integer J for which $N = 2^J \leq \sqrt{n} < 2^{J+1}$. This integer represents the finest resolution level one can afford for a sample of size n.
(2) Compute the frequencies associated with the 3D-histogram on the regular grid $N \times N$ and identify them with the empirical scaling coefficients at resolution level J. In other words, set

$$\tilde{\alpha}_{Jk_1k_2} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1} \left\{ \frac{k_1-1}{N} < \frac{R_i}{n} \leq \frac{k_1}{N}, \quad \frac{k_2-1}{N} < \frac{S_i}{n} \leq \frac{k_2}{N} \right\}$$

for all $k_1, k_2 \in \{1, \ldots, N\}$.
(3) To avoid problems at the edges of the unit square, the $N \times N$ matrix $A = (\tilde{\alpha}_{Jk_1k_2})$ is expanded into a $3N \times 3N$ block matrix B obtained by symmetrizing A along all of its borders. Specifically,

$$B = \begin{pmatrix} {}_*A_* & A_* & {}_*A_* \\ A_* & A & A_* \\ {}_*A_* & A_* & {}_*A_* \end{pmatrix},$$

where

$$A_* = (\tilde{\alpha}_{Jk_1(N+1-k_2)}), \quad {}_*A = (\tilde{\alpha}_{J(N+1-k_1)k_2}), \quad {}_*A_* = (\tilde{\alpha}_{J(N+1-k_1)(N+1-k_2)}).$$

Various ways of handling edge effects are considered by Autin et al. (2008), who conclude that this strategy is preferable.
(4) Apply the Fast Wavelet Transform on B and extract the $(2,2)$ block of size $N \times N$ corresponding to the original A. A collection $(\tilde{c}_j)$ of estimates of c at

resolution levels $j = J, J - 1, \ldots, 0$ is then defined by setting

$$\tilde{c}_j(u, v) = \sum_{k \in \mathbb{Z}^2} \tilde{\alpha}_{jk} \phi_{jk}(u, v), \quad u, v \in (0, 1).$$

(5) Select $c^* = \tilde{c}_{J-1}$ as the estimator of the unknown copula density c.

This procedure is illustrated in Figure 5 for the same data used in Figure 3. In this case, $n = 2000$ and hence $J = 5$. It is clear that the wavelet filter is very good at smoothing out the histogram, even at resolution level $J - 1$. It is so efficient in fact that even at resolution level $J - 2$, the key features of the underlying copulas are beginning to fade out. As mentioned earlier, this is due to the relative lack of regularity of these two specific copulas, which have heavy tails. By contrast, the
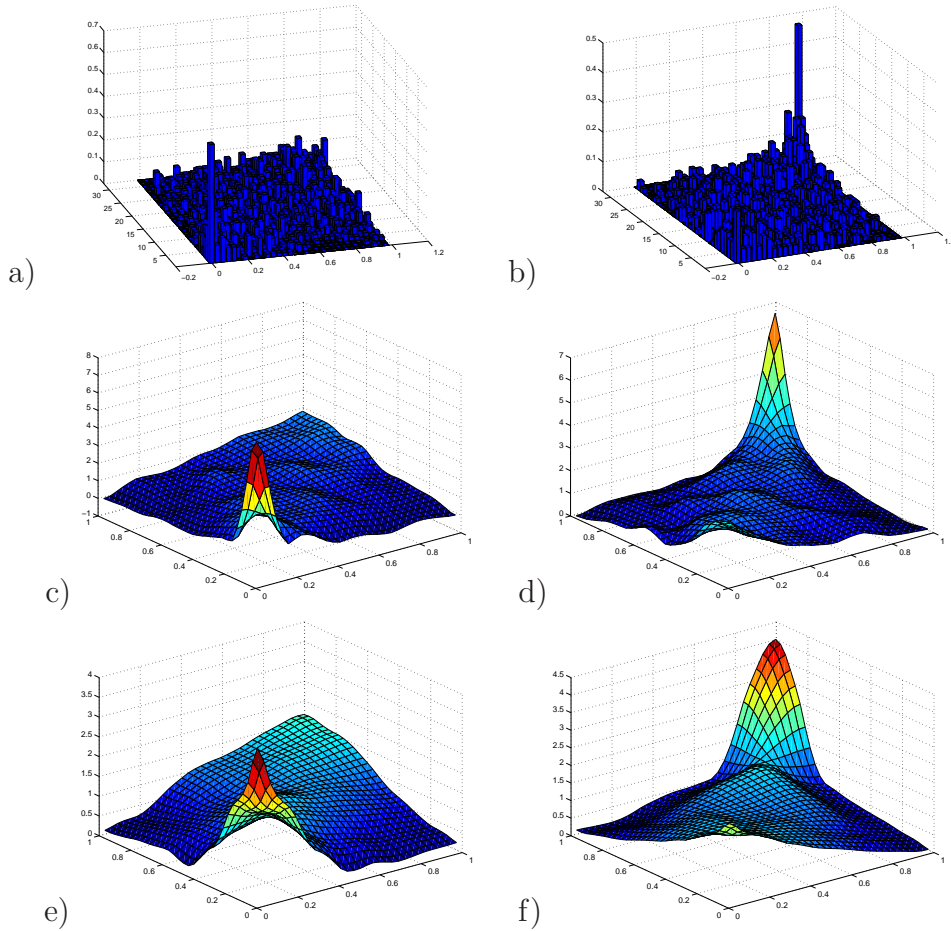

Fig. 5. Wavelet smoothing of $n = 2000$ pairs of normalized ranks from a Clayton(1) [top row] and a Gumbel(1.5) [bottom row]; panels a-b show the 3D-histograms on a $32 \times 32 = 2^J \times 2^J$ grid with $J = 5$; panels c)–d) and e)–f) show the wavelet estimators $\hat{c}_j$ at levels $j = J - 1$, $J - 2$, respectively.

wavelet estimators $\hat{c}_j$ with $j = J - 2, \ldots, J - 4$ would look very much like $\hat{c}_{J-1}$ when the data stem from the Gaussian or Frank copulas, for example. The reason for this phenomenon will be clear from the mathematical results that follow.

4   Mathematical framework

Let ϕ be a given scaling function, and let ψ be the associated wavelet. It is assumed henceforth that both functions are real-valued and have compact support $[0, L]$ for some $L > 0$. For every $j \in \mathbb{N}$, let ϕ_{jk}, $\psi_{jk}^{(1)}$, $\psi_{jk}^{(2)}$, and $\psi_{jk}^{(3)}$ be defined as in (4) for every $k = (k_1, k_2) \in \mathbb{Z}^2$. By construction, the set

$$\{\phi_{j_0 k}, \psi_{j\ell}^{(1)}, \psi_{j\ell}^{(2)}, \psi_{j\ell}^{(3)} : j \geq j_0, k \in \mathbb{Z}^2, \ell \in \mathbb{Z}^2\}$$

is an orthonormal basis of $L_2(\mathbb{R}^2)$ for arbitrary $j_0 \in \mathbb{N}$.

Given a copula density c, one can then expand it in the form (3) with

$$\alpha_{j_0 k} = \int_{(0,1)^2} c(u, v) \phi_{j_0 k}(u, v) dv du, \quad k \in \mathbb{Z}^2.$$

Observe that in view of (1), the change of variables $u = F(x)$ and $v = G(y)$ yields

$$\alpha_{j_0 k} = \int \phi_{j_0 k}(F(x), G(y)) h(x, y) dy dx = E_h\{\phi_{j_0 k}(F(X), G(Y))\},$$

where E_h denotes the expectation with respect to the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from density h. If the marginal distributions F and G were known, a natural (moment-based) estimator of $\alpha_{j_0 k}$ would then be given by

$$\hat{\alpha}_{j_0 k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{j_0 k}(F(X_i), G(Y_i)). \tag{5}$$

When F and G are unknown, a nonparametric analogue is obtained upon replacing F and G by their empirical counterparts, F_n and G_n. In view of relation (2), the estimator is thus rank-based, viz.

$$\tilde{\alpha}_{j_0 k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{j_0 k}(F_n(X_i), G_n(Y_i)) = \frac{1}{n} \sum_{i=1}^{n} \phi_{j_0 k}\left(\frac{R_i}{n}, \frac{S_i}{n}\right).$$

The wavelet-based estimator of c is then given by

$$\tilde{c}_{j_0}(u, v) = \sum_{k \in \mathbb{Z}^2} \tilde{\alpha}_{j_0 k} \phi_{j_0 k}(u, v), \quad u, v \in (0, 1) \tag{6}$$

9

in which the integer j_0 is the smoothing index of the method.

Note that just as Deheuvels' empirical copula is not a copula, $\tilde{c}_{j_0}$ is not necessarily a copula density. In particular, $\tilde{c}_{j_0}$ may sometimes be negative on parts of its domain and fail to integrate to 1. If in applications, an intrinsic copula density estimate is deemed necessary, it can be derived from $\tilde{c}_{j_0}$ by truncation and normalization.

From a numerical point of view, it is important to note that the sum over k in (6) is finite, because the wavelet lives on a compact support. As a result, one needs only compute $\lfloor 2^{2j_0} L^2 \rfloor$ terms in practice. In the special case where the copula density must be estimated at a single point $(u_0, v_0) \in (0,1)^2$, only $\lfloor L^2 \rfloor$ terms need be computed. For these reasons, the set on which k varies is left unspecified in the Appendix.

Of course, the performance of the procedure depends on the choice of the level j_0. The latter should be determined in some optimal way. The properties of the estimate with respect to the pointwise and quadratic loss functions are considered in the following section.

5 Optimality Results

The purpose of this section is to study the performance of $\tilde{c}_{j_0}$ as an estimator of the underlying copula density c. Two different loss functions will be used to this end. The first one is the mean integrated squared error (MISE), defined by

$$\mathrm{MISE}(\tilde{c}_{j_0}, c) = E_h \left[\int_0^1 \int_0^1 \{\tilde{c}_{j_0}(u, v) - c(u, v)\}^2 dv du \right].$$

The second loss function is the pointwise error at an arbitrary but fixed pair (u_0, v_0), i.e.,

$$\mathrm{PTW}(\tilde{c}_{j_0}, c) = E_h[\{\tilde{c}_{j_0}(u_0, v_0) - c(u_0, v_0)\}^2].$$

In view of the decomposition (3) for c, one can write

$$\mathrm{MISE}(\tilde{c}_{j_0}, c) = \mathrm{MISE}(\tilde{c}_{j_0}, c_{j_0}) + \int_0^1 \int_0^1 \{D_{j_0} c(u, v)\}^2 dv du,$$

$$\mathrm{PTW}(\tilde{c}_{j_0}, c) = \mathrm{PTW}(\tilde{c}_{j_0}, c_{j_0}) + \{D_{j_0} c(u_0, v_0)\}^2.$$

In each case, the first summand measures the error due to sampling while the second is a systematic error associated with the fact that the infinite-dimensional object c has been approximated by its trend at level j_0.

10

First consider the bias term arising from the approximation. This term can be bounded as soon as the copula density is sufficiently regular. Roughly speaking, regularity pertains to the size of the horizontal, vertical and oblique details in decomposition (3) for c. More precisely, suppose that c belongs to the ball of radius $M > 0$ in the Besov space B_s^p with parameters $s > 0$ and $p \geq 2$, or $s > 2/p - 1$ and $p \in [1, 2]$. For an introduction to Besov spaces, see Triebel (1992).

Proceeding as in the proof of Lemma 1 in Butucea and Tribouley (2006), one gets

$$\int_0^1 \int_0^1 \{D_{j_0} c(u, v)\}^2 dvdu \leq M\, 2^{-2j_0 s^*}, \tag{7}$$

where $s^* = s + 1 - 2/p$ if $p \in [1, 2]$ and $s^* = s$ otherwise. Moreover, the same bound applies to the pointwise error, but for the ball of radius M in the Besov space B_s^∞.

Now turn to the sampling error term. Introduce

$$\hat{c}_{j_0}(u, v) = \sum_{k \in \mathbb{Z}^2} \hat{\alpha}_{j_0 k} \phi_{j_0 k}(u, v), \quad u, v \in (0, 1)$$

with $\hat{\alpha}_{j_0 k}$ defined as in (5). This is the estimator that would be used if the marginal distributions F and G were known. One can then write

$$\mathrm{MISE}(\tilde{c}_{j_0}, c_{j_0}) \leq 2\, \mathrm{MISE}(\tilde{c}_{j_0}, \hat{c}_{j_0}) + 2\, \mathrm{MISE}(\hat{c}_{j_0}, c_{j_0}).$$

A similar decomposition is also valid for the pointwise loss.

Now it is well known that

$$\mathrm{MISE}(\hat{c}_{j_0}, c_{j_0}) \vee \mathrm{PTW}(\hat{c}_{j_0}, c_{j_0}) \leq K_1 \frac{2^{2j_0}}{n}, \tag{8}$$

for some constant $K_1 > 0$ depending only on ϕ and either

$$||c||_2 = \int c(u, v)^2 dvdu \quad \text{or} \quad ||c||_\infty = \sup_{u, v \in (0, 1)} |c(u, v)|,$$

as the case may be; see, e.g., Kerkyacharian and Picard (1992) for details. Thus to ensure that the variance of $\hat{c}_{j_0}$ tends to zero asymptotically, j_0 must be chosen so that $2^{j_0} \ll \sqrt{n}$.

The additional cost caused by the need to resort to rank statistics is studied in the following theorem, whose proof is given in the Appendix.

11

Theorem 1 *Let ϕ be a scaling function having m derivatives and for arbitrary resolution level $j_0 \in \mathbb{N}$, let $\tilde{c}_{j_0}$ be the estimator of a copula density c defined by (6). If j_0 is such that*

$$2^{j_0} \le \left(\frac{n}{\log(n)} \right)^{1/2 - 1/(2m)} < 2^{j_0 + 1},$$

then there exists a constant $K_2 > 0$ (depending only on ϕ and on either $||c||_2$ or $||c||_\infty$ as the case may be) such that

$$\text{MISE}(\tilde{c}_{j_0}, \hat{c}_{j_0}) \vee \text{PTW}(\tilde{c}_{j_0}, \hat{c}_{j_0}) \le K_2 \frac{2^{2j_0}}{n} \left(2^{2j_0} \frac{\log(n)}{n} + 2^{-j_0} \log(n) \right).$$

Note the slight additional restriction on j_0 involving m. The case $m = 2$ corresponds to the constraints on the bandwidth found by Fermanian and Scaillet (2005), who use a kernel-based estimation method.

Combining bounds $(7) - (8)$ with the result of Theorem 1, one can see that the error term associated with the use of ranks is negligible with respect to the usual error term as soon as $2^{j_0} \gg \log(n)$. Accordingly, the following result holds.

Theorem 2 *Fix $s > 0$ and $p \ge 2$ or $s > 2/p - 1$ and $p \in [1, 2]$. Denote by $B_s^p(M)$ the Besov ball of radius $M > 0$. Let also ϕ be a scaling function having m derivatives with $m > 1 + 1/s^*$. Given a sample of size n from a bivariate distribution whose underlying copula has density c, choose $j^* \in \mathbb{N}$ such that*

$$2^{j^*} \le n^{1/(2 + 2s^*)} < 2^{j^* + 1}$$

and denote by $\tilde{c}_{j^*}$ the estimator of c of resolution j^* given by (6) in terms of ϕ. Then there exists a constant $K > 0$ depending only on ϕ and M such that

$$\sup_{c \in B_s^p(M)} n^{s^*/(1 + s^*)} \text{MISE}(\tilde{c}_{j^*}, c) \le K.$$

The same bound is valid for the pointwise error, with $B_s^\infty(M)$ instead of $B_s^p(M)$.

The net result of this analysis is that under the hypotheses of Theorem 2, the procedure of estimation described above is *optimal* on the Besov ball $B_s^p(M)$ for the quadratic loss as well as on the Hölder ball $B_s^\infty(M)$ for the pointwise loss function.

6 Illustrations

Two applications of the proposed methodology are presented below. In each case, the graphs were produced in a matter of seconds with MatLab's wavelet toolbox.

These applications were selected to illustrate a number of practical points.

As a first example, consider the data of Frees and Valdez (1998), which were later reanalyzed by Genest et al. (1998), Klugman and Parsa (1999), Chen and Fan (2005) and Genest et al. (2006), among others. The data consist of the indemnity payment (LOSS) and the allocated loss adjustment expense (ALAE) for 1500 general liability claims. As in all previous studies, the 34 claims subject to censoring were discarded.

Six graphical representations of these data are provided in Figure 6. In the top row, panels a), b), and c) show the original data on the logarithmic scale, the rank-rank plot, and the 3D-histogram based on a 16×16 grid, respectively. The bottom row shows three copula estimates. They are:

(1) panel d): the raw wavelet-based copula density estimate, i.e., without boundary correction;
(2) panel e): the copula density estimate c^* (with boundary correction) described in Section 3;
(3) panel f): the Gumbel copula density whose parameter estimate $\hat{\tau} = 0.31$ is the empirical value of Kendall's tau.

The need to apply a boundary correction to the wavelet-based estimation procedure is apparent from Figure 6. The difficulty with the uncorrected estimator is that it misses peaks along the boundaries of the unit square. This is due to the fact that the
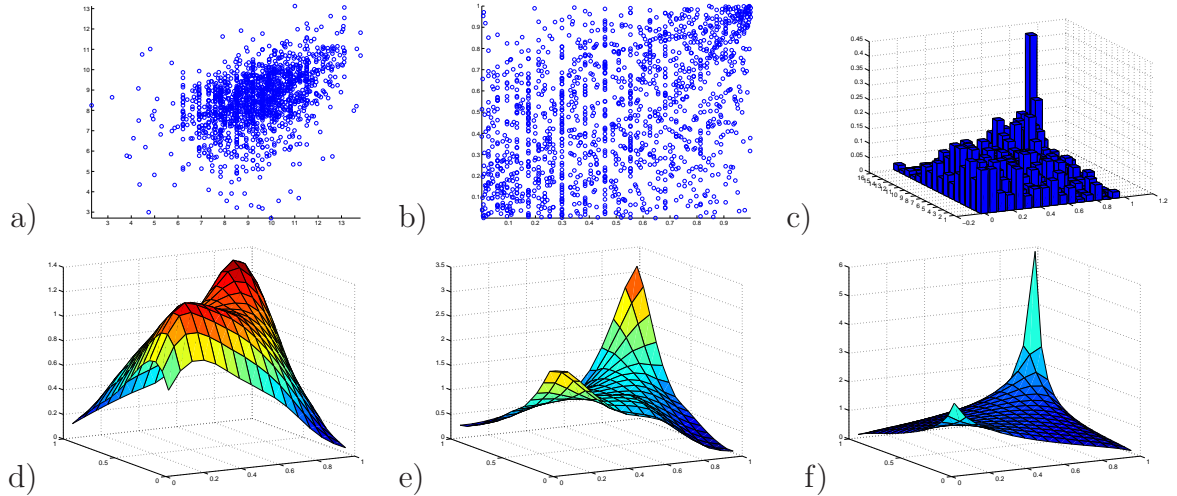

Fig. 6. Six graphical representations of the LOSS/ALAE data of Frees and Valdez (1998). Panel a): Original data on the logarithmic scale; panel b): rank-rank plot; panel c): 3D-histogram on a 16×16 grid; panel d): wavelet-based copula density estimate without boundary correction; panel e) recommended wavelet-based copula density estimate c^* (with boundary correction); panel f) Gumbel copula with parameter 1.4.

13

nonparametric procedure tries to produce a smooth copula estimate both within and outside the unit square (where the copula vanishes). Once the boundary correction is applied, however, the wavelet-based estimate c^* resembles the Gumbel copula density with parameter $\tau = 0.31$. The advantage of the wavelet-based estimator is that it suggests this choice without having to make parametric assumptions. In other words, while the graph in panel f) represents a good estimator under the assumption that the copula is from the Gumbel family, the wavelet-based procedure does not need to make any such assumption to arrive at the same conclusion.

The various authors who analyzed this data set concluded that the Gumbel copula provides an adequate representation of the underlying dependence structure. This parametric family of extreme-value copulas captures the fact that almost all large indemnity payments generate important adjustment expenses (e.g., investigation and legal costs) while the effort invested in the treatment of a small claim is more variable. Accordingly, the copula exhibits positive but asymmetric dependence.

The second example uses time-series data on the daily prices for one-month-ahead futures on light sweet crude oil and natural gas. The log-returns for these two commodities were analyzed by Grégoire et al. (2008) for the period extending from July 1, 2003 to July 19, 2006. They found that once standardized to mean 0 and variance 1, the log-returns X_t for oil could be treated as a random sample from a Student distribution with 13.745 degrees of freedom. The log-returns for gas, however, exhibit heteroscedasticity; they can be modeled by a GARCH$(1, 6)$ model whose residuals Y_t follow a skewed t-distribution (Azzalini and Capitanio, 2003).

Of special interest in this application is the dependence between the two commodities, once the effect of time has been filtered out. This dependence is represented by the pairs of ranks derived from the residuals (X_t, Y_t). Nine graphical representations of these data are provided in Figure 7. In the top row, panels a), b), and c) show the pairs (X_t, Y_t), the associated rank-rank plot, and the 3D-histogram based on a 16×16 grid, respectively. The second row shows the proposed copula density estimator c^* using three different wavelets: the Haar, Daubechies(4) and Adelson wavelet, corresponding to panels d), e) and f), respectively. Finally, panels g)–i) on the third row exhibit three parametric copula densities: the Frank copula with parameter 3.22, the Gaussian copula with parameter $r = 0.522$, and the Student copula with $r = 0.522$ and 22 degrees of freedom, respectively. The latter is the copula ultimately selected by Grégoire et al. (2008) on the basis of goodness-of-fit tests, although the Gaussian and Frank copulas were mentioned as very close contenders.
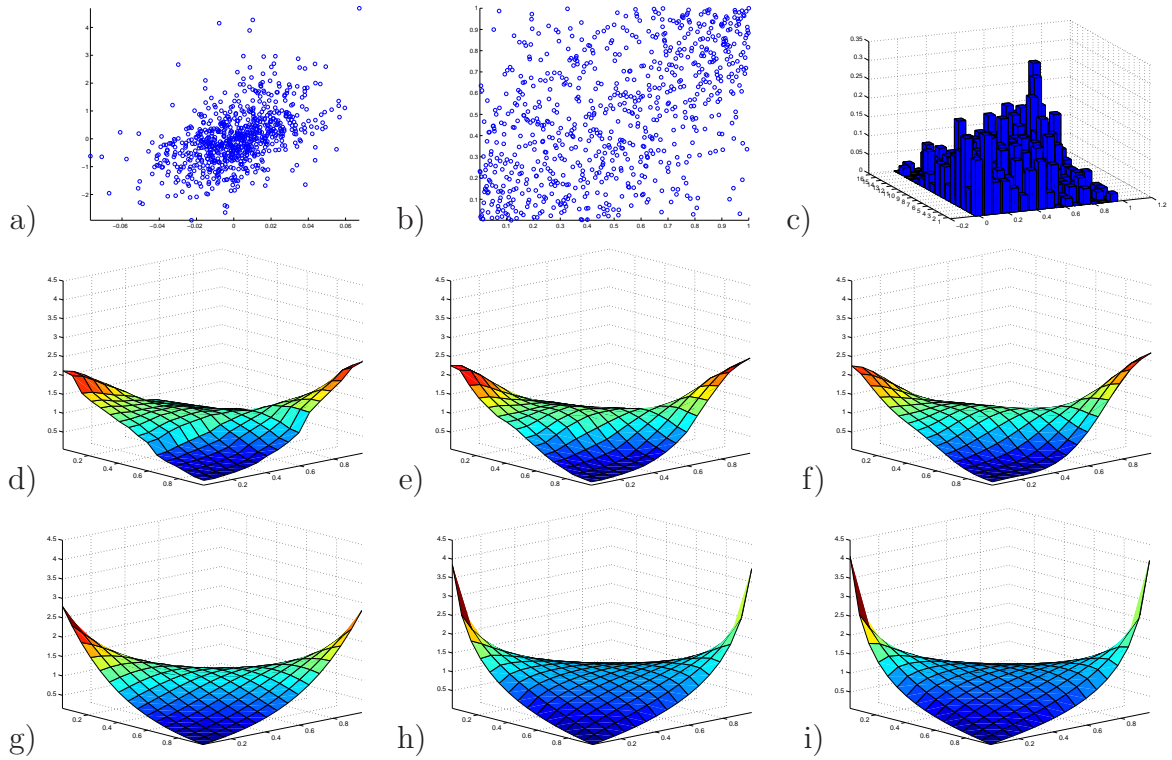
Fig. 7. Nine graphical representations of the oil/gas data of Grégoire et al. (2008). Panel a): Pairs of residuals after time-series filtering; panel b): rank-rank plot; panel c): 3D-histogram on a 16×16 grid; panels d)–f): copula density estimates with Haar, Daubechies(4) and Adelson wavelet, respectively; panel g): Frank copula with parameter 3.22; panel h): Gaussian copula with parameter 0.522; panel i) Student copula with parameter 0.522 and 22 degrees of freedom.

Three points are apparent from the graphs:

(1) in this specific application, the 3D-histogram conveys very little information beyond the fact that most of the probability mass accumulates along the main diagonal $(u = v)$;

(2) there is little to choose between the three wavelets: in most applications, they yield essentially the same estimator, although the density associated with Haar wavelet tends to be a little more jagged (in that sense, the wavelet plays much the same role as the kernel in classical density estimation);

(3) the third row confirms that the three choices of parametric copulas considered by Grégoire et al. (2008) are acceptable and quite similar in many respects (although the Student copula has heavier tail behavior than the other two).

15

As a further confirmation of point 3, Table 1 reports the average squared error

$$\text{ASE}(c^*, c_{\hat{\theta}}) = \frac{1}{M^2} \sum_{\ell_1=1}^{M} \sum_{\ell_2=1}^{M} \left\{ c^* \left(\frac{\ell_1}{M}, \frac{\ell_2}{M} \right) - c_{\hat{\theta}} \left(\frac{\ell_1}{M}, \frac{\ell_2}{M} \right) \right\}^2$$

and the relative average squared error

$$\text{RASE}(c^*, c_{\hat{\theta}}) = \text{ASE}(c^*, c_{\hat{\theta}}) \bigg/ \frac{1}{M^2} \sum_{\ell_1=1}^{M} \sum_{\ell_2=1}^{M} c_{\hat{\theta}}^2 \left(\frac{\ell_1}{M}, \frac{\ell_2}{M} \right),$$

between the recommended wavelet estimator c^* and a parametric copula density $c_{\hat{\theta}}$ in which θ has been estimated by a rank-based method. In Table 1, $M = 16$ and the parametric copula estimates are those considered by Grégoire et al. (2008). The copulas are listed in order of preference, based on the ASE and RASE criteria. It is clear that the Gumbel and Clayton dependence structures can be ruled out. It is more difficult to choose between the other three.

These findings are consistent with the conclusions of Grégoire et al. (2008). On the basis of goodness-of-fit tests, their order of preference is

$$\text{Student}(22) \succ \text{Gaussian} \succ \text{Frank} \succ \text{Gumbel} \succ \text{Clayton};$$

the associated P-values (in percentage) are 3.00, 1.40, 0.95, 0.75 and 0, respectively. This slight discrepancy may be due to the fact that the procedure on which c^* is based tends to oversmooth high peaks.

One way or another, it is obvious from the various graphical displays that throughout the study period, the prices of oil and natural gas exhibited a strong dependence, particularly in the tails. In addition, the association between the residuals from the time-series models is clearly positive and approximately symmetric. That these commodities should be increasingly dependent is not surprising, given that both natural gas and crude oil are used to generate electricity and heating. In addition, natural gas is used to extract oil from tar sands.

Copula	Parameter	ASE	RASE ($\times 100$)
Frank	3.474	0.0184	1.43
Gaussian	0.522	0.0406	3.14
Student(22)	0.522	0.0480	3.69
Gumbel	1.538	0.1306	9.14
Clayton	1.076	0.2795	18.08

Table 1
Average squared error (ASE) and relative average squared error (RASE) between the proposed wavelet estimator c^* and five copulas with specific choices of parameter.

7 Discussion

This paper has proposed a rank-based procedure for the estimation of a copula density based on the method of wavelets. The estimator is easy to implement, e.g., with ready-to-use wavelet toolboxes available in MatLab or R. It involves only two "strategic" choices: a wavelet function and a multiresolution level j^*. It was seen through an example that the choice of wavelet is fairly inconsequential. In the algorithm of Section 3, j^* was arbitrarily fixed at level $J-1$, and this was seen to produce acceptable results, except when the underlying dependence structure exhibits very high peaks. Although it was shown in Section 4 that the estimator $c^* = c_{j^*}$ is optimal in a minimax sense, the choice of j^* depends on a priori information about the degree of regularity, s, of the unknown copula density.

To circumvent these difficulties, one could rely on non-linear wavelet estimation methods. In addition to estimating the trend, these adaptive techniques take the details into account whenever they are deemed to be sufficiently important. This makes it possible to capture with more accuracy the high peaks of the underlying density since details are important in that neighborhood. Different thresholding techniques adapted to the copula density estimation problem are considered in the companion paper by Autin et al. (2008).

While a wavelet-based estimate of a copula density is useful for describing the underlying dependence structure, it does not eliminate the need for parametric copula models, particularly for prediction purposes. It does, however, provide an additional graphical tool for model selection. In future work, one might also consider using a wavelet-based copula density estimate c^* as a benchmark for goodness-of-fit testing. A simple solution would be to select the parametric copula which is closest to c^* according to the ASE criterion. As it turns out, however, $\mathrm{ASE}(c^*, c_\theta)$ is a biased estimate of $\|c_{j^*} - c_\theta\|_2^2$, so additional work along these lines is required. In addition, comparisons would need to be done with existing goodness-of-fit tests for copula distributions surveyed by Genest et al. (2008).

Appendix

Because the wavelet basis is orthonormal, Theorem 1 is an obvious consequence of the following proposition.

Proposition 1 *Assume that the scaling function ϕ is $m-$differentiable and let j be the integer such that*
$$2^j \leq (n/\log(n))^{1/2-1/2m} < 2^{j+1}.$$
Then there exists a constant $K > 0$ such that for given $(u_0, v_0) \in (0, 1)^2$, one has

$$E_h \left[\sum_k (\tilde{\alpha}_{jk} - \hat{\alpha}_{jk})^2 \right] \vee E_h \left[\sum_k (\tilde{\alpha}_{jk} - \hat{\alpha}_{jk}) \phi_{jk}(u_0, v_0) \right]^2$$
$$\leq K \frac{2^{2j}}{n} \left(2^{2j} \frac{\log(n)}{n} + 2^{-j} \log(n) \right).$$

The proof of this result is given below. It relies on two key points:

(1) It follows from the localization of the wavelet basis that the estimation of the scaling coefficients only depends on a small number of observations: for fixed k, only $n2^{-j}$ of the n available data points are needed, as will be seen in the first step of the proof.
(2) The wavelet basis may be chosen to be sufficiently regular to control the distance both between $F_n(X_i)$ and $F(X_i)$, and between $G_n(Y_i)$ and $G(Y_i)$. Through a Taylor series expansion, this leads to the desired result for almost all levels j.

For clarity, the following argument is presented in the bivariate context, as in the rest of this article. Note, however, that a d-variate generalization is possible, using multivariate expansions.

7.1 Proof of Proposition 1

For arbitrary $k = (k_1, k_2) \in \mathbb{Z}^2$, introduce

$$\xi_k(X_i, Y_i) = \phi_{jk}(F_n(X_i), G_n(Y_i)) - \phi_{jk}(F(X_i), G(Y_i)).$$

Observe that this quantity may be written alternatively as

$$\xi_k(X_i, Y_i) = \xi_{k_1}(X_i)\xi_{k_2}(Y_i) + \xi_{k_1}(X_i)\phi_{jk_2}(G(Y_i)) + \xi_{k_2}(Y_i)\phi_{jk_1}(F(X_i)) \tag{9}$$

18

in terms of

$$\xi_{k_1}(X_i) = \phi_{jk_1}(F_n(X_i)) - \phi_{jk_1}(F(X_i)) \quad \text{and} \quad \xi_{k_2}(Y_i) = \phi_{jk_2}(G_n(Y_i)) - \phi_{jk_2}(G(Y_i)).$$

The proof of the theorem requires the evaluation of

$$E_h\left[\sum_k (\tilde{\alpha}_{jk} - \hat{\alpha}_{jk})^2\right] = \sum_k E\left(\frac{1}{n}\sum_{i=1}^n \xi_k(X_i, Y_i)\right)^2 \tag{10}$$

and

$$E_h\left(\sum_k (\tilde{\alpha}_{jk} - \hat{\alpha}_{jk})\, \phi_{jk}(u_0, v_0)\right)^2$$

$$\leq \sum_{k,p}\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{\ell=1}^n E_h|\xi_k(X_i, Y_i)\xi_p(X_\ell, Y_\ell)|\right) |\phi_{jk}(u_0, v_0)\phi_{jp}(u_0, v_0)|. \tag{11}$$

7.1.1   First step: Finding bounds on the number of terms in the sums over i

In equations (10) and (11), the right-hand terms involve a sum over $i \in \{1, \ldots, n\}$. The scaling function ϕ being well localized, the number of non-zero terms in each of these sums is of a smaller order than n. To find a sharper bound on the size of $I = \{i : \xi_k(X_i, Y_i) \neq 0\}$, use the expansion (9). Clearly, one has

$$|I| \leq |\{i : \xi_{k_1}(X_i) \neq 0, \quad \xi_{k_2}(Y_i) \neq 0\}|.$$

Observe that for each $i \in \{1, \ldots, n\}$, $U_i = F(X_i)$ is a uniform random variable on $(0, 1)$. Let $X_{(1)} < \cdots < X_{(n)}$ be the ordered sample, and set $U_{(i)} = F(X_{(i)})$. The problem reduces to finding the size of the set

$$I(j) = \left\{i : \xi_{k_1}(X_{(i)}) = \phi_{jk_1}\left(\frac{i}{n}\right) - \phi_{jk_1}(U_{(i)}) \neq 0\right\}.$$

Fix $\epsilon > 0$. Clearly, $I(j) \subset I(j, 0) \cup I_1(j, \epsilon) \cup I_2(j, \epsilon)$ with

$$I(j, 0) = \left\{i : \phi_{jk_1}\left(\frac{i}{n}\right) \neq 0\right\},$$

$$I_1(j, \epsilon) = \left\{i : \phi_{jk_1}\left(\frac{i}{n}\right) = 0, \quad \phi_{jk_1}(U_{(i)}) \neq 0, \quad \left|U_{(i)} - \frac{i}{n}\right| < \epsilon\right\},$$

$$I_2(j, \epsilon) = \left\{i : \phi_{jk_1}\left(\frac{i}{n}\right) = 0, \quad \phi_{jk_1}(U_{(i)}) \neq 0, \quad \left|U_{(i)} - \frac{i}{n}\right| \geq \epsilon\right\}$$

$$\subset \left\{i : |F_n(X_{(i)}) - F(X_{(i)})| \geq \epsilon\right\}.$$

19

Recall that the support of ϕ is $[0, L]$ and write $I_{jk_1} = [k_1 2^{-j}, (k_1 + L)2^{-j}]$ for the support of ϕ_{jk_1}. It follows that $|I(j, 0)| = \lfloor L\,n2^{-j} \rfloor$ and $|I_1(j, \epsilon)| = \lfloor 2n\epsilon \rfloor$. A finer bound than n is not needed for the size of $I_2(j, \epsilon)$ because ϵ will be chosen in such a way that for every i, the probability of the event $\{|F_n(X_{(i)}) - F(X_{(i)})| \geq \epsilon\}$ is as small as desired. Since

$$\left| \sum_{i=1}^{n} \xi_k(X_i, Y_i) \right| \leq \left| \sum_{i \in I(j,0) \cup I_1(j,\epsilon)} \xi_k(X_i, Y_i) + \sum_{i \in I_2(j,\epsilon)} \xi_k(X_i, Y_i) \right|$$

$$\leq \sum_{i \in I(j,0) \cup I_1(j,\epsilon)} |\xi_k(X_i, Y_i)| + n\, \mathbf{1}_{\{\|\hat{F} - F\|_\infty > \epsilon\}}$$

the Dvoretzky–Kiefer–Wolfowitz inequality can be invoked with $\epsilon = (\delta \log(n)/(2n))^{1/2}$ and δ as large as desired. Therefore, one gets

$$E_h \left(\frac{1}{n} \sum_{i=1}^{n} \xi_k(X_i, Y_i) \right)^2 \leq 2 E_h \left(\frac{1}{n} \sum_{i \in I_1(j,0) \cup I_1(j,\epsilon)} |\xi_k(X_i, Y_i)| \right)^2 + 2\, n^{-\delta}.$$

Observe that this choice of ϵ leads to $|I(j, 0) \cup I_1(j, \epsilon)| \leq \lfloor (L + 2\sqrt{\delta})n2^{-j} \rfloor$.

7.1.2 Second step: Finding bounds on the expected values

The following lemma is the key tool for determining upper bounds on the right-hand terms of equations (10) and (11). The proof of the lemma itself is deferred to the end of this section.

Lemma 1 *Assume that the scaling function ϕ is $m-$differentiable for some integer m. Then there exists a constant $K > 0$, depending on ϕ and on $\|c\|_2$, such that for all $i \neq \ell$ and $j \in \mathbb{N}$ satisfying $2^j \leq (n/\log(n))^{1/2 - 1/2m} < 2^{j+1}$, one has*

$$|E_h \xi_k(X_i, Y_i)\xi_p(X_i, Y_i)| \ \vee \ \sum_k |E_h \xi_k(X_i, Y_i)\xi_k(X_\ell, Y_\ell)| \leq K\, 2^{3j} \left(\frac{\log(n)}{n} \right).$$

If in addition the copula density c is uniformly bounded, then for all $i \neq \ell$,

$$E_h \left[\xi_k(X_i, Y_i)\xi_p(X_\ell, Y_\ell) \right] \leq K \frac{\log(n)}{n} \ \text{ and } \ E_h \left[\xi_k(X_i, Y_i)\xi_p(X_i, Y_i) \right] \leq K\, 2^{2j} \frac{\log(n)}{n},$$

where $K > 0$ is another constant that now depends on $\|c\|_\infty$ also.

Since the support of the scaling function is compact, there are at most $\lfloor L^2 2^{2j} \rfloor$ terms in the sums over k appearing in the right-hand terms of equation (10). Accordingly, one can find a constant $K > 0$ such that

$$E_h \left[\sum_k (\tilde{\alpha}_{jk} - \hat{\alpha}_{jk})^2 \right]$$

$$\leq \frac{1}{n^2} \sum_k \sum_{i \in I(j,0) \cup I_1(j,\epsilon)} E_h \left(\xi_k(X_i, Y_i) \right)^2$$

$$+ \frac{1}{n^2} \sum_k \sum_{i \neq \ell \in I(j,0) \cup I_1(j,\epsilon)} |E_h \xi_k(X_i, Y_i) \xi_k(X_\ell, Y_\ell)| + K\, 2^{2j} n^{-\delta}$$

$$\leq K \frac{1}{n^2} 2^{2j} (n2^{-j}) 2^{3j} \frac{\log(n)}{n} + K \frac{1}{n^2} (n2^{-j})^2 2^{3j} \frac{\log(n)}{n} + K\, n^{-\delta+1},$$

which leads to the result stated in Theorem 1. Moreover, for the pointwise error, there are then only $\lfloor L^2 \rfloor$ terms in the sum over k such that $\phi_{jk}(u_0, v_0)$ is non-zero. This yields

$$E_h \left(\sum_k (\tilde{\alpha}_{jk} - \hat{\alpha}_{jk}) \phi_{jk}(u_0, v_0) \right)^2$$

$$\leq \frac{1}{n^2} \sum_{k,p} \sum_{i \in I(j,0) \cup I_1(j,\epsilon)} |E_h \xi_k(X_i, Y_i) \xi_p(X_i, Y_i)| \, \|\phi_{jk} \phi_{jp}\|_\infty$$

$$+ \frac{1}{n^2} \sum_{k,p} \sum_{i \neq l \in I(j,0) \cup I_1(j,\epsilon)} |E_h \xi_k(X_i, Y_i) \xi_p(X_\ell, Y_\ell)| \, \|\phi_{jk} \phi_{jp}\|_\infty + K\, n^{-\delta}$$

$$\leq K \frac{1}{n^2} (n2^{-j}) 2^{3j} \frac{\log(n)}{n} 2^{2j} + K \frac{1}{n^2} (n2^{-j})^2 \frac{\log(n)}{n} 2^{2j} + K\, n^{-\delta+1},$$

for some other constant $K > 0$. Thus the proof of Theorem 1 is complete.

7.2 Proof of Lemma 1

Let $\Delta(X_i) = F_n(X_i) - F(X_i)$. Assuming that ϕ is continuously $m-$differentiable and denoting $\phi^{(r)}$ its derivative of order r, one can write

$$\xi_{k_1}(X_i) = \sum_{r=1}^{m-1} \hat{z}_{k_1}^r(X_i) + \hat{w}_{k_1}(X_i),$$

where

$$\hat{z}_{k_1}^r(X_i) = 2^{jr}\Delta(X_i)^r\phi_{jk_1}^{(r)}(F(X_i)),$$

$$\hat{w}_{k_1}(X_i) = 2^{jm}\int_{F_n(X_i)}^{F(X_i)}\phi_{jk_1}^{(m)}(t)(F_n(X_i) - t)^{m-1}dt.$$

For fixed $\epsilon > 0$, write

$$\Delta(X_i) = \Delta(X_i)\left(\mathbf{1}_{|\Delta(X_i)|\leq\epsilon} + \mathbf{1}_{|\Delta(X_i)|>\epsilon}\right).$$

Denote by LD a quantity bounded above by $K\,2^{jm'}\,\mathbf{1}_{|\Delta(X_i)|>\epsilon}$ for some integer $m' \leq m$ and some $K > 0$ depending on the sup-norm of the derivative of ϕ. Then

$$|\hat{z}_{k_1}^r(X_i)| \leq 2^{jr}\epsilon^r|\phi_{jk_1}^{(r)}(F(X_i))| + LD,$$

$$|\hat{w}_{k_1}(X_i)| \leq \|\phi^{(m)}\|_\infty 2^{(m+1/2)j}\epsilon^m + LD.$$

Now assume that ϵ is such that $\epsilon \leq 2^{-j}$. Then

$$|\xi_{k_1}(X_i)| \leq K\left(2^{(m+1/2)j}\epsilon^m + 2^j\epsilon\max_{r\in\{1,\ldots,m-1\}}|\phi_{jk_1}^{(r)}(F(X_i))|\right) + LD, \tag{12}$$

for some other constant $K > 0$. Obviously, the same kind of expansion holds when the variable Y_i is considered. Using the expansion (9), one finds

$$\xi_k(X_i, Y_i)\xi_p(X_\ell, Y_\ell) = T_1 + T_2 + T_3,$$

where

$$T_1 = \xi_{k_1}(X_i)\xi_{k_2}(Y_i)\xi_{p_1}(X_\ell)\xi_{p_2}(Y_\ell),$$

T_2 is a sum of four terms of the type

$$TT_2 = \xi_{k_1}(X_i)\xi_{k_2}(Y_i)\xi_{p_1}(X_\ell)\phi_{jp_2}(G(Y_\ell)),$$

and T_3 is a sum of four terms of the type

$$TT_3 = \xi_{k_1}(X_i)\xi_{p_1}(X_\ell)\phi_{jk_2}(G(Y_i))\phi_{jp_2}(G(Y_\ell)).$$

7.2.1   Case where c is possibly unbounded

Using Hölder's Inequality, one can compute easily a bound for the expectation terms, viz.

$$E_h[|\phi_{jk_1}^{(a_1)}(F(X))|^{b_1}|\phi_{jk_2}^{(a_2)}(G(Y))|^{b_2}] \leq \||\phi_{jk_1}^{(a_1)}(u)|^{b_1}|\phi_{jk_2}^{(a_2)}(v)|^{b_2}\|_2\|c\|_2$$
$$\leq 2^{j(b_1/2+b_2/2-1)}\|c\|_2 \|\phi^{(a_1)}\|_2 \|\phi^{(a_2)}\|_2,$$

where $b_1, b_2 \in (0, \infty)$ are arbitrary. Observe that if b_1 or b_2 vanishes, the bound is better because the expectation is taken over the marginal of c, which is bounded, viz.

$$E_h|\phi_{jk_1}^{(a_1)}(F(X))|^{b_1} \leq 2^{j(b_1/2-1)} \|\phi^{(a_1)}\|_1.$$

Let K be some new constant only depending on the L_1-norm of the derivatives of ϕ and on $\|c\|_2$. Using (12), an upper bound on $|E_h(T_1)|$ is then given by

$$K\left(2^{(4m+2)j}\epsilon^{4m} + 2^{(3m+2)j}\epsilon^{3m+1} + 2^{(2m+3)j}\epsilon^{2m+2} + 2^{(m+4)j}\epsilon^{m+3} + 2^{5j}\epsilon^4 + E_h(LD)\right)$$

when $i = \ell$, and by

$$K\left(2^{(4m+2)j}\epsilon^{4m} + 2^{(3m+2)j}\epsilon^{3m+1} + 2^{(2m+3)j}\epsilon^{2m+2} + 2^{(m+3)j}\epsilon^{m+3} + 2^{4j}\epsilon^4 + E_h(LD)\right)$$

when $i \neq \ell$. Similarly, one gets

$$|E_h(TT_2)| \leq K\left(2^{(3m+1)j}\epsilon^{3m} + 2^{(2m+3)j}\epsilon^{2m+1} + 2^{(m+3)j}\epsilon^{m+2} + 2^{4j}\epsilon^3 + E_h(LD)\right)$$

when $i = \ell$ and

$$|E_h(TT_2)| \leq K\left(2^{(3m+1)j}\epsilon^{3m} + 2^{(2m+3)j}\epsilon^{2m+1} + 2^{(m+2)j}\epsilon^{m+2} + 2^{3j}\epsilon^3 + E_h(LD)\right)$$

when $i \neq \ell$. Furthermore,

$$|E_h(TT_3)| \leq K\left(2^{2mj}\epsilon^{2m} + 2^{(m+2)j}\epsilon^{m+1} + 2^{3j}\epsilon^2 + E_h(LD)\right)$$

when $i = \ell$. When $i \neq \ell$, the term TT_3 has to be studied carefully in order to avoid a logarithmic term. To this end, the upper bound (12) must be replaced with the following equality

$$\xi_{k_1}(X_i) = \hat{z}_{k_1}^1(X_i) + R_{k_1}(X_i)$$

with

$$|R_{k_1}(X_i)| \leq K\left(2^{(m+1/2)j}\epsilon^m + 2^{2j}\epsilon^2 \max_{r \in \{2, \ldots, m-1\}} |\phi_{jk_1}^{(r)}(F(X_i))|\right) + LD$$

and yet another constant $K > 0$. As a result, one finds

$$E_h(TT_3) = E_h[\hat{z}_{k_1}^1(X_i)\hat{z}_{p_1}^1(X_\ell)\phi_{jk_2}(G(Y_i))\phi_{jp_2}(G(Y_\ell))] + R = t_k t_p + R,$$

23

where
$$|R| \leq K \left(2^{2mj}\epsilon^{2m} + 2^{(m+1)j}\epsilon^{m+1} + E_h(LD) \right)$$
for some new constant $K > 0$, and
$$t_k = 2^j \int \Delta(F^{-1}(u))\phi_{jk_1}^{(1)}(u)\ \phi_{jk_2}(v)c(u,v)dvdu.$$

Now in view of the fact that the wavelet basis is orthogonal, note that Parseval's Identity implies that
$$\sum_{k_2} \left(\int \Phi(v)\phi_{jk_2}(v)dv \right)^2 \leq \int \Phi^2(v)dv,$$
for any function $\Phi \in L_2(\mathbb{R})$. Applying this inequality to
$$\Phi_{uk_1}(v) = \int \Delta(F^{-1}(u))\phi_{jk_1}^{(1)}(u)c(u,v)du,$$
it follows that for $k = p$,

$$\begin{aligned}
\sum_k t_k^2 &= 2^{2j}\sum_{k_1}\sum_{k_2}\left(\int \Phi_{uk_1}(v)\phi_{jk_2}(v)dv \right)^2 \leq 2^{2j}\sum_{k_1}\left(\int \Phi_{uk_1}^2(v)dv \right) \\
&\leq 2^{2j}\sum_{k_1}\int \left[\int \Delta(F^{-1}(u))\phi_{jk_1}^{(1)}(u)c(u,v)du \right]^2 dv \\
&\leq 2^{2j}\|\Delta\|_\infty^2 \sum_{k_1}\int |\phi_{jk_1}^{(1)}(u)|^2 du \int c^2(u,v)dvdu,
\end{aligned}$$

where the last step is a consequence of the Cauchy–Schwartz Inequality.

In all of the above, the term LD may vary from line to line but remains bounded above by a constant $K > 0$ times
$$2^{jm'}(\mathbf{1}_{|F_n(X_i)-F(X_i)|>\epsilon} + \mathbf{1}_{|\hat{G}(Y_i)-G(Y_i)|>\epsilon}).$$
Applying the Dvoretzky–Kiefer–Wolfowitz Inequality, one finds

$$\begin{aligned}
E_h(LD) &\leq K2^{jm'}\left[P\left(|F_n(X_i) - F(X_i)| > \epsilon\right) + P\left(|G_n(Y_i) - G(Y_i)| > \epsilon\right) \right] \\
&\leq K2^{jm'}\left[P\left(\|F_n - F\|_\infty > \epsilon\right) + P\left(\|G_n - G\|_\infty > \epsilon\right) \right] \\
&\leq Kn^{-2\delta},
\end{aligned}$$

where δ could be as large as desired, provided that $\epsilon \geq (\delta'\log(n)/n)^{1/2}$ for $\delta' > m'/2 + 2\delta$. This is compatible with the condition on j given in the statement of Lemma 1.

Take such an ϵ. If $2^j \leq (n/\log(n))^{1/2-1/2m}$, the larger term is $2^{3j}\epsilon^2$ when $i = \ell$. If $i \neq \ell$, the larger term given by t_k is of the same order. This completes the proof of the first part of Lemma 1.

7.2.2 Case where c is uniformly bounded

When $\|c\|_\infty$ is finite, a sharper bound holds for the expectation terms, viz.

$$E_h[|\phi_{jk_1}^{(a_1)}(F(X))|^{b_1}|\phi_{jk_2}^{(a_2)}(G(Y))|^{b_2}] \leq \||\phi_{jk_1}^{(a_1)}(u)|^{b_1}|\phi_{jk_2}^{(a_2)}(v)|^{b_2}\|_1\|c\|_\infty$$
$$\leq 2^{j(b_1/2+b_2/2-2)}\|c\|_\infty \, \|\phi^{(a_1)}\|_{2b_1}^{b_1} \, \|\phi^{(a_2)}\|_{2b_2}^{b_2}.$$

The proof is the same as before, except that (for the same choice of ϵ) one gets

$$|E_h(T_1)| + |E_h(TT_2)| + |E_h(TT_3)| \leq K \left(2^{4j}\epsilon^4 + 2^{3j}\epsilon^3 + 2^{2j}\epsilon^2\right)$$

when $i = \ell$, and

$$|E_h(T_1)| + |E_h(TT_2)| + |E_h(TT_3)| \leq K \left(2^{3j}\epsilon^4 + 2^j\epsilon^3 + \epsilon^2\right)$$

when $i \neq \ell$. In this case, the constant $K > 0$ also depends on $\|c\|_\infty$. This completes the proof of Lemma 1.

Acknowledgements

References

Autin, F., Le Pennec, E., and Tribouley, K. (2008). Thresholding methods to estimate the copula density. *Submitted available online at* http://www.cmi.univ-mrs.fr/~autin/DONNEES/COPULAS.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t*-distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65:367–389.

Biau, G. and Wegkamp, M. H. (2005). A note on minimum distance estimation of copula densities. *Statist. Probab. Lett.*, 73:105–114.

Butucea, C. and Tribouley, K. (2006). Nonparametric homogeneity tests. *J. Statist. Plann. Inference*, 136:597–639.

Chen, X. and Fan, Y. (2005). Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *Canad. J. Statist.*, 33:389–414.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance.* Wiley, New York.

Daubechies, I. (1992). *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés: Un test non paramétrique d'indépendance. *Acad. Royal Bel., Bull. Class. Sci., 5$^{\mathrm{e}}$ série*, 65:274–292.

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

Fermanian, J.-D. and Scaillet, O. (2005). Some statistical pitfalls in copula modelling for financial applications. In Klein, E., editor, *Capital Formation, Governance and Banking*, pages 59–74. Nova Science Publishing, New York.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *N. Am. Actuar. J.*, 2:1–25.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1998). Comment on "Understanding relationships using copulas," by Edward w. Frees and Emiliano a. Valdez, January 1998. *N. Am. Actuar. J.*, 2:143–149.

Genest, C., Quessy, J.-F., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the integral probability transformation. *Scand. J. Statist.*, 33:337–366.

Genest, C., Rémillard, B., and Beaudoin, D. (2008). Goodness-of-fit tests for copulas: A review and a power study. *Insurance Math. Econom.*, page in press.

Gijbels, I. and Mielniczuk, J. (1990). Estimating the density of a copula function. *Comm. Statist. Theory Methods*, 19:445–464.

Grégoire, V., Genest, C., and Gendron, M. (2008). Using copulas to model price dependence in energy markets. *Energy Risk*, page in press.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman & Hall, London.

Kerkyacharian, G. and Picard, D. (1992). Density estimation in Besov spaces. *Statist. Probab. Lett.*, 13:15–24.

Klugman, S. and Parsa, R. (1999). Fitting bivariate loss distributions with copulas. *Insurance Math. Econom.*, 24:139–148.

Mallat, S. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11:674–693.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management.*

Princeton University Press, Princeton, NJ.

Meyer, Y. (1992). *Wavelets and operators*, volume 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. Translated from the 1990 French original by D. H. Salinger.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York, second edition.

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

Triebel, H. (1992). *Theory of Function Spaces. II*, volume 84 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel.