Journal of Multivariate Analysis 101 (2010) 200–222

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Thresholding methods to estimate copula density

F. Autin [a], E. Le Pennec [b], K. Tribouley [c,*]

[a] CMI, Université Aix-Marseille 1, 39 rue F. Joliot Curie, 13453 Marseille Cedex 13, France

[b] LPMA, Université Paris 7, 175 rue du Chevaleret, 75013 Paris, France

[c] LPMA et MODALX, Université Paris 10, 200 avenue de la République, 92001 Nanterre Cedex, France

ARTICLE INFO

Article history:
Received 31 January 2008
Available online 22 August 2009

AMS subject classification:
62G10
62G20
62G30

Keywords:
Copula density
Wavelet method
Thresholding rules
Minimax theory
Maxiset theory

ABSTRACT

This paper deals with the problem of multivariate copula density estimation. Using wavelet methods we provide two shrinkage procedures based on thresholding rules for which knowledge of the regularity of the copula density to be estimated is not necessary. These methods, said to be adaptive, have proved to be very effective when adopting the minimax and the maxiset approaches. Moreover we show that these procedures can be discriminated in the maxiset sense. We provide an estimation algorithm and evaluate its properties using simulation. Finally, we propose a real life application for financial data.

1. Introduction

In risk management, in the areas of finance, insurance and climatology, for example, a new tool has been developed to model the dependence structure of data: the copula. **A copula** is a multivariate joint distribution defined on the d-dimensional unit cube $[0, 1]^d$ such that every marginal distribution is uniform on the interval $[0, 1]$. Sklar's Theorem [1] allows us to separately study the laws of the coordinates X^m for $m = 1, \dots d$, of any d-vector X, and the dependence between the coordinates.

Theorem 1. *Let $d \geq 2$ and H be a d-variate distribution function. If each marginal distribution F_m, $m = 1, \dots d$, of H is continuous, a unique d-variate copula C exists, so that*

$$\forall (x_1, \dots, x_d) \in \mathbb{R}^d, \quad H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

The **copula model** has been extensively studied within a parametric framework. Numerous classes of parametric copulas, parametric distribution functions C, have been proposed. For instance there is the elliptic family, which contains the Gaussian copulas and the Student copulas, and the Archimedean family, which contains the Gumbel copulas, the Clayton copulas and the Frank copulas. The first step of such a parametric approach is to select the parametric family of the copula being considered. This is a **modeling** task that may require finding new copula and methodologies to simulate the

* Corresponding author.
E-mail addresses: autin@cmi.univ-mrs.fr (F. Autin), lepennec@math.jussieu.fr (E. Le Pennec), ktriboul@u-paris10.fr, karine.tribouley@u-paris10.fr (K. Tribouley).

0047-259X/$ – see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.jmva.2009.07.009

corresponding data. Usual **statistical inference** (estimation of the parameters, goodness-of-fit test, etc) can only take place in a second step. Both tasks have been extensively studied.

We propose here to study the copula model within a non-parametric framework. Our aim is to make very mild assumption about the copula. Thus, contrary to the parametric setting, no a priori model of the phenomenon is needed. For practitioners, non-parametric estimators could be seen as a **benchmark** that makes it possible to select the right parametric family by comparing them to an **agnostic** estimate. In fact, most of the time, practitioners observe the scatter plot of $\{(X_i, Y_i), i = 1, \dots, n\}$, or $\{(R_i, S_i), i = 1, \dots, n\}$ where R and S are the rank statistics of (X, Y), and then attempt, on the basis of these observations, only to guess the family of parametric copulas the target copula belongs to. Providing good non-parametric estimators of the copula makes this task easier and provides a more rigorous way to describe the copula.

In our study, we propose non-parametric procedures to estimate the **copula density** c associated with the copula C. More precisely, we consider the following model. We assume that we are observing an n-sample $(X_1^1, \dots, X_1^d), \dots, (X_n^1, \dots, X_n^d)$ of independent data with the same distribution H (and the same density h) as $(X^1, \dots, X^d)$. Referring to the marginal distributions of the coordinates of the vector $(X^1, \dots, X^d)$ as $F_1, \dots, F_d$, we are interested in estimating the copula density c defined as the derivative (if it exists) of the copula distribution

$$c(u_1, \dots, u_d) = \frac{h(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdots f_d(F_d^{-1}(u_d))}$$

where $F_p^{-1}(u_p) = \inf\{x \in R : F_p(x) \geq u_p\}$, $1 \leq p \leq d$ and $u = (u_1, \dots, u_d) \in [0, 1]^d$. This would be a classical density model if the marginal distributions, and thus the direct observations, $(U_i^1 = F_1(X_i^1), \dots, U_i^d = F_d(X_i^d))$ for $i = 1, \dots, n$, were known. Unfortunately, this is not the case. We can observe that this model is somewhat similar to the non-parametric regression model with unknown random design studied in Kerkyacharian and Picard [2] with their **warped wavelet families**.

Two wavelet-based methods are presented: a **Local Thresholding Method** and a **Global Thresholding Method**. Both are extensions of the methods studied by Donoho et al. [3,4] and Kerkyacharian et al. [5] in the classical density estimation framework. The copula density c is estimated using a specific multiscale basis representation of $[0, 1]^d$, the wavelet representation. Each wavelet coefficient is estimated individually and possibly *thresholded* (set to 0) if it is considered to be non-significant. The two methods differ in their definition of non-significant: one is local, and individually considers considering individually each estimated coefficient; the other is global, and simultaneously considers all coefficients at each scale. Contrary to the kernel-based method, these methods do not require a fine-tuning of the smoothing parameters. The definition of non-significant is not dependent on the (unknown) regularity of the copula: the procedures are data driven and automatically provide an estimator close to the best possible estimators. We can observe that this includes the estimators that require precise knowledge of the regularity of the copula.

We first measure the performance for both estimators on all copula densities that are bounded and that belong to a very large class of regularity. The good behavior of our procedures is due to the approximation properties of the wavelet basis. A regular copula can be approximated by few non-zero-wavelet coefficients leading to estimators with both a small bias and small variance. The wavelet representation is connected to well-known regularity spaces: Besov spaces, in particular, that contain Sobolev spaces or Holder spaces, can be defined through the wavelet coefficients. The first results of this paper are the proofs that the rate of convergence of our estimators are:

(1) optimal in the minimax sense (up to a logarithmic factor),
(2) the same as in the standard density model. Using pseudo-data instead of direct observations does not damage the quality of the procedures.

It should be observed that the same behavior also arises for linear wavelet procedures (see Genest et al. [6]). However, the linear procedure is not adaptive in the sense that we need to know the regularity index of the copula density to obtain optimal procedures. This paper provides a solution to this drawback.

Following the maxiset approach, we then characterize the precise set of copula densities estimated at a given polynomial rate for our procedures. We verify that the local one outperforms the others, in the sense that this is the procedure for which the set of copula densities estimated at a given rate is the largest.

One of the main difficulties of copula density estimation lies in the fact that most of the pertinent information is located near the boundaries of $[0, 1]^d$ (at least for the most common copulas like the Gumbel copula or the Clayton copula). In the theoretical construction, we use a family of wavelets especially designed for this case: they extend only within the compact set $[0, 1]^d$, do not thus cross the boundary and are optimal in terms of the approximation. In the practical construction, boundaries remain an issue. In fact, the theoretically optimal wavelets are rarely implemented and when they are, they are not as efficient as in the theory. We propose an appropriate symmetrization/periodization process of the original data here in order to deal with this problem. We also enhance the scheme by adding some translation invariance. We numerically verify the good behavior of the proposed scheme for simulated data with the usual parametric copula families. We then illustrate an application on financial data by proposing a method to choose the parametric family and the parameters based on a preliminary non-parametric estimator used as a benchmark. The last result of this paper is thus to propose an implementation that is very easy to use and that provides good estimators.

The paper is organized as follows. Section 2 describes the multidimensional wavelet basis used in the sequel. Section 3 is devoted to the description of thresholding estimation procedures for which performances are studied in Section 4 for the minimax approach and in Section 5 for the maxiset approach. Section 6 deals with the practical results. Proofs of main theorems are given in Section 7, while proofs of propositions and technical lemmas are included in the Appendix.

2. Wavelet setting

Our multivariate wavelet basis is built thanks to the tensorial product of the wavelet basis on the interval proposed by Cohen et al. [7]. More precisely, for any $j_0 \in \mathbb{N}$, we consider

$$\{\phi_{j_0,k}\}_{k\in\{1,\dots,2^{j_0}\}} \bigcup \{\psi_{j,k}\}_{j\geq j_0, k\in\{1,\dots,2^j\}}$$

the basis of $L_2([0,1])$ obtained by Cohen et al. [7] from a compactly supported function ϕ and its corresponding wavelet ψ. Here $h_{j,k}(\cdot)$ denotes the function $2^{j/2}h(2^j\cdot -k)$ for $h(\cdot)$ being either $\phi(\cdot)$ or $\psi(\cdot)$. We define then the multivariate wavelets as

$$\phi_{j,k}(x_1,\dots,x_d) = \phi_{j,k_1}(x_1)\dots\phi_{j,k_d}(x_d),$$

$$\psi^{\epsilon}_{j,k}(x_1,\dots,x_d) = \prod_{m=1}^{d} \phi^{1-\epsilon_m}_{j,k_m}(x_m)\psi^{\epsilon_m}_{j,k_m}(x_m),$$

for all $\epsilon = (\epsilon_1,\dots,\epsilon_d) \in S_d = \{0,1\}^d \setminus \{(0,\dots,0)\}$. Indeed, with $k = (k_1,\dots,k_d)$ a multicomponent vector, the set

$$\{\phi_{j_0,k}, \psi^{\epsilon}_{j,\ell} | j \geq j_0, k \in \{1,\dots,2^{j_0}\}^d, \ell \in \{1,\dots,2^j\}^d, \epsilon \in S_d\}$$

is an orthonormal basis of $L_2([0,1]^d)$ for any $j_0 \in \mathbb{N}$ (see for example Meyer [8]). It follows that any real function h of $L_2([0,1]^d)$ can be expanded as

$$\forall x \in [0,1]^d, \quad h(x) = \sum_{k\in\{1,\dots,2^{j_0}\}^d} h_{j_0,k}\phi_{j_0,k}(x) + \sum_{j\geq j_0}\sum_{k\in\{1,\dots,2^j\}^d}\sum_{\epsilon\in S_d} h^{\epsilon}_{j,k}\psi^{\epsilon}_{j,k}(x),$$

where the scaling coefficient $h_{j_0,k}$ and the wavelet coefficient $h^{\epsilon}_{j,k}$ are given by

$$h_{j_0,k} = \int_{[0,1]^d} h(x)\phi_{j_0,k}(x)\mathrm{d}x \quad \text{and} \quad h^{\epsilon}_{j,k} = \int_{[0,1]^d} h(x)\psi^{\epsilon}_{j,k}(x)\mathrm{d}x.$$

Roughly speaking, the expansion of the analyzed function on the wavelet basis splits into the "trend" at the level j_0 and the sum of the "details" for all the larger levels $j, j \geq j_0$. For more details on the multivariate setting in the density model, see Tribouley [9].

To simplify the notation, we omit the range of k and ϵ in the summation from now on. However, note that for any level j, the summation extends over a finite number of terms $2^{jd} \times (2^d - 1)$.

3. Estimation procedures

For a copula density c belonging to $L_2([0,1]^d)$, it is equivalent to estimate c and to estimate its wavelet coefficients. It turns out that this can be easily done. Observe that, for any d-variate function Φ

$$E_c(\Phi(U_1,\dots,U_d)) = E_h\left(\Phi(F_1(X^1),\dots,F_d(X^d))\right)$$

or equivalently

$$\int_{[0,1]^d} \Phi(u)c(u)\mathrm{d}u = \int_{\mathbb{R}^d} \Phi(F_1(x_1),\dots,F_d(x_d))h(x_1,\dots,x_d)\mathrm{d}x_1\dots\mathrm{d}x_d.$$

This means that the wavelet coefficients of the copula density c on the wavelet basis are equal to the coefficients of the joint density h on the warped wavelet family

$$\{\phi_{j_0,k}(F_1(\cdot),\dots,F_d(\cdot)), \psi^{\epsilon}_{j,\ell}(F_1(\cdot),\dots,F_d(\cdot)) | j \geq j_0, k \in \{0,\dots,2^{j_0}\}^d, \ell \in \{0,\dots,2^j\}^d, \epsilon \in S_d\}.$$

The corresponding empirical coefficients are

$$\widehat{c_{j_0,k}} = \frac{1}{n}\sum_{i=1}^{n} \phi_{j_0,k}(F_1(X_i^1),\dots,F_d(X_i^d))$$

and

$$\widehat{c^{\epsilon}_{j,k}} = \frac{1}{n}\sum_{i=1}^{n} \psi^{\epsilon}_{j,k}(F_1(X_i^1),\dots,F_d(X_i^d)). \tag{1}$$

These coefficients cannot be evaluated since the distributions functions associated to the marginal distributions $F_1, \ldots, F_d$ are unknown. We propose to replace these unknown distributions functions by their corresponding empirical distributions functions $\widehat{F_1}, \ldots \widehat{F_d}$. The modified empirical coefficients are

$$\widetilde{c_{j_0,k}} = \frac{1}{n}\sum_{i=1}^{n}\phi_{j_0,k}(\widehat{F_1}(X_i^1), \ldots, \widehat{F_d}(X_i^d))$$

and

$$\widetilde{c_{j,k}^{\epsilon}} = \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}^{\epsilon}(\widehat{F_1}(X_i^1), \ldots, \widehat{F_d}(X_i^d))$$

where the empirical distribution functions are given by

$$\forall t \in R, \quad \widehat{F_p}(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_i^p \leq t\}, \quad p = 1, \ldots, d.$$

The most unaffected way to estimate the density c is to reconstruct the function from its modified empirical coefficients. We consider here the very general family of **truncated estimators** of c defined by

$$\widetilde{c_T} := \widetilde{c_T}(j_n, J_n) = \sum_k \widetilde{c_{j_n,k}}\phi_{j_n,k} + \sum_{j=j_n}^{J_n}\sum_{k,\epsilon}\omega_{j,k}^{\epsilon}\widetilde{c_{j,k}^{\epsilon}}\psi_{j,k}^{\epsilon}, \tag{2}$$

where the indices (j_n, J_n) are such that $j_n \leq J_n$ and where, for any (j, k, ϵ), $\omega_{j,k}^{\epsilon}$ belongs to $\{0, 1\}$. Notice that the weight $\omega_{j,k}^{\epsilon}$ may or may not depend on the observations.

The later case has been considered by Genest et al. [6] who proposed to use a **linear procedure**

$$\widetilde{c_L} := \widetilde{c_L}(j_n) = \sum_k \widetilde{c_{j_n,k}}\phi_{j_n,k} \tag{3}$$

for a suitable choice of j_n. The accuracy of this linear procedure relies on the fast uniform decay of the wavelets coefficients across the scale as soon as the function is uniformly regular. The trend at the chosen level j_n becomes a sufficient approximation. The optimal choice of j_n depends on the regularity of the unknown function to be estimated and thus the procedure is not data driven.

We propose here to use some **nonlinear procedures** based on hard thresholding methods (see for instance Cohen et al. [10], Kerkyacharian and Picard [11], and Donoho and Johnstone [3])) that overcome this issue. In hard thresholding procedures, the "small" coefficients are killed by setting the corresponding weight $\omega_{j,k}^{\epsilon}$ to 0. They differ by the definition of "small". We study here two strategies: a local one, where each coefficient is considered individually, and a global one, where all the coefficients at the same scale are considered globally.

For a given threshold level $\lambda_n > 0$ and a set of indices (j_n, J_n), the local hard threshold weights $\omega_{j,k}^{\epsilon,L}$ and the global hard threshold weights $\omega_{j,k}^{\epsilon,G}$ are defined respectively by

$$\omega_{j,k}^{\epsilon,HL} = \mathbf{1}\{|\widetilde{c_{j,k}^{\epsilon}}| > \lambda_n\}. \quad \text{and} \quad \omega_{j,k}^{\epsilon,HG} = \mathbf{1}\left\{\sum_k |\widetilde{c_{j,k}^{\epsilon}}|^2 > 2^{jd}\lambda_n^2\right\}.$$

Let us put $\widetilde{c_{j,k}^{\epsilon,HL}} = \omega_{j,k}^{\epsilon,HL}\widetilde{c_{j,k}^{\epsilon}}$ and $\widetilde{c_{j,k}^{\epsilon,HG}} = \omega_{j,k}^{\epsilon,HG}\widetilde{c_{j,k}^{\epsilon}}$. The corresponding local hard thresholding estimators $\widetilde{c_{HL}}$ and global hard thresholding estimators $\widetilde{c_{HG}}$ are defined respectively by

$$\widetilde{c_{HL}} := \widetilde{c_{HL}}(j_n, J_n, \lambda_n) = \sum_k \widetilde{c_{j_n,k}}\phi_{j_n,k} + \sum_{j=j_n}^{J_n}\sum_{k,\epsilon}\widetilde{c_{j,k}^{\epsilon,HL}}\psi_{j,k}^{\epsilon}. \tag{4}$$

and

$$\widetilde{c_{HG}} := \widetilde{c_{HG}}(j_n, J_n, \lambda_n) = \sum_k \widetilde{c_{j_n,k}}\phi_{j_n,k} + \sum_{j=j_n}^{J_n}\sum_{k,\epsilon}\widetilde{c_{j,k}^{\epsilon,HG}}\psi_{j,k}^{\epsilon}. \tag{5}$$

The nonlinear procedures given in (4) and (5) depend on the level indices (j_n, J_n) and on the threshold value λ_n. In the next section, we define a criterion to measure the performance of our procedures and explain how to choose those parameters to achieve optimal performance.

4. Minimax results

4.1. Minimax approach

The **minimax theory** is a classical way to analyze the performance of estimation procedures which has been extensively developed since the 1980's. In the minimax setting, the practitioner chooses a loss function $\ell(.)$ that quantifies the loss of a misestimation and a functional class $\mathcal{F}$ which is supposed to contain the estimated function c. He measures then the worst case loss of the estimator $\widetilde{c}$:

$$\sup_{c\in\mathcal{F}} E\ell(\widetilde{c}-c)$$

and compares it with the best possible value of this quantity, called the **minimax risk**,

$$R(\mathcal{F}) = \inf_{\widetilde{c}} \sup_{c\in\mathcal{F}} E\,\ell(\widetilde{c}-c).$$

The infimum is taken over all possible estimators. If both coincide, the procedure is **minimax optimal** on the class $\mathcal{F}$. A lot of minimax results for standard statistical models and many families of functional spaces as Sobolev spaces, Holder spaces, and others as the family of Besov spaces have been now established (see for instance Ibragimov and Khasminski [12] or Kerkyacharian and Picard [10]).

4.2. Besov bodies

We deal here with wavelet methods; it is thus standard to consider as functional classes the Besov bodies characterized by the wavelet coefficients as follows

Definition 1 (*Strong Besov Bodies*). For any $s > 0$, a function c belongs to the Besov body $\mathcal{B}^s_{2\infty}$ if and only if its sequence of wavelet coefficients $c^{\epsilon}_{j,k}$ satisfies

$$\sup_{J\geq 0} 2^{2Js} \sum_{j>J} \sum_{k,\epsilon} (c^{\epsilon}_{j,k})^2 < \infty.$$

These spaces can be seen as extensions of classical regularity spaces. For example, any function that is s times differentiable belongs to $\mathcal{B}^s_{2\infty}$ (see for instance Donoho and Johnstone [3]). In this paper, we focus on the quadratic loss and these Besov bodies for which the minimax risks are known:

$$\forall c \in \mathcal{B}^s_{2\infty}, \quad \sup_n \inf_{\widetilde{c}} n^{\frac{2s}{2s+d}}\, E\|\widetilde{c}-c\|_2^2 < \infty$$

where the infimum is taken other any estimator of the density c. Notice that this defines a **minimax rate** that measures the best possible decay of the error when the number of samples n varies.

4.3. Optimality

If the wavelet is regular enough, Genest et al. [6] prove that the linear procedure $\widetilde{c_L} = \widetilde{c_L}(j_n^*)$ defined in (3) is minimax optimal on the Besov body $B^s_{2\infty}$ for all $s > 0$ provided j_n^* is chosen so that:

$$2^{j_n^*-1} < n^{\frac{1}{2s+d}} \leq 2^{j_n^*}.$$

As hinted in the previous section, this result is not fully satisfactory because the optimal procedure depends on the regularity s of the density which is generally unknown.

The thresholding procedures described in (4) and (5) do not suffer from this drawback: the same choice of parameters j_n, J_N and λ_n yields an almost minimax optimal estimator simultaneously for any $B^s_{2,\infty}$. The following theorem (which is a direct consequence of Theorem 3 established in the following section) ensures indeed that

Theorem 2. *Assume that the wavelet is continuously differentiable and let $s > 0$. For any choice of level j_n and J_n and threshold λ_n such that*

$$2^{j_n-1} < (\log(n))^{1/d} \leq 2^{j_n}, \qquad 2^{J_n-1} < \left(\frac{n}{\log n}\right)^{1/d} \leq 2^{J_n}, \qquad \lambda_n = \sqrt{\frac{\kappa \log(n)}{n}}$$

for some κ large enough,

$$\forall s > 0, \quad c \in \mathcal{B}^s_{2\infty} \cap L_\infty([0,1]^d) \Rightarrow \sup_n \left(\frac{n}{\log(n)}\right)^{\frac{2s}{2s+d}} E\|\widetilde{c}-c\|_2^2 < \infty$$

where $\widetilde{c}$ stands either for the hard local thresholding procedure $\widetilde{c_{HL}}(j_n, J_n, \lambda_n)$ or for the hard global thresholding procedure $\widetilde{c_{HG}}(j_n, J_n, \lambda_n)$.

Observe that, when $s > d/2$, the embedding $\mathcal{B}^s_{2\infty} \subsetneq L_\infty([0,1]^d)$ is satisfied. Thus the assumption $c \in \mathcal{B}^s_{2\infty} \cap L_\infty([0,1]^d)$ in Theorem 2 could be replaced with the assumption $c \in \mathcal{B}^s_{2\infty}$.

We immediately deduce

Corollary 4.1. *The hard local thresholding procedure $\widetilde{c_{HL}}$ and the hard global thresholding procedure $\widetilde{c_{HG}}$ are adaptive minimax optimal up to a logarithmic factor on the Besov bodies $\mathcal{B}^s_{2\infty}$ for the quadratic loss function.*

Notice that this logarithmic factor is nothing but the classical "price" of adaptivity.

4.4. Criticism on the minimax point of view

The minimax theory requires the choice of the functional space $\mathcal{F}$ (or the choice of a sequence of functional spaces $\mathcal{F}_s$). The arbitrariness of this choice is the main drawback of the minimax approach. Indeed, Corollary 4.1 establishes that no other procedures could be uniformly better on the spaces $\mathcal{B}^s_{2\infty}$ but it does not address two important questions. What about a different choice of spaces? Both of our thresholding estimators achieve the minimax rate on the spaces $\mathcal{B}^s_{2,\infty}$ but is there a way to distinguish their performance? To answer to these questions, we propose to explore the **maxiset approach**.

5. Maxiset results

5.1. Maxiset approach

The maxiset point of view developed by Cohen et al. [13] is inspired by the approximation theory. This new way to analyze the performance of estimation procedures fixes the procedures instead of the space. The space of functions (called the maxiset) for which a given procedure attains a prescribed rate of convergence is studied. The larger the space the better the estimator. The maxiset point of view is more optimistic than the minimax point of view in the sense that the maxiset approach points out **all** the functions estimated by a fixed procedure at a **given rate** instead of looking at a **worst case** behavior on a **given class**.

The maxiset of a fixed estimation procedure $\widetilde{c}$ associated with the rate of convergence r_n, denoted $\mathcal{MS}(\widetilde{c}, r_n)$, is defined through the following equivalence

$$\sup_n r_n^{-1} E\|\widetilde{c} - c\|_2^2 < \infty \iff c \in \mathcal{MS}(\widetilde{c}, r_n).$$

where we still consider the quadratic loss. Remark that if an estimator $\widetilde{c}$ of c achieves the (minimax) rate r_n on a functional space $\mathcal{F}$ then $\mathcal{F}$ is included in the maxiset $\mathcal{MS}(\widetilde{c}, r_n)$. Minimax procedures on the same target space can thus differ by their maxisets, providing a way to compare them: the best procedure is the procedure admitting the largest maxiset.

Many papers have considered the maxiset approach in the white noise model (see Cohen et al. [13] or Autin et al. [14]) and the density estimation model (see Autin [15]). In both models, the hard local procedure appears to be the best one amongst a large family of shrinkage procedures, called the **elitist rules**, and the corresponding maxisets involve **weak Besov spaces**.

5.2. Weak Besov spaces

These spaces are special cases of Lorentz spaces defined by properties of the wavelet coefficients. We define here the local weak Besov spaces $\mathcal{W}_L(r)$ and the global weak Besov spaces $\mathcal{W}_G(r)$ by

Definition 2 (*Local Weak Besov Spaces*). For any $0 < r < 2$, a function $c \in L_2([0,1]^d)$ belongs to the local weak Besov space $\mathcal{W}_L(r)$ if and only if its sequence of wavelet coefficients $c^\epsilon_{j,k}$ satisfies the following equivalent properties:

- $\sup_{0<\lambda\le 1} \lambda^{r-2} \sum_{j\ge 0} \sum_{k,\epsilon} (c^\epsilon_{j,k})^2 \mathbf{1}\{|c^\epsilon_{j,k}| \le \lambda\} < \infty$,
- $\sup_{0<\lambda\le 1} \lambda^{r} \sum_{j\ge 0} \sum_{k,\epsilon} \mathbf{1}\{|c^\epsilon_{j,k}| > \lambda\} < \infty$.

and

Definition 3 (*Global Weak Besov Spaces*). For any $0 < r < 2$, a function $c \in L_2([0,1]^d)$ belongs to the global weak Besov space $\mathcal{W}_G(r)$ if and only if its sequence of wavelet coefficients $c^\epsilon_{j,k}$ satisfies the following equivalent properties:

- $\sup_{0<\lambda\le 1} \lambda^{r-2} \sum_{j\ge 0} \sum_{k,\epsilon} (c^\epsilon_{j,k})^2 \mathbf{1}\{\sum_k (c^\epsilon_{j,k})^2 \le 2^{dj}\lambda^2\} < \infty$,
- $\sup_{0<\lambda\le 1} \lambda^{r} \sum_{j\ge 0} 2^{dj} \sum_{\epsilon} \mathbf{1}\{\sum_k (c^\epsilon_{j,k})^2 > 2^{dj}\lambda^2\} < \infty$.

As for the definition of the Besov bodies, the definition depends on the wavelet basis. However, as established by Meyer [8] and Cohen et al. [13], this dependency is quite weak. Note that the equivalences between the properties used in the definitions of the weak Besov spaces can be proved as in Cohen et al. [13].

These spaces are clearly related to the Besov bodies $\mathcal{B}^s_{2,\infty}$. Indeed some computation proves that $\mathcal{B}^s_{2,\infty} \subset \mathcal{W}_G\left(\frac{2d}{2s+d}\right)$ and $\mathcal{B}^s_{2,\infty} \subset \mathcal{W}_L\left(\frac{2d}{2s+d}\right)$. In Section 7.3, we prove the following strict inclusion property

Proposition 1. *For any $0 < r < 2$, $\mathcal{W}_G(r) \subsetneq \mathcal{W}_L(r)$.*

5.3. Performances and comparison of our procedures

In this section, we study the maxiset of the linear procedure and the maxisets of the thresholding procedures described in Section 1. We focus on the near minimax optimal procedures: we use the following choices of parameters

$$2^{j_n-1} < (\log(n))^{1/d} \leq 2^{j_n}, \qquad 2^{J_n-1} < \left(\frac{n}{\log(n)}\right)^{1/d} \leq 2^{J_n}$$
$$2^{j_n^*-1} < \left(\frac{n}{\log(n)}\right)^{\frac{1}{2s+d}} \leq 2^{j_n^*}, \qquad \lambda_n = \sqrt{\frac{\kappa \log(n)}{n}}$$

for some $\kappa > 0$ and we study the linear estimator $\widetilde{c_L} = \widetilde{c_L}(j_n^*)$, the local thresholding estimator $\widetilde{c_{HL}} = \widetilde{c_{HL}}(j_n, J_n, \lambda_n)$ and the global thresholding estimator $\widetilde{c_{HG}} = \widetilde{c_{HG}}(j_n, J_n, \lambda_n)$.

Let us fix $s > 0$. We focus on the rate $r_n = \left(n^{-1}\log(n)\right)^{2s/(2s+d)}$ which is the (near) minimax rate achieved on the space $B^s_{2\infty}$. The following theorem exhibits the maxisets of the procedures with this target rate r_n.

Theorem 3. *Let $s > 0$, and assume that $c \in L_\infty([0,1]^d)$. For a large enough κ, we get*

$$\sup_n \left(\frac{n}{\log(n)}\right)^{\frac{2s}{2s+d}} E\|\widetilde{c_L} - c\|_2^2 < \infty \iff c \in \mathcal{B}^s_{2\infty}, \tag{6}$$

$$\sup_n \left(\frac{n}{\log(n)}\right)^{\frac{2s}{2s+d}} E\|\widetilde{c_{HL}} - c\|_2^2 < \infty \iff c \in \mathcal{B}^{\frac{ds}{2s+d}}_{2\infty} \cap \mathcal{W}_L\left(\frac{2d}{2s+d}\right), \tag{7}$$

$$\sup_n \left(\frac{n}{\log(n)}\right)^{\frac{2s}{2s+d}} E\|\widetilde{c_{HG}} - c\|_2^2 < \infty \iff c \in \mathcal{B}^{\frac{ds}{2s+d}}_{2\infty} \cap \mathcal{W}_G\left(\frac{2d}{2s+d}\right). \tag{8}$$

Note that the same spaces arise if we assume that the marginal distributions are known (see Autin et al. [16]). This is also a nice result to prove that the lack of direct observations does not make the problem harder.

The following strict embedding,

$$\mathcal{B}^s_{2\infty} \subsetneq \mathcal{B}^{\frac{ds}{2s+d}}_{2\infty} \cap \mathcal{W}_G\left(\frac{2d}{2s+d}\right)$$

implies

Corollary 5.1. *Let $s > 0$ and let us consider the target rate*

$$r_n = \left(\frac{\log(n)}{n}\right)^{\frac{2s}{2s+d}}. \tag{9}$$

Then we get

$$\mathcal{MS}(\widetilde{c_L}, r_n) \subsetneq \mathcal{MS}(\widetilde{c_{HG}}, r_n) \subsetneq \mathcal{MS}(\widetilde{c_{HL}}, r_n).$$

In other words, in the maxiset point of view and when the quadratic loss is considered, the thresholding rules outperform the linear procedure. Moreover, the hard local thresholding estimator $\widetilde{c_{HL}}$ appears to be the best estimator among the considered procedures since it strictly outperforms the hard global thresholding estimator $\widetilde{c_{HG}}$.

6. Applied results

In this section, we deal with numerical aspects of the thresholding estimation. Although we have used wavelets on the interval in the theory, they are seldom available in numerical packages. We propose here ways to overcome this drawback. We test then our methodology on simulated datasets and we verify that there is a best numerical scheme. We test it in the context of the parametric estimation. Finally, we apply the chosen procedure to financial data.

6.1. Algorithms

For the sake of simplicity, the estimation algorithms are described in the bivariate case but their extension to other dimension is straightforward. We assume that a sample $\{(X_i, Y_i)\}_{1\leq i\leq n}$ of size n is given.

All estimators proposed in this paper can be summarized in an algorithm having seven steps:

(1) Rank the X_i, Y_i with

$$R_i = \sum_{l=1}^{n} \mathbf{1}\{X_l \leq X_i\} \quad \text{and} \quad S_i = \sum_{l=1}^{n} \mathbf{1}\{Y_l \leq Y_i\}.$$

(2) Compute the maximal scale index $J_n = \lfloor \frac{1}{2} \log_2(\frac{n}{\log n}) \rfloor$.

(3) Compute the empirical scaling coefficients at the maximal scale index J_n:

$$\widetilde{c_{J_n,k_1,k_2}} = \frac{1}{n} \sum_{i=1}^{n} \phi_{J_n,k_1,k_2}\left(\frac{R_i}{n}, \frac{S_i}{n}\right) \quad \text{for } 1 \leq k_1 \leq 2^{J_n} \text{ and } 1 \leq k_2 \leq 2^{J_n}.$$

(4) Compute the empirical wavelet coefficients $\widetilde{c^{\epsilon}_{j,k_1,k_2}}$ from these scaling coefficients with the fast 2D wavelet transform algorithm.

(5) Threshold these coefficients according to the global thresholding rule or the local thresholding rule to obtain the estimated wavelet coefficients $\widetilde{c^{\epsilon,T}_{j,k_1,k_2}}$.

(6) Compute the estimated scaling coefficients $\widetilde{c^{T}_{J_n,k_1,k_2}}$ at scale index J_n by the fast 2D wavelet inverse transform algorithm.

(7) Construct the estimated copula density $\widetilde{c}$ using the formula

$$\widetilde{c} = \sum_{k_1,k_2} \widetilde{c_{J_n,k_1,k_2}} \phi_{J_n,k_1,k_2}.$$

Unfortunately only the steps (1), (2) and (5) are as straightforward as they seem to be. Two issues make the other steps more complex: the handling of the boundaries and the discrete nature of computer results.

The later issue is the easiest to solve. As in most numerical scheme, we fix a grid resolution of $1/N$ much smaller than 2^{-J_n} and approximate the estimated copula density at step (7) on the induced grid $(i/N, j/N)$. Although the scaling functions are not always known explicitly, a very good approximation can be computed on this grid and we assume from now on that this effect is negligible. The norms E_q appearing in the numerical results (see Tables A.1–A.4) are thus empirical norms $\| \cdot \|_{N,q}$ on this grid. In our experiments, we take $N = 4 \times 2^{J_n}$. Notice that step (3) also requires an evaluation of the scaling function using a similar approximation.

The former issue, the boundary handling, is the most important one. Indeed, for most copula densities, the interesting behavior arises in the corners which are the most difficult parts to handle numerically. In our theorems, we use the wavelet on the interval defined by Cohen et al. [7]. We test this scheme numerically and we compare it with other choices of boundary handling.

The classical construction of the wavelet yields a basis over $\mathbb{R}^d$ while we only have samples on $[0, 1]^d$.

- A first choice is to consider the function of $[0, 1]^d$ to be estimated as a function of $\mathbb{R}^d$ which is 0 outside $[0, 1]^d$. This choice is called **zero padding**.
- A second choice is to suppose that we observe the restriction on $[0, 1]^d$ of a 1-periodic function, this is equivalent to work in the classical periodic wavelet setting. This choice called **periodization** is very efficient when the function is really periodic.
- We propose also to modify the periodization and assume that we observe the restriction over $[0, 1]^d$ of a even 2-periodic function. As this introduces a symmetrization over the existing borders, we call this method **symmetrization**. It avoids the introduction of discontinuities along the border. Notice that nevertheless this symmetrization introduces discontinuities for the derivatives at the boundaries.
- The last choice is the use of the tailored wavelet on the interval proposed by Cohen et al. [7] and the corresponding **boundary** corrected wavelet transform. Remark that this transform is more involved than the classical one.

Once this choice is made, we use the corresponding fast wavelet transform. The resulting estimated copula density is the restriction to $[0, 1]^d$ of the estimated function.

Wavelet thresholding methods in a basis suffer from a griding effect. Often, isolated wavelets are seen in the estimated signal. To reduce this effect, we propose to use the **cycle spinning** trick proposed by Donoho and Johnstone. The copula density is estimated simultaneously in a collection of basis obtained by translations of a single wavelet basis and the resulting estimators are averaged. In our numerical experiments, we have performed this operation using 25 different translations and observed a significant improvement of the results.

6.2. Simulation

We focus on usual parametric families of copulas: the FGM, the Gaussian, the Student, the Clayton, the Frank and the Gumbel families. We give results for two very different values of n (the number of data): $n = 500$ which is very small for a bidimensional problem and $n = 2000$ which is usual in non-parametric estimation.

We test both methods of estimation (local thresholding and global thresholding) and, for each method, four different ways to solve the boundaries problems (zero padding, periodization, symmetrization and interval wavelets). In our experiments, the first marginal distribution is an exponential with parameter 4 and the second marginal distribution is the standard Gaussian. Let us remark here that the results obtained by our algorithm do not depend on the marginal distributions.

To evaluate the quality of our results, we consider three empirical loss functions derived from the L_1 norm, the L_2 norm and the L_∞ norm, that is to say

$$E_q = \|\widetilde{c} - c_0\|_{N,q} \quad \text{for } q = 1, 2, \infty,$$

where c_0 is the "true" copula density and $N \times N$ is the number of points of the grid (as described in the previous subsection). Tables A.3 and A.4 summarize the relative errors given by

$$RE^q = \frac{\|\widetilde{c} - c_0\|_{N,q}}{\|c_0\|_{N,q}} \quad \text{for } q = 1, 2, \infty.$$

These relative errors are computed with 100 repetitions of the experiment. The associated standard deviation is also given (in parentheses).

Tables A.1 and A.2 show that the zero padding method, the periodization method and, surprisingly, the boundary corrected method (which is, theoretically, the optimal construction) provide similar results. Moreover, they lead generally to much larger errors than the ones obtained by the symmetric periodization. This method appears to be the best one in order to solve the boundaries effects. This remark is valid for both sample size ($n = 500, 2000$). Although the zero padding method is the default method in the Matlab Wavelet Toolbox, it suffers from a severe drawback: it introduces strong discontinuities along the borders of $[0, 1]^d$. The periodization method suffers from the same drawback than the zero padding method as soon as the function is not really periodic. Fig. A.1 emphasizes the superiority of the symmetric periodization method in the case where the unknown copula density is a normal copula. While the copula estimated with symmetric extension remains close to the shape of the true copula up to a resolution issue, this is not the case for the two other estimated copulas. In the periodized version, the height of the extreme peaks is reduced and two spurious peaks corresponding to the periodization of the real peaks appear. The zero padded version is slightly better as it shows only the reduced height artifact. The bad performance of the boundary corrected method arises from a different issue: the difficulty of implementing a discrete numerical scheme corresponding exactly to the theoretical continuous construction. It explains also why this construction is only seldom implemented.

Tables A.3 and A.4 display the empirical L_1, L_2 and L_∞ estimation error for the symmetric extension for respectively $n = 500$ and $n = 2000$. They show that the best results are obtained for the L_2 norm for which the method has been designed. The second best results are obtained for the L_1 norm because a bound on the L_2 norm implies a bound on the L_1 norm. The estimation problem in L_∞ is much more challenging as it is not a consequence of the estimation in L_2.

Observe that the behavior strongly depends on the copula itself. This is coherent with the theory that states that the more "regular" the copula is, the more efficient the estimator will be. The copulas that are the least well estimated (Normal with parameter 0.9, Student with parameter 0.5 and Gumbel with parameter 8.33) are the most "irregular" ones: they are very "peaky". They are therefore not regular enough to be estimated correctly by the proposed method.

A final remark should be given on the difficulty to evaluate such errors. Whereas the L_1 norm is finite equal to 1 for all true copula, the L_2 and L_∞ norms can be very large (even infinite) because of their peaks. This is not an issue from the numerical point of view as we are restricted to a grid of step $1/N$ on which one can ensure the finiteness of the copula. Nevertheless the induced "empirical" norm can be substantially different from the integrated norm. Thus the error for $n = 500$ to $n = 2000$ are not strictly equivalent as the function can be much more complex for the resolution induced by $n = 2000$ than for $n = 500$.

6.3. Parametric estimation

Practitioners often use non-parametric estimators as a benchmark to choose the copula and its parameters among a family. We test our estimator in this framework by computing empirical distances

$$E_q(\theta, 0) = \|\hat{c} - c_\theta\|_{N,q} \quad \text{for } q = 1, 2, \infty$$

between the benchmark denoted $\hat{c}$ and a copula density c_θ varying in a fixed parametric family of copula densities $\mathcal{C}_0$. The corresponding natural estimator of the parameter θ is thus

$$\hat{\theta}_0^q = \arg\min_\theta E_q(\theta, 0).$$

Table A.9 gives the estimator θ for each norm with the a priori knowledge of the parametric family $\mathcal{C}_0$ from which the data are issued. As a benchmark, we have used the local thresholding with symmetrization.

From the theoretical point of view, this way to estimate θ is wrong as the empirical estimators of the error are biased. A much better choice would have been to estimate the error by its corresponding U-statistics as proposed in Gayraud and Tribouley [17]. Nevertheless, the numerical results are quite good as soon as the copula can be estimated efficiently. On the one hand, when the copula are too irregular, the corresponding estimate is a smoothed version and the estimate parameter corresponds to this smoothed version. On the other hand, when the Kendall's tau is small enough, the estimated parameters are close to the true parameter even if a slight bias toward a smoother copula can be observed.

6.4. Real data applications

We apply the thresholding methods on various financial series to identify the behavior of the dependence (or non-dependence). All data correspond to daily closing market quotations and are from 01/07/1987 to 31/01/2007. As usual, we consider the log-return of the data.

Notice that we apply our procedures even though the independence assumption is not necessarily satisfied by our data. We first propose estimators of the bivariate copula density associated with two financial series using the adaptive thresholding procedures (see Figs. A.2–A.5). Next, the non-parametric estimator denoted $\hat{c}$ is used as a benchmark and we derive a new estimator by choosing the copula amongst a parametric family of copula that minimizes the error between itself and the benchmark $\hat{c}$. Note that, contrary to the previous section, we do not want to impose an a priori knowledge on the parametric family. Nevertheless, we focus on copulas which belong to the Gaussian, Student, Gumbel, Clayton or Frank families. More precisely, we consider the following parametric classes of copulas

$$\mathcal{C}_1 = \{c \in \mathcal{N}_\theta, \theta = [-0.99 : 0.01 : 0.99]\}$$
$$\mathcal{C}_2 = \{c \in \mathcal{T}_\theta, \theta = [-0.99 : 0.01 : 0.99, 1 : 1 : 100]\}$$
$$\mathcal{C}_3 = \{c \in \mathcal{G}_\theta, \theta = [1 : 0.01 : 2]\}$$
$$\mathcal{C}_4 = \{c \in \mathcal{C}_\theta, \theta = [0 : 0.01 : 2]\}$$
$$\mathcal{C}_5 = \{c \in \mathcal{F}_\theta, \theta = [-2 : 0.01 : 2]\}$$

and we propose to estimate the parameter θ for each class $\mathcal{C}_p$ of copula densities, as in the previous subsection, by

$$\hat{\theta}_p^q = \arg\min_\theta E_q(\theta, p) \quad \text{for } p = 1, \ldots, 5.$$

We derive estimators of c among all the candidates $\left\{c_{\hat{\theta}_p^q}, p = 1, \ldots 5\right\}$ for each contrast $q = 1, 2, \infty$. Tables A.5–A.8 give

- the estimate $\tilde{\theta}^q$ for $q = 1, 2, \infty$ defined by
$$\tilde{\theta}^q = \arg\min_{p=1,\ldots,5} \left(\arg\min_\theta E_q(\theta, p)\right),$$
- the parametric family $\mathcal{C}_{\hat{p}}$ corresponding to the smallest error,
- the associated relative errors defined by
$$RE^q(\tilde{\theta}^q) = 100 \frac{\|\hat{c} - c_{\tilde{\theta}^q}\|_{N,q}}{\|c_{\tilde{\theta}^q}\|_{N,q}}$$
where c is in $\mathcal{C}_{\hat{p}}$.

We have tested a lot of financial series and have selected four revealing examples. In our tests, the Clayton family or the Gumbel family have never been selected; the selected family is always either the Gaussian family, the Student family or the Frank family.

The first observation is that the parametric families are quite well adapted since the relative error between the best fits and the non-parametric benchmark RE^q is always (much) smaller than 10% (except for the L_∞ norm). As expected, the results are quite similar for both thresholding methods. There is however a significant bias from the metric point of view toward the block approach. This bias can be seen, for example, in Fig. A.3, where the peaks have disappeared. Remark that this phenomenon occurs when the unknown copula density is not uniformly regular (when it does present high peaks). When this is not the case, as in the DowJones versus Ftse100uk, the local approach is more adapted. Nevertheless, the parameters estimated by the two different methods remain close.

The second observation is that the choice of the contrast is crucial to estimate the parameter θ: there are significant differences between $\hat{\theta}_1^p, \hat{\theta}_2^p, \hat{\theta}_\infty^p$. This is usual in density estimation as they do not measure the same behavior. The L_1 norm is our preferred choice. It seems natural in a density context and corresponds to a more robust criterion than the L_2 norm for which our theorems have been obtained. The L_∞ focuses on pointwise difference and, thus, is not adapted to the task. Nevertheless, the choice of the best family seems not to depend on the choice of the contrast: each type of parametric family is linked to a specific structure of dependence and are different enough to be identified whatever the criterion is.

We conclude this section with a few comments on the selected examples. The estimated copula for Cac versus Brent indicates that those series are independent. The copula densities DowJones versus Oncedor and Brent versus ExonMobil are both detected as Frank copulas but with opposite behaviors. Both results can be interpreted. It is obvious that the series Brent and ExonMobil should exhibit a strong dependence with a strong correlation. The negative dependence between Oncedor and the financial indices can be explained by the fact that Oncedor (gold) is a hedge when the stock market collapses. Remark that we observe the same kind of dependence of Oncedor for others composite indices such as Ftse100uk, Cac. The more delicate case is for the copula DowJones versus Ftse100uk. It is a very peaky copula and thus quite hard to estimate. We think nevertheless that the local thresholding method produces a nice estimate.

6.5. Conclusion

When the unknown copula density is uniformly regular (in the sense that it is not too peaky on the corners), the thresholding wavelet procedures associated with the symmetrization extension produce a good non-parametric estimation. If the copula presents strong peaks at the corner (for instance the Clayton copula with a large Kendall tau), our method is much less efficient. We think that improvements will come from a new family of wavelet adapted to singularity on the corners.

As shown in the numerical experiments, our procedures can be used in the popular two steps decision procedure: first use a non-parametric estimator to decide which copula family to consider and second estimate the parameters within this family. We do not claim that the plug-in method used with our estimate as a benchmark is optimal (it is slightly biased), but it provides a simple single framework. We did not study here the properties of such an estimator or of the corresponding goodness-of-fit test problem. We refer to Gayraud and Tribouley [17] for this last statistical issue.

7. Proofs

We first state two propositions needed to establish the main results. Next, we prove Theorem 3 in two steps by proving both implications. Last, we prove Proposition 1 and Corollary 5.1.

From now on, K denotes any constant that does not depend on j, k and n. Its value may change from one line to another and may depends on the wavelet, on $\|c\|_\infty$ and $\|c\|_2$.

7.1. Preliminaries

These preliminary results concern the estimation of the wavelet coefficients and the scaling coefficients (denoted $c_{j,k}^{\epsilon_0}$ with $\epsilon_0 = (0, \ldots, 0)$ to unify the notation). Proposition 3 shows that the accuracy of estimates is as sharp as if the direct observations were available.

Proposition 2. *Assume that the copula density belongs to $L_\infty([0,1]^d)$ and let $\delta > 0$. There exists a constant $K > 0$ such that for any j such that $2^j \le 2\left(\frac{n}{\log(n)}\right)^{1/d}$, and for any (k, ϵ)*

$$\mathbb{P}\left(|\widetilde{c_{j,k}^{\epsilon}} - \widehat{c_{j,k}^{\epsilon}}| > \lambda_n\right) \le K n^{-\delta} \tag{10}$$

$$\mathbb{P}\left(\sum_k (\widetilde{c_{j,k}^{\epsilon}} - \widehat{c_{j,k}^{\epsilon}})^2 > L^d 2^{dj} \lambda_n^2\right) \le K n^{1-\delta} (\log(n))^{-1} \tag{11}$$

provided κ is chosen large enough.

It is clear that (11) is a direct consequence of (10). The proof of (10) is relegated to the Appendix. From (10) we immediately deduce

Proposition 3. *Under the same assumptions on j and c as in Proposition 2, there exists a constant $K > 0$ such that for any (k, ϵ)*

$$E\left[\left(\widetilde{c_{j,k}^{\epsilon}} - \widehat{c_{j,k}^{\epsilon}}\right)^2\right] \le K \frac{\log(n)}{n}.$$

7.2. Proof of Theorem 3

First, we prove the result for the linear estimator. Secondly, we prove the result for the local thresholding method. We do not prove the result for the global thresholding method since the techniques are the same except that the required large deviation inequality is given by (11) instead of (10).

7.2.1. Proof of Equivalence (6)

Let c be a copula density function belonging to $L_\infty([0,1]^d)$ and satisfying for any n,

$$\mathbb{E}\|\widetilde{c_L} - c\|_2^2 \le K \left(\frac{\log(n)}{n}\right)^{\frac{2s}{2s+d}} \tag{12}$$

for some constant $K > 0$. Let us prove that c also belongs to the space $\mathcal{B}_{2\infty}^s$. Let us recall that the smoothing index used for the linear procedure is j_n^* and it satisfies $2^{1-j_n^*} > \left(n^{-1} \log(n)\right)^{1/(2s+d)}$. Since

$$\mathbb{E}\|\widetilde{c_L} - c\|_2^2 = \mathbb{E}\left\|\widetilde{c_L} - \sum_k c_{j_n^*,k}\phi_{j_n^*,k}\right\|_2^2 + \left\|\sum_{j \ge j_n^*}\sum_{k,\epsilon} c_{j,k}^{\epsilon}\psi_{j,k}^{\epsilon}\right\|_2^2,$$

the assumption (12) implies

$$\sum_{j\geq j_n^*}\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2 \leq \mathbb{E}\|\widetilde{c_L}-c\|_2^2 \leq K\ (2^{-2j_n^*})^s.$$

So $c\in\mathcal{B}_{2\infty}^s$.

Conversely, let us suppose that $c\in\mathcal{B}_{2\infty}^s$. Then, using the same techniques as in Genest et al. [6], we can show that for any n

$$\mathbb{E}\|\widetilde{c_L}-c\|_2^2 \leq K\left(\frac{\log(n)}{n}\right)^{\frac{2s}{2s+d}}$$

which ends the proof. The proof in Genest et al. [6] is given in the case $d=2$ and uses a sharp control on the estimated coefficients.

7.2.2. Proof of Equivalence (7) (first step: $\Longrightarrow$)

When the direct observations $(F_1(X_i^1),\ldots,F_d(X_i^d))$ are available, we use the estimator $\widehat{c_{HL}}$ built in the same way as $\widetilde{c_{HL}}$ but with the sequence of coefficients $\widehat{c_{j,k}^{\epsilon}}$ defined in (1) and with the threshold $\lambda_n/2$ instead of λ_n. Let j_n, J_n be positive integers and $\lambda_n>0$. We get

$$E\|\widetilde{c_{HL}}-c\|_2^2 \leq 2E\|\widetilde{c_{HL}}-\widehat{c_{HL}}\|_2^2+2E\|\widehat{c_{HL}}-c\|_2^2.$$

First, we study the error term due to the pseudo-observations

$$\begin{aligned}T &= E\|\widetilde{c_{HL}}-\widehat{c_{HL}}\|_2^2\\ &= E\left[\sum_k(\widetilde{c_{j_nk}^{\epsilon_0}}-\widehat{c_{j_nk}^{\epsilon_0}})^2\right]+E\left[\sum_{j_n}^{J_n}\sum_{k,\epsilon}(\widetilde{c_{j,k}^{\epsilon}}-\widehat{c_{j,k}^{\epsilon}})^2\mathbf{1}\{|\widetilde{c_{j,k}^{\epsilon}}|>\lambda_n\}\mathbf{1}\left\{|\widehat{c_{j,k}^{\epsilon}}|>\frac{\lambda_n}{2}\right\}\right]\\ &\quad+E\left[\sum_{j_n}^{J_n}\sum_{k,\epsilon}(\widehat{c_{j,k}^{\epsilon}})^2\mathbf{1}\{|\widetilde{c_{j,k}^{\epsilon}}|\leq\lambda_n\}\mathbf{1}\left\{|\widehat{c_{j,k}^{\epsilon}}|>\frac{\lambda_n}{2}\right\}\right]+E\left[\sum_{j_n}^{J_n}\sum_{k,\epsilon}(\widetilde{c_{j,k}^{\epsilon}})^2\mathbf{1}\{|\widetilde{c_{j,k}^{\epsilon}}|>\lambda_n\}\mathbf{1}\left\{|\widehat{c_{j,k}^{\epsilon}}|\leq\frac{\lambda_n}{2}\right\}\right]\\ &= T_1+T_2+T_3+T_4.\end{aligned}$$

Using Proposition 3, we have

$$T_1\leq K\frac{\log(n)}{n}2^{dj_n}\leq K\frac{(\log(n))^2}{n}. \tag{13}$$

To study T_2, we apply Cauchy–Schwarz inequality and we obtain

$$\begin{aligned}T_2 &= E\left[\sum_{j_n}^{J_n}\sum_{k,\epsilon}(\widetilde{c_{j,k}^{\epsilon}}-\widehat{c_{j,k}^{\epsilon}})^2\mathbf{1}\left\{|\widetilde{c_{j,k}^{\epsilon}}|>\lambda_n\right\}\mathbf{1}\left\{|\widehat{c_{j,k}^{\epsilon}}|>\frac{\lambda_n}{2}\right\}\left(\mathbf{1}\left\{|c_{j,k}^{\epsilon}|\leq\frac{\lambda_n}{4}\right\}+\mathbf{1}\left\{|c_{j,k}^{\epsilon}|>\frac{\lambda_n}{4}\right\}\right)\right]\\ &\leq\sum_{j_n}^{J_n}\sum_{k,\epsilon}\left[E(\widetilde{c_{j,k}^{\epsilon}}-\widehat{c_{j,k}^{\epsilon}})^4\right]^{1/2}\left[\mathbb{P}\left(|\widehat{c_{j,k}^{\epsilon}}-c_{j,k}^{\epsilon}|>\frac{\lambda_n}{4}\right)\right]^{1/2}+\sum_{j_n}^{J_n}\sum_{k,\epsilon}E(\widetilde{c_{j,k}^{\epsilon}}-\widehat{c_{j,k}^{\epsilon}})^2\mathbf{1}\left\{|c_{j,k}^{\epsilon}|>\frac{\lambda_n}{4}\right\}.\end{aligned}$$

Observe that, for any j,k,ϵ, we have

$$|\widetilde{c_{j,k}^{\epsilon}}|\vee|\widehat{c_{j,k}^{\epsilon}}|\leq 2^{jd/2}(\|\psi\|_\infty^d\vee\|\phi\|_\infty^d). \tag{14}$$

For any $\delta>0$, we use now the standard Bernstein Inequality to obtain

$$\mathbb{P}\left(|\widehat{c_{j,k}^{\epsilon}}-c_{j,k}^{\epsilon}|>\frac{\lambda_n}{4}\right)\leq Kn^{-\delta}. \tag{15}$$

This inequality is valid for a choice of κ large enough. Let us now fix r in $]0,2[$. Applying Proposition 3 and using (14), we have

$$\begin{aligned}T_2 &\leq K\sum_{j_n}^{J_n}\sum_{k,\epsilon}2^{jd}\left[\mathbb{P}\left(|\widehat{c_{j,k}^{\epsilon}}-c_{j,k}^{\epsilon}|>\frac{\lambda_n}{4}\right)\right]^{1/2}+\sum_{j_n}^{J_n}\sum_{k,\epsilon}E(\widetilde{c_{j,k}^{\epsilon}}-\widehat{c_{j,k}^{\epsilon}})^2\mathbf{1}\left\{|c_{j,k}^{\epsilon}|>\frac{\lambda_n}{4}\right\}\\ &\leq K\left(2^{2dJ_n}n^{-\delta/2}+u_n\left[\left(\frac{\lambda_n}{4}\right)^r\sum_{j_n}^{J_n}\sum_{k,\epsilon}\mathbf{1}\left\{|c_{j,k}^{\epsilon}|>\frac{\lambda_n}{4}\right\}\right]\right)\end{aligned} \tag{16}$$

where $u_n = (\lambda_n/4)^{-r}(\log(n)/n)$. Similarly, we have

$$T_3 \le E\left[\sum_{j_n}^{J_n}\sum_{k,\epsilon}(\widehat{c_{j,k}^{\epsilon}})^2\mathbf{1}\{|\widetilde{c_{j,k}^{\epsilon}}| \le \lambda_n\}\mathbf{1}\left\{|\widehat{c_{j,k}^{\epsilon}}| > \frac{\lambda_n}{2}\right\} \times \left(\mathbf{1}\left\{|c_{j,k}^{\epsilon}| \le \frac{\lambda_n}{4}\right\} + \mathbf{1}\left\{|c_{j,k}^{\epsilon}| > \frac{\lambda_n}{4}\right\}\right)\right]$$
$$\le E\left[\sum_{j_n}^{J_n}\sum_{k,\epsilon}(\widehat{c_{j,k}^{\epsilon}})^2\mathbf{1}\left\{|\widehat{c_{j,k}^{\epsilon}}| > \frac{\lambda_n}{2}\right\}\mathbf{1}\left\{|c_{j,k}^{\epsilon}| \le \frac{\lambda_n}{4}\right\}\right] + E\left[\sum_{j_n}^{J_n}\sum_{k,\epsilon}(\widehat{c_{j,k}^{\epsilon}})^2\mathbf{1}\{|\widetilde{c_{j,k}^{\epsilon}}| \le \lambda_n\}\mathbf{1}\left\{|c_{j,k}^{\epsilon}| > \frac{\lambda_n}{4}\right\}\right]$$
$$\le K\sum_{j_n}^{J_n}\sum_{k,\epsilon}2^{dj}\mathbb{P}\left(|\widehat{c_{j,k}^{\epsilon}} - c_{j,k}^{\epsilon}| > \frac{\lambda_n}{4}\right) + \left(\frac{\lambda_n}{4}\right)^r\sum_{j_n}^{J_n}\sum_{k,\epsilon}v_n\mathbf{1}\left\{|c_{j,k}^{\epsilon}| > \frac{\lambda_n}{4}\right\}$$

where

$$v_n = 2\left(\frac{\lambda_n}{4}\right)^{-r}\left[E(\widetilde{c_{j,k}} - \widehat{c_{j,k}})^2 + E(\widetilde{c_{j,k}})^2\mathbf{1}\{|\widetilde{c_{j,k}}| \le \lambda_n\}\right]$$
$$\le 2\left(Ku_n + 4^r\lambda_n^{2-r}\right).$$

This implies

$$T_3 \le K\left(\frac{2^{2dJ_n}}{n^{\delta}} + (u_n + \lambda_n^{2-r})\left[\left(\frac{\lambda_n}{4}\right)^r\sum_{j_n}^{J_n}\sum_{k,\epsilon}\mathbf{1}\left\{|c_{j,k}^{\epsilon}| > \frac{\lambda_n}{4}\right\}\right]\right). \tag{17}$$

Using (14) and Proposition 2, we get

$$T_4 \le K\sum_{j_n}^{J_n}\sum_{k,\epsilon}2^{dj}\mathbb{P}\left(|\widetilde{c_{j,k}^{\epsilon}} - \widehat{c_{j,k}^{\epsilon}}| > \frac{\lambda_n}{2}\right) \le K2^{2dJ_n}n^{-\delta}. \tag{18}$$

Combining the bounds of (13) and (16)–(18) and choosing j_n, J_n as indicated in Theorem 2, we get for $\delta \ge 6$

$$E\|\widetilde{c_{HL}} - c\|_2^2 \le 2\,E\|\widehat{c_{HL}} - c\|_2^2 + K\rho_n$$

where

$$\rho_n = \frac{(\log(n))^2}{n} + \left(\frac{\log n}{n}\right)^{1-\frac{r}{2}}\left(\frac{\lambda_n}{4}\right)^r\sum_{j_n}^{J_n}\sum_{k,\epsilon}\mathbf{1}\left\{|c_{j,k}^{\epsilon}| > \frac{\lambda_n}{4}\right\} + \frac{1}{n(\log(n))^2}.$$

On the one hand, let us suppose that c belongs to the weak Besov space $\mathcal{W}_L(\frac{2d}{2s+d})$. For $r := 2d/(2s+d)$,

$$\left(\frac{\lambda_n}{4}\right)^r\sum_{j_n}^{J_n}\sum_{k,\epsilon}\mathbf{1}\left\{|c_{j,k}^{\epsilon}| > \frac{\lambda_n}{4}\right\} \le K.$$

It follows that

$$\rho_n \le K\left(\frac{\log(n)}{n}\right)^{\frac{2s}{2s+d}}.$$

Using the standard result when direct observations are available, we also have

$$E\|\widehat{c_{HL}} - c\|_2^2 \le K\left(\frac{\log(n)}{n}\right)^{\frac{2s}{2s+d}}$$

for $c \in \mathcal{W}_L(\frac{2d}{2s+d}) \cap \mathcal{B}^s_{2\infty}$. This ends the proof of the first part of (7) of Theorem 3.

7.2.3. Proof of Equivalence (7) (second step: $\Longleftarrow$)

Suppose that there exists M such that for any n,

$$E\|\widetilde{c_{HL}} - c\|_2^2 \le M\left(n^{-1}\log(n)\right)^{\frac{2s}{2s+d}}.$$

Since

$$\sum_{j>J_n}\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2 \le E\|\widetilde{c_{HL}} - c\|_2^2,$$

and, setting J_n as indicated in Theorem 2, we obtain

$$\sum_{j>J_n}\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2 \le M\left(\frac{\log(n)}{n}\right)^{\frac{2s}{2s+d}} \le M\left(2^{d(1-J_n)}\right)^{\frac{2s}{2s+d}} \le K\left(2^{-2J_n}\right)^{\frac{ds}{2s+d}}.$$

Using Definition 1 of the strong Besov bodies, we deduce that c belongs to $\mathcal{B}_{2\infty}^{\frac{ds}{2s+d}}$. Let us now study the sum of squares of the small detail coefficients

$$\sum_{j\ge 0}\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2\mathbf{1}\left\{|c_{j,k}^{\epsilon}|\le\frac{\lambda_n}{2}\right\} = \left[\sum_{j<j_n}+\sum_{j=j_n}^{J_n}+\sum_{j>J_n}\right]\left[\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2\mathbf{1}\{|c_{j,k}^{\epsilon}|\le\lambda_n/2\}\right]$$
$$\le H_1+H_2+H_3. \tag{19}$$

We have already proved that $c\in\mathcal{B}_{2\infty}^{\frac{ds}{2s+d}}$. Setting λ_n as indicated in Theorem 2, we deduce

$$H_3\le\sum_{j>J_n}\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2\le K2^{-2J_n\frac{ds}{2s+d}}\le K\left(\frac{\lambda_n}{2}\right)^{\frac{4s}{2s+d}}. \tag{20}$$

Taking j_n as in Theorem 2, we get

$$H_1\le K\sum_{j<j_n}2^{dj}\left(\frac{\lambda_n}{2}\right)^2\le K\log(n)\left(\frac{\lambda_n}{2}\right)^2\le K\left(\frac{\lambda_n}{2}\right)^{\frac{4s}{2s+d}}. \tag{21}$$

Observe that

$$H_2=E\left[\sum_{j_n}^{J_n}\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2\mathbf{1}\left\{|c_{j,k}^{\epsilon}|\le\frac{\lambda_n}{2}\right\}\left(\mathbf{1}\{|\widetilde{c_{j,k}^{\epsilon}}|\le\lambda_n\}+\mathbf{1}\{|\widetilde{c_{j,k}^{\epsilon}}|>\lambda_n\}\right)\right].$$

Remembering that

$$E\left[\sum_{j_n}^{J_n}\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2\mathbf{1}\{|\widetilde{c_{j,k}^{\epsilon}}|\le\lambda_n\}\right]\le E\|\widetilde{c_{HL}}-c\|_2^2$$

and using Proposition 2 and (15), we get

$$H_2\le E\|\widetilde{c_{HL}}-c\|_2^2+\sum_{j_n}^{J_n}\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2\mathbb{P}\left(|\widetilde{c_{j,k}^{\epsilon}}-\widehat{c_{j,k}^{\epsilon}}|>\frac{\lambda_n}{4}\right)+\sum_{j_n}^{J_n}\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2\mathbb{P}\left(|\widehat{c_{j,k}^{\epsilon}}-c_{j,k}^{\epsilon}|>\frac{\lambda_n}{4}\right)$$
$$\le M\left(\frac{\log(n)}{n}\right)^{\frac{2s}{2s+d}}+K\|c\|_2^2n^{-\delta}\le K\left(\frac{\lambda_n}{2}\right)^{\frac{4s}{2s+d}} \tag{22}$$

for δ larger than 1. Combining (20)–(22), we conclude that $c\in\mathcal{W}_L(r)$ with r such that $2-r=4s/(2s+d)$. Hence, we end the proof of the indirect direction of (7).

7.3. Proofs of Proposition 1 and Corollary 5.1

The proof of the **inclusion** given in Proposition 1 follows immediately from the definitions of the functional spaces. Let $c_{j,k}^{\epsilon}$ denote the sequence of wavelet coefficients of a function c. We have

$$\sup_{0<\lambda\le 1}\lambda^{r-2}\sum_{j\ge 0}\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2\mathbf{1}\{|c_{j,k}^{\epsilon}|\le\lambda\}$$
$$=\sup_{0<\lambda\le 1}\lambda^{r-2}\sum_{j\ge 0}\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2\mathbf{1}\{|c_{j,k}^{\epsilon}|\le\lambda\}\left[\mathbf{1}\left\{\sum_k(c_{j,k}^{\epsilon})^2\le 2^{dj}\lambda^2\right\}+\mathbf{1}\left\{\sum_k(c_{j,k}^{\epsilon})^2>2^{dj}\lambda^2\right\}\right]$$
$$\le\sup_{0<\lambda\le 1}\lambda^{r-2}\sum_{j\ge 0}\sum_{k,\epsilon}(c_{j,k}^{\epsilon})^2\mathbf{1}\left\{\sum_k(c_{j,k}^{\epsilon})^2\le 2^{dj}\lambda^2\right\}+K\sup_{0<\lambda\le 1}\lambda^r\sum_{j\ge 0}2^{dj}\sum_{\epsilon}\mathbf{1}\left\{\sum_k(c_{j,k}^{\epsilon})^2>2^{dj}\lambda^2\right\}.$$

It follows from Definition 3 that

$$c\in\mathcal{W}_G(r)\Rightarrow c\in\mathcal{W}_L(r).$$

To establish the **strict inclusions**, we build a sparse function belonging to $\mathcal{B}_{2\infty}^{\frac{ds}{2s+d}} \cap \mathcal{W}_L(\frac{2d}{2s+d})$ but not belonging to $\mathcal{W}_G(\frac{2d}{2s+d})$. Let us choose a real number α such that $d/2 \le \alpha < s+d/2$. Let us consider the sparse sequence $c_{j,k}^{\epsilon}$ in which all coefficients $c_{j,k}^{\epsilon}$ are set to 0 except for the $\lfloor 2^{j(2d\alpha)/(2s+d)} \rfloor$ first ones at each scale that are set to $(2^d-1)^{-1}2^{-\alpha j}$. Let c be the corresponding function. For all $0 < \lambda \le 1$, let j_λ be such that $2^{j_\lambda} = \left((2^d-1)\lambda\right)^{-1/\alpha}$. We get

$$\begin{aligned}\sum_{j\ge 0}\sum_{k,\epsilon} \mathbf{1}\{|c_{j,k}^{\epsilon}| > \lambda\} &= \sum_{j<j_\lambda}\sum_{k,\epsilon} \mathbf{1}\{|c_{j,k}^{\epsilon}| > \lambda\} \\ &\le K 2^{\frac{2d\alpha}{2s+d}j_\lambda} \le K\lambda^{-\frac{2d}{2s+d}}\end{aligned}$$

implying

$$\sup_{0<\lambda\le 1} \lambda^{\frac{2d}{2s+d}} \sum_{j\ge 0}\sum_{k,\epsilon} \mathbf{1}\{|c_{j,k}^{\epsilon}| > \lambda\} < \infty.$$

Thus the function c belongs to the local weak Besov space $\mathcal{W}_L(\frac{2d}{2s+d})$. Next, let us put $\alpha' = (4\alpha s + 2sd + d^2)/(2(2s+d))$. We observe that $\alpha' < s + d/2$ since $\alpha < s + d/2$. For all $0 < \lambda \le 1$ let j_λ^* be such that $2^{j_\lambda^*} = \left((2^d-1)\lambda\right)^{-1/\alpha'}$. We get

$$\begin{aligned}\sum_{j\ge 0} 2^{dj} \sum_{\epsilon} \mathbf{1}\left\{\sum_{k} (c_{j,k}^{\epsilon})^2 > 2^{dj}\lambda^2\right\} &\ge (2^d-1)\sum_{j<j_\lambda^*} 2^{dj} \\ &\ge 2^{dj_\lambda^*-1} \ge K\lambda^{-\frac{d}{\alpha'}}\end{aligned}$$

and thus

$$\sup_{0<\lambda\le 1} \lambda^{\frac{2d}{2s+d}} \sum_{j\ge 0} 2^{dj} \sum_{\epsilon} \mathbf{1}\left\{\sum_{k} (c_{j,k}^{\epsilon})^2 > 2^{dj}\lambda^2\right\} = \infty.$$

It follows that the function c does not belong to the global weak Besov space $\mathcal{W}_G(\frac{2d}{2s+d})$. This ends the proof of Proposition 1. Notice that the function c belongs to the strong Besov body $\mathcal{B}_{2\infty}^{\frac{ds}{2s+d}}$ because for any (j, ϵ)

$$\sum_{k,\epsilon} (c_{j,k}^{\epsilon})^2 \le 2^{\frac{2d\alpha}{2s+d}j} 2^{-2\alpha j} \le 2^{-\frac{2ds}{2s+d}j}$$

so

$$\sup_{J\ge 0} 2^{\frac{2ds}{2s+d}J} \sum_{j\ge J}\sum_{k,\epsilon} (c_{j,k}^{\epsilon})^2 < \infty,$$

which proves Corollary 5.1.

Appendix

In this section we prove (10) of Proposition 2. In the sequel, we fix the indices j and $k = (k_1, \ldots, k_d)$ and we take without loss of generality $\epsilon = 2^d - 1$. For any $i = 1, \ldots, n$ (the observation index) and any $m = 1, \ldots d$ (the coordinate index), let us introduce

$$\begin{aligned}\Delta(X_i^m) &= \widehat{F_m}(X_i^m) - F_m(X_i^m), \\ \xi_j(X_i^m) &= \psi_{j,k_m}(\widehat{F_m}(X_i^m)) - \psi_{j,k_m}(F_m(X_i^m)), \\ N_j(m) &= \#\left\{i \in \{1, \ldots, n\};\ \xi_j(X_i^m) \neq 0\right\},\end{aligned}$$

and

$$\begin{aligned}\xi_j(X_i^1, \ldots, X_i^m) &= \psi_{j,k}^{\epsilon}(\widehat{F_1}(X_i^1), \ldots, \widehat{F_d}(X_i^d)) - \psi_{j,k}^{\epsilon}(F_1(X_i^1), \ldots, F_d(X_i^d)) \\ N_j &= \#\left\{i \in \{1, \ldots, n\};\ \xi_j(X_i^1, \ldots, X_i^d) \neq 0\right\}.\end{aligned}$$

As previously remarked in Genest et al. [6] in the case $d = 2$, we have

$$\xi_j(X_i^1, \ldots, X_i^d) = \prod_{m=1}^{d} \xi_j(X_i^m) + \sum_{m_1=1}^{d}\left[\psi_{j,k_{m_1}}^{\epsilon}(F_{m_1}(X_i^{m_1})) \prod_{\substack{m=1\\ m\neq m_1}}^{d} \xi_j(X_i^m)\right]$$

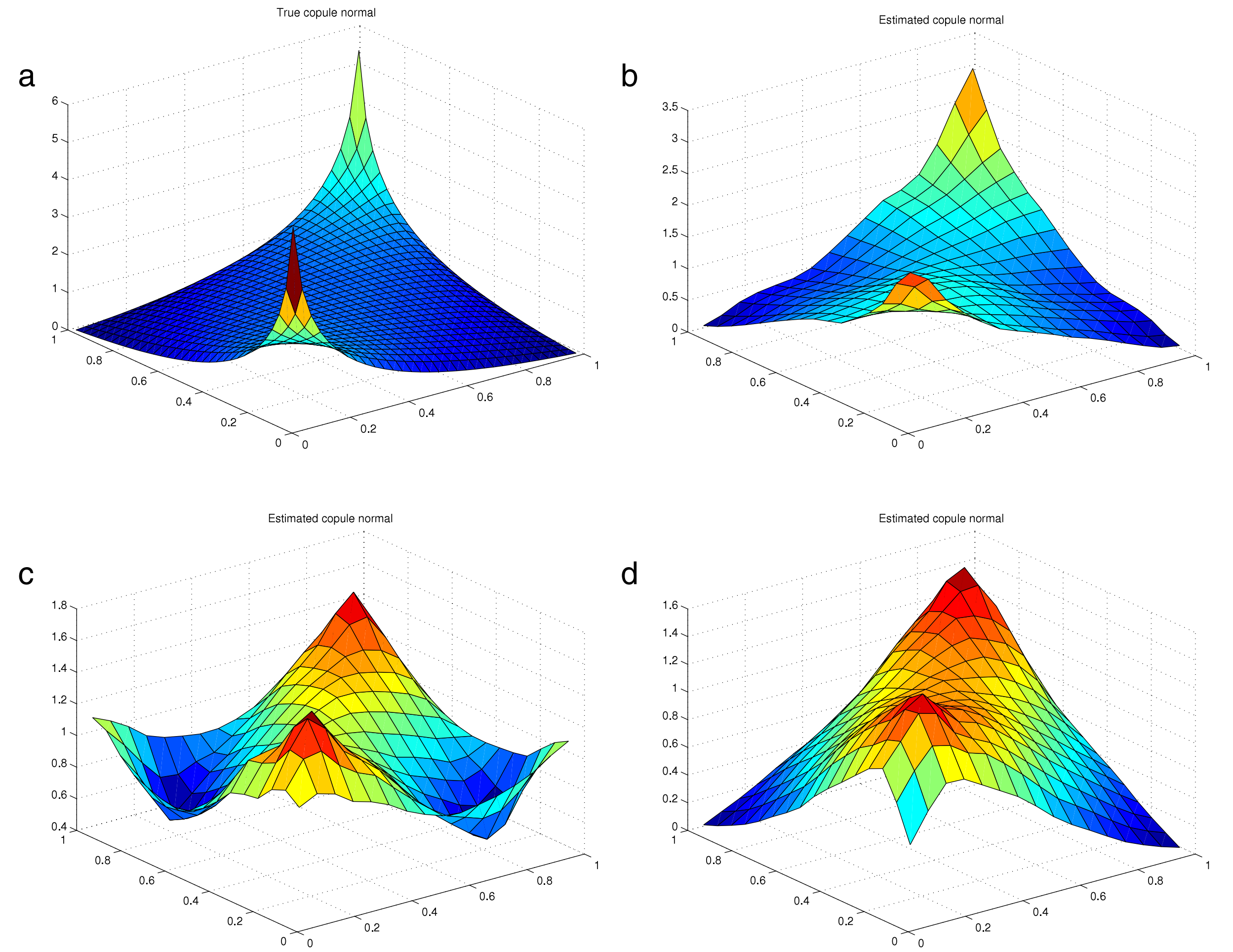


Fig. A.1. Estimation of the normal copula density of parameter 0.5 with $n = 2000$ (local thresholding): (a) true copula, (b) estimated copula with symmetrization, (c) estimated copula with periodization, (d) estimated copula with zero padding.

$$+ \sum_{\substack{m_1,m_2=1\\ m_1\neq m_2}}^{d} \left[\psi^{\epsilon}_{j,k_{m_1}}(F_{m_1}(X_i^{m_1}))\psi^{\epsilon}_{j,k_{m_2}}(F_{m_2}(X_i^{m_2})) \prod_{\substack{m=1\\ m\neq m_1,m_2}}^{d} \xi_j(X_i^m) \right]$$

$$+\cdots+ \sum_{m_1=1}^{d} \left[\xi_j(X_i^{m_1}) \prod_{\substack{m=1\\ m\neq m_1}}^{d} \psi^{\epsilon}_{j,k_m}(F_m(X_i^m)) \right] \tag{A.1}$$

In the sequel, for $m = 1, \ldots, d$, $T_{m,j}(X_i)$ denotes any term of the type

$$\left[\psi^{\epsilon}_{j,k_1}(F_1(X_i^1)) \times \cdots \times \psi^{\epsilon}_{j,k_{d-m}}(F_{d-m}(X_i^{d-m}))\right]\left[\xi_j(X_i^{d-m+1}) \times \cdots \times \xi_j(X_i^d)\right]$$

i.e. such that there are exactly m factors $\xi_j(X_i^{\cdot})$ appearing in the product. The cardinality of such terms $T_{m,j}(X_i)$ is equal to $C_d^m = \frac{d!}{m!(d-m)!}$. Observe that the number of terms in (A.1) is $2^d - 1$. It is fundamental to notice that there is no term $T_{0,j}(X_i) = \prod_{m=1,\ldots,d} \psi^{\epsilon}_{j,k_m}(F_m(X_i^m))$.

A.1. Technical lemmas

We begin by technical lemmas.

Lemma 1. *There exists a universal constant K_0 such that for any $m \in \{1, \ldots, d\}$*

$$\forall t > 0, \mathbb{P}(\max_{1\leq i\leq n} |\Delta(X_i^m)| > t) \leq K_0 \exp(-2nt^2).$$

Table A.1
Relative L_2 estimation error for $n = 500$.

Copula		Method	Boundary handling			
$c(\cdot)$	par.		sym	per	ZeroPad	Boundary
FGM	1.0	Local	0.007 (0.003)	0.079 (0.005)	0.129 (0.010)	0.096 (0.013)
		Block	0.006 (0.002)	0.077 (0.008)	0.141 (0.006)	0.074 (0.004)
Normal	0.0	Local	0.002 (0.002)	0.0004 (0.0004)	0.122 (0.005)	0.042 (0.009)
		Block	0.002 (0.002)	0.0004 (0.0006)	0.105 (0.001)	0.013 (0.002)
Normal	0.5	Local	0.031 (0.007)	0.161 (0.011)	0.179 (0.010)	0.158 (.0008)
		Block	0.032 (0.008)	0.154 (0.011)	0.202 (0.005)	0.189 (0.007)
Normal	**0.9**	Local	**0.156 (0.011)**	**0.391 (0.008)**	**0.418 (0.006)**	**0.406 (0.007)**
		Block	**0.140 (0.009)**	**0.381 (0.005)**	**0.491 (0.022)**	**0.406 (0.007)**
Student	**(0.5, 1)**	Local	**0.326 (0.018)**	**0.460 (0.008)**	**0.544 (0.009)**	**0.488 (0.010)**
		Block	**0.324 (0.026)**	**0.458 (0.010)**	**0.585 (0.004)**	**0.475 (0.015)**
Clayton	0.8	Local	0.075 (0.013)	0.225 (0.010)	0.252 (0.011)	0.213 (0.011)
		Block	0.095 (0.012)	0.216 (0.011)	0.279 (0.005)	0.216 (0.007)
Frank	4	Local	0.021 (0.006)	0.149 (0.015)	0.212 (0.015)	0.140 (0.009)
		Block	0.013 (0.006)	0.134 (0.009)	0.193 (0.006)	0.140 (0.007)
Gumbel	**8.3**	Local	**0.701 (0.002)**	**0.849 (0.001)**	**0.866 (0.001)**	**0.854 (0.001)**
		Block	**0.698 (0.002)**	**0.852 (0.001)**	**0.878 (0.001)**	**0.858 (0.001)**
Gumbel	1.25	Local	0.038 (0.010)	0.104 (0.005)	0.172 (0.009)	0.125 (0.013)
		Block	0.052 (0.007)	0.109 (0.004)	0.173 (0.004)	0.104 (0.003)

Table A.2
Relative L_2 estimation error for $n = 2000$.

Copula		Method	Boundary handling			
$c(\cdot)$	par.		sym	per	ZeroPad	Boundary
FGM	1.0	Local	0.004 (0.001)	0.066 (0.004)	0.090 (0.004)	0.064 (.003)
		Block	0.004 (0.002)	0.060 (0.003)	0.107 (0.004)	0.065 (.002)
Normal	0.0	Local	0.0006 (0.0005)	0.0001 (0.0001)	0.082 (0.001)	0.011 (0.002)
		Block	0.0006 (0.0007)	0.0001 (0.0001)	0.091 (0.002)	0.010 (0.001)
Normal	0.5	Local	0.017 (0.003)	0.145 (0.004)	0.142 (0.005)	0.141 (.002)
		Block	0.017 (0.003)	0.133 (0.004)	0.152 (0.005)	0.146 (.003)
Normal	**0.9**	Local	**0.138 (0.005)**	**0.389 (0.003)**	**0.402 (0.003)**	**0.395 (0.003)**
		Block	**0.133 (0.004)**	**0.381 (0.003)**	**0.426 (0.005)**	**0.391 (.002)**
Student	**(0.5, 1)**	Local	**0.296 (0.006)**	**0.452 (0.004)**	**0.516 (0.004)**	**0.459 (0.003)**
		Block	**0.288 (0.006)**	**0.447 (0.003)**	**0.523 (0.003)**	**0.450 (0.004)**
Clayton	0.8	Local	0.060 (0.005)	0.207 (0.005)	0.213 (0.004)	0.206 (.003)
		Block	0.060 (0.005)	0.197 (0.003)	0.225 (0.007)	0.212 (0.005)
Frank	4	Local	0.0121 (0.003)	0.124 (0.005)	0.119 (0.004)	0.132 (0.003)
		Block	0.007 (0.002)	0.114 (0.003)	0.122 (0.005)	0.137 (0.004)
Gumbel	**8.3**	Local	**0.697 (0.002)**	**0.851 (0.001)**	**0.866 (0.001)**	**0.855 (0.001)**
		Block	**0.697 (0.001)**	**0.852 (0.001)**	**0.864 (0.001)**	**0.853 (0.001)**
Gumbel	1.25	Local	0.024 (0.004)	0.102 (0.003)	0.139 (0.003)	0.103 (0.002)
		Block	0.033 (0.004)	0.099 (0.003)	0.150 (0.004)	0.101 (0.001)

Lemma 1 is a consequence of Dvoreski–Kiefer–Wolfowitz Inequality.

Lemma 2. *Let $\delta > 0$ and let n be an integer such that $n \log(n) \geq 2(\delta^{-1} \vee 1)$. Then, there exists $K_1 > 0$ such that for any level j satisfying*

$$2^j \leq \frac{1}{3}\left(\frac{2n}{\delta \log(n)}\right)^{1/2},$$

and for any $m \in \{1, \ldots, d\}$,

$$\mathbb{P}(N_j(m) > (L+3)n2^{-j}) \vee \mathbb{P}(N_j > d(L+3)n2^{-j}) \leq K_1 n^{-\delta}. \tag{A.2}$$

For the interested reader, the detailed proofs of these lemmas are given in Autin et al. [16].

Table A.3
Relative L_1, L_2 and L_∞ estimation errors for $n = 500$.

Copula		Method	Empirical loss function		
$c(\cdot)$	par.		L_1	L_2	L_∞
FGM	1.0	Local	0.062 (0.014)	0.007 (0.003)	0.189 (0.051)
		Block	0.061 (0.011)	0.006 (0.002)	0.175 (0.047)
Normal	0.0	Local	0.038 (0.017)	0.002 (0.002)	0.145 (0.062)
		Block	0.038 (0.018)	0.002 (0.002)	0.129 (0.058)
Normal	0.5	Local	0.118 (0.012)	0.031 (0.007)	0.539 (0.066)
		Block	0.112 (0.016)	0.032 (0.008)	0.555 (0.051)
Normal	**0.9**	Local	**0.287 (0.026)**	**0.156 (0.011)**	**0.648 (0.020)**
		Block	**0.205 (0.021)**	**0.140 (0.009)**	**0.644 (0.018)**
Student	**(0.5, 1)**	Local	**0.290 (0.022)**	**0.326 (0.018)**	**0.791 (0.026)**
		Block	**0.259 (0.018)**	**0.324 (0.026)**	**0.797 (0.035)**
Clayton	0.8	Local	0.119 (0.014)	0.075 (0.013)	0.658 (0.051)
		Block	0.125 (0.018)	0.095 (0.012)	0.740 (0.040)
Frank	4	Local	0.129 (0.017)	0.021 (0.006)	0.329 (0.075)
		Block	0.092 (0.020)	0.013 (0.006)	0.321 (0.069)
Gumbel	**8.3**	Local	**0.682 (0.015)**	**0.701 (0.002)**	**0.914 (0.001)**
		Block	**0.629 (0.012)**	**0.698 (0.002)**	**0.915 (0.001)**
Gumbel	1.25	Local	0.099 (0.011)	0.038 (0.010)	0.625 (0.104)
		Block	0.105 (0.012)	0.052 (0.007)	0.749 (0.044)

Table A.4
Relative L_1, L_2 and L_∞ estimation errors for $n = 2000$.

Copula		Method	Empirical loss function		
$c(\cdot)$	par.		L_1	L_2	L_∞
FGM	1.0	Local	0.0448 (0.00821)	0.0036 (0.0012)	0.1414 (0.0382)
		Block	0.04887 (0.0096)	0.0037 (0.0015)	0.1463 (0.0527)
Normal	0.0	Local	0.0181 (0.0087)	0.00063 (0.0005)	0.0673 (0.0332)
		Block	0.0190 (0.0092)	0.0006 (0.0007)	0.0669 (0.0284)
Normal	0.5	Local	0.0830 (0.0078)	0.0176 (0.0032)	0.4374 (0.0465)
		Block	0.0923 (0.0104)	0.0177 (0.0029)	0.4089 (0.0673)
Normal	**0.9**	Local	**0.2048 (0.0160)**	**0.1376 (0.00522)**	**0.6400 (0.0114)**
		Block	**0.1622 (0.0113)**	**0.1330 (0.0045)**	**0.6389 (0.0106)**
Student	**(0.5, 1)**	Local	**0.2159 (0.0107)**	**0.2966 (0.0056)**	**0.7712 (0.0110)**
		Block	**0.1955 (0.0095)**	**0.2881 (0.0058)**	**0.7669 (0.0133)**
Clayton	0.8	Local	0.0862 (0.0068)	0.0603 (0.0053)	0.625 (0.0239)
		Block	0.1096 (0.0096)	0.0596 (0.0054)	0.6091 (0.0308)
Frank	4	Local	0.0983 (0.0131)	0.01208 (0.0032)	0.2635 (0.0569)
		Block	0.0702 (0.0096)	0.0075 (0.0017)	0.2508 (0.0608)
Gumbel	**8.3**	Local	**0.6283 (0.0086)**	**0.6975 (0.0015)**	**0.9145 (0.0009)**
		Block	**0.6223 (0.0058)**	**0.6971 (0.0012)**	**0.9143 (0.0007)**
Gumbel	1.25	Local	0.0720 (0.0075)	0.0240 (0.0041)	0.5377 (0.0568)
		Block	0.0721 (0.0085)	0.0336 (0.0042)	0.6688 (0.0421)

Lemma 3. *Let us assume that c belongs to $L_\infty([0,1]^d)$ and let $(j, N) \in \mathbb{N}^2$. For any $1 \le p \le q \le d$, for any subsets $\mathcal{S}_p$ and $\mathcal{S}_{q-p}$ of $\{1, \dots, d\}$ with cardinality equal to p and $q - p$ having no common component, let us put for $i = 1 \dots, n$,*

$$Z_i(\mathcal{S}_p, \mathcal{S}_{q-p}) = \prod_{m \in \mathcal{S}_p} \psi_{j,k_m}(F_m(X_i^m)) \prod_{m' \in \mathcal{S}_{q-p}} (\psi^{(1)})_{j,k_{m'}}(F_{m'}(X_i^{m'})), \tag{A.3}$$

with the following notation $\psi_{j,k}^{(1)}(.) = 2^{j/2}\psi'(2^j. - k)$. For any $\mu \ge 2K_3 2^{-jq/2}$, we have

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N} Z_i(\mathcal{S}_p, \mathcal{S}_{q-p})\right| > \mu\right) \le 2\exp\left(-K_2 N\left(\mu^2 \wedge \mu 2^{1-jq/2}\right)\right)$$

where K_2, K_3 are constants such that

$$K_3 \ge (L+1)^q \|c\|_\infty \|\psi\|_\infty^p \|\psi'\|_\infty^{q-p}, \qquad K_2 \le \frac{1}{8}\|\psi\|_\infty^{-p}\|\psi'\|_\infty^{p-q}\left(K_3^{-1} \vee 6\right).$$

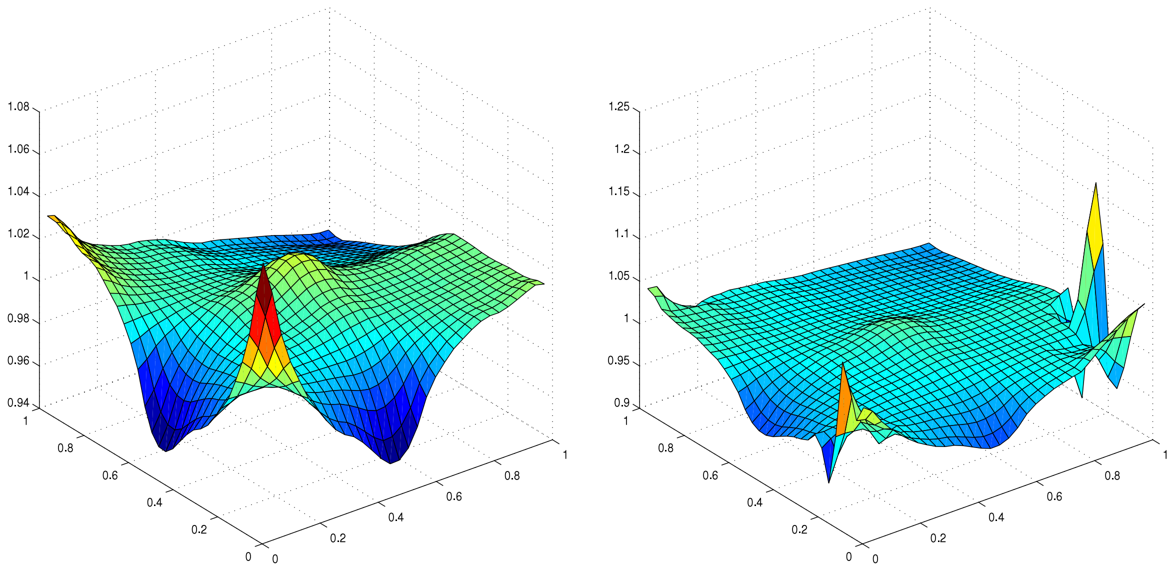


Fig. A.2. Brent/Cac: Block Thresholding Method (left) and Local Thresholding Method (right).

Table A.5
Brent/Cac: distances between the benchmarks and the parametric families.

		$\hat{\theta}_1$	E_1	$\hat{\theta}_2$	E_2	$\hat{\theta}_\infty$	E_∞
Gaussian	Block	−0.01	0.0068	−0.01	0.0001	−0.01	0.0449
Gaussian	Local	−0.01	0.0080	−0.01	0.0002	0.01	0.0847
Student	Block	(−0.11, 91)	0.0640	(−0.11, 91)	0.0103	(−0.11, 91)	0.6639
Student	Local	(0.07, 40)	0.0226	(0.07, 40)	0.0010	(0.02, 100)	0.1279
Clayton	Block	0.01	0.0125	0.01	0.0002	0.01	0.0395
Clayton	Local	0.01	0.0135	0.01	0.0004	0.01	0.0942
Frank	Block	0.01	0.0103	0.01	0.0002	0.01	0.0467
Frank	Local	0.01	0.0115	0.01	0.0003	0.07	0.0825
Gumbel	Block	1.00	0.0093	1.00	0.0002	1.00	0.0462
Gumbel	Local	1.00	0.0106	1.00	0.0003	1.00	0.0963
All	Block	−0.01	Gaussian 0.68%	−0.01	Gaussian 0.01%	0.01	Clayton 4.28%
All	Local	−0.01	Gaussian 0.79%	−0.01	Gaussian 0.02%	0.07	Frank 7.98%

Lemma 3 is a direct application of the Bernstein Inequality with

$$\left|EZ_i(\mathcal{S}_p, \mathcal{S}_{q-p})\right| = \left|\int_{[0,1]^d} \prod_{m\in\mathcal{S}_p} \psi_{j,k_m}(u_m) \prod_{m'\in\mathcal{S}_{q-p}} \psi^{(1)}_{j,k_{m'}}(u_{m'})c(u_1,\ldots,u_d)du_1\times\cdots\times du_d\right|$$
$$\leq K_3 2^{-jq/2}$$

and in the same way,

$$Var(Z_i(\mathcal{S}_p, \mathcal{S}_{q-p})) \leq (L+1)^q \|c\|_\infty \|\psi\|_\infty^{2p} \|\psi'\|_\infty^{2(q-p)}$$

and

$$\left|Z_i(\mathcal{S}_p, \mathcal{S}_{q-p})\right| \leq \|\psi\|_\infty^p \|\psi'\|_\infty^{q-p} 2^{jq/2}.$$

A.2. Proof of Proposition 2

By Equality (A.1), we have for any $\lambda > 0$

$$\mathbb{P}\left(|\widehat{c^{\epsilon}_{j,k}} - \widetilde{c^{\epsilon}_{j,k}}| > \lambda\right) \leq \sum_{m=1}^{d} C_d^m L_m$$

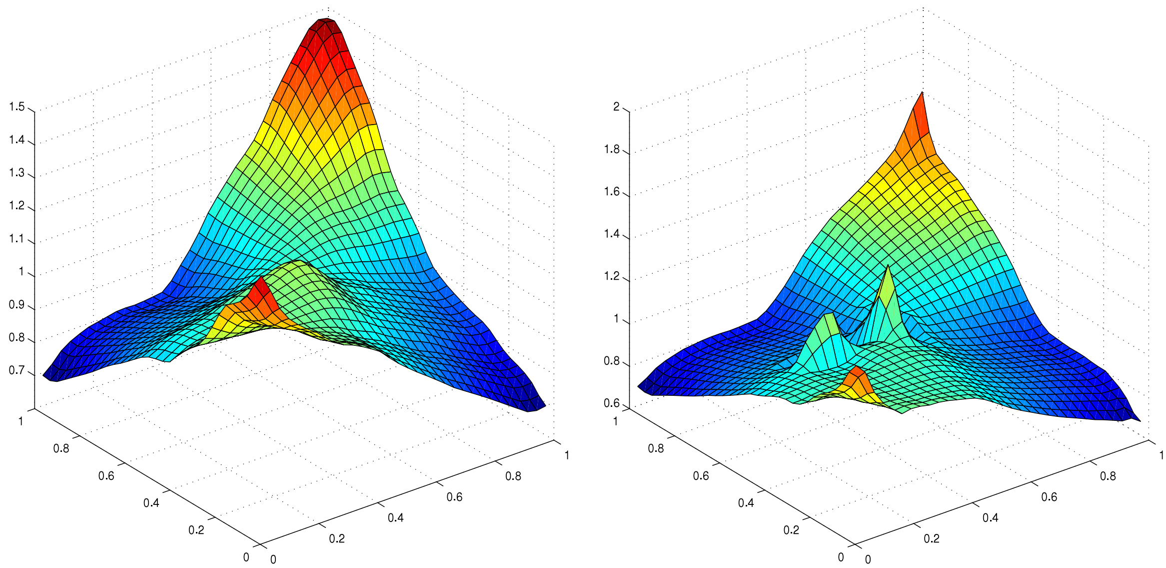


Fig. A.3. Brent/ExonMobil: Block Thresholding Method (left) and Local Thresholding Method (right).

Table A.6
Brent/ExonMobil: distances between the benchmarks and the parametric families.

		$\hat{\theta}_1$	E_1	$\hat{\theta}_2$	E_2	$\hat{\theta}_\infty$	E_∞
Gaussian	Block	0.15	0.0396	0.14	0.0030	0.10	0.1337
Gaussian	Local	0.14	0.0492	0.13	0.0041	0.10	0.1437
Student	Block	(0.14, 37)	0.0376	(0.13, 81)	0.0030	(0.08, 61)	0.1329
Student	Local	(0.14, 95)	0.0491	(0.13, 95)	0.0041	(0.09, 80)	0.1411
Clayton	Block	0.15	0.0706	0.12	0.0099	0.05	0.1879
Clayton	Local	0.14	0.0799	0.11	0.0109	0.05	0.1967
Frank	Block	0.76	0.0301	0.83	0.0017	0.85	0.0957
Frank	Local	0.75	0.0393	0.80	0.0027	0.54	0.1355
Gumbel	Block	1.10	0.0436	1.07	0.0069	1.02	0.2309
Gumbel	Local	1.10	0.0529	1.06	0.0076	1.02	0.2298
All	Block	0.76	Frank 3.01%	0.83	Frank 0.17%	0.85	Frank 6.61%
All	Local	0.75	Frank 3.93%	0.80	Frank 0.27%	0.54	Frank 10.64%

where

$$L_m = \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} T_{m,j}(X_i)\right| > \frac{\lambda}{2^d - 1}\right).$$

Using a Taylor expansion, the following inequality holds for any $i \in \{1, \ldots, n\}$ and any $m' \in \{1, \ldots, d\}$

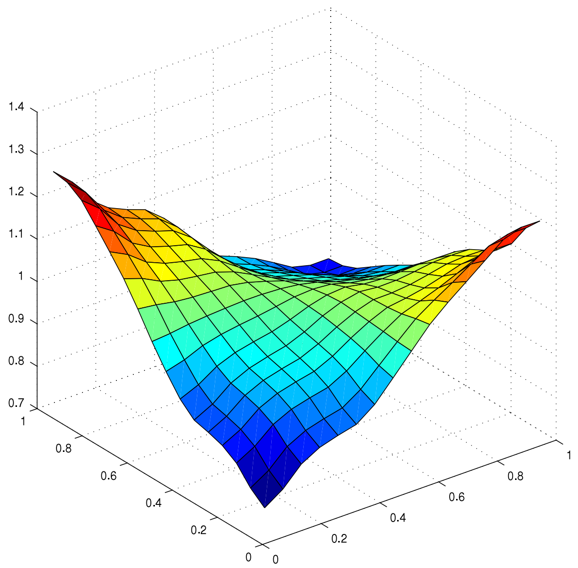
$$|\xi_j(X_i^{m'})| \leq 2^j |\Delta(X_i^{m'})| (\psi^{(1)})_{j,k_{m'}}(F_{m'}(X_i^{m'})) + 2^{\frac{3j}{2}-1} |\Delta(X_i^{m'})|^2 \|\psi'\|_\infty. \tag{A.4}$$

This implies that, for an associated $\mathcal{S}_{d-m}$,

$$|T_{m,j}(X_i)| \leq \|\psi'\|_\infty^m \sum_{\substack{m'=0 \\ \mathcal{S}_{m-m'} \cap \mathcal{S}_{d-m'} = \emptyset}}^{m} 2^{j(m+m'/2)} \left(\max_{m'=1,\ldots,m} |\Delta(X_i^{m'})|\right)^{m+m'} |Z_i(\mathcal{S}_{d-m}, \mathcal{S}_{m-m'})|.$$

For $m = 1, \ldots d$, let us introduce the events

$$\mathcal{D}_{0,m} = \left\{\max_{1 \leq i \leq n} |\Delta(X_i^m)| \leq \sqrt{\frac{\delta \log(n)}{2n}}\right\}, \qquad \mathcal{D}_{1,m} = \left\{N_j(m) \leq n_j = (L+3)n2^{-j}\right\}$$



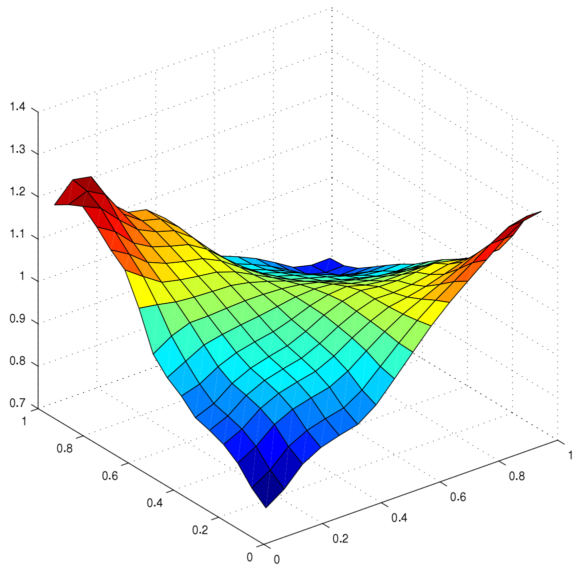


Fig. A.4. DowJones/Oncedor: Block Thresholding Method (left) and Local Thresholding Method (right).

Table A.7
DowJones/Oncedor: distances between the benchmarks and the parametric families.

		$\hat{\theta}_1$	E_1	$\hat{\theta}_2$	E_2	$\hat{\theta}_\infty$	E_∞
Gaussian	Block	−0.11	0.0233	−0.10	0.0010	−0.07	0.0765
Gaussian	Local	−0.11	0.0243	−0.10	0.0011	−0.07	0.0765
Student	Block	(−0.11, 61)	0.0233	(−0.10, 61)	0.0011	(−0.06, 61)	0.0859
Student	Local	(−0.11, 80)	0.0239	(−0.10, 80)	0.0011	(−0.06, 63)	0.0859
Clayton	Block	0.01	0.0801	0.01	0.0104	0.01	0.2924
Clayton	Local	0.01	0.0805	0.01	0.0105	0.01	0.2924
Frank	Block	−0.57	0.0148	−0.56	0.0003	−0.50	0.0456
Frank	Local	−0.58	0.0155	−0.57	0.0004	−0.48	0.0433
Gumbel	Block	1.00	0.0755	1.00	0.0090	1.00	0.2316
Gumbel	Local	1.00	0.0760	1.00	0.0092	1.00	0.2316
All	Block	−0.57	Frank 1.48%	−0.56	Frank 0.03%	−0.50	Frank 3.69%
All	Local	−0.58	Frank 1.54%	−0.57	Frank 0.03%	−0.48	Frank 3.53%

and,

$$\mathcal{D}_0 = \bigcap_{m=1}^{d} \mathcal{D}_{0,m}, \qquad \mathcal{D}_1 = \bigcap_{m=1}^{d} \mathcal{D}_{1,m}.$$

It follows that for any $\mathcal{S}_p$ and any $\mathcal{S}_{q-p}$

$$\begin{aligned} L_m &\le \mathbb{P}\left(\left(\left|\frac{1}{n}\sum_{i=1}^{n} T_{m,j}(X_i)\right| > \frac{\lambda}{2^d-1}\right) \cap \mathcal{D}_0 \cap \mathcal{D}_1\right) + \mathbb{P}\left(\mathcal{D}_0^c\right) + \mathbb{P}\left(\mathcal{D}_1^c\right) \\ &\le \sum_{m'=0}^{m} \mathbb{P}\left(\left|\frac{1}{n_j}\sum_{i=1}^{n_j} Z_i(\mathcal{S}_{d-m}, \mathcal{S}_{m-m'})\right| > \mu\right) + \mathbb{P}\left(\mathcal{D}_0^c\right) + \mathbb{P}\left(\mathcal{D}_1^c\right) \end{aligned}$$

where

$$\mu = 2^{-j(m+m'/2)} \left(\frac{2n}{\delta \log(n)}\right)^{\frac{m+m'}{2}} \frac{2^j \|\psi'\|_\infty^{-m} (L+3)^{-1} \lambda}{(2^d-1)(m+1) C_m^{\lfloor m/2 \rfloor}}.$$

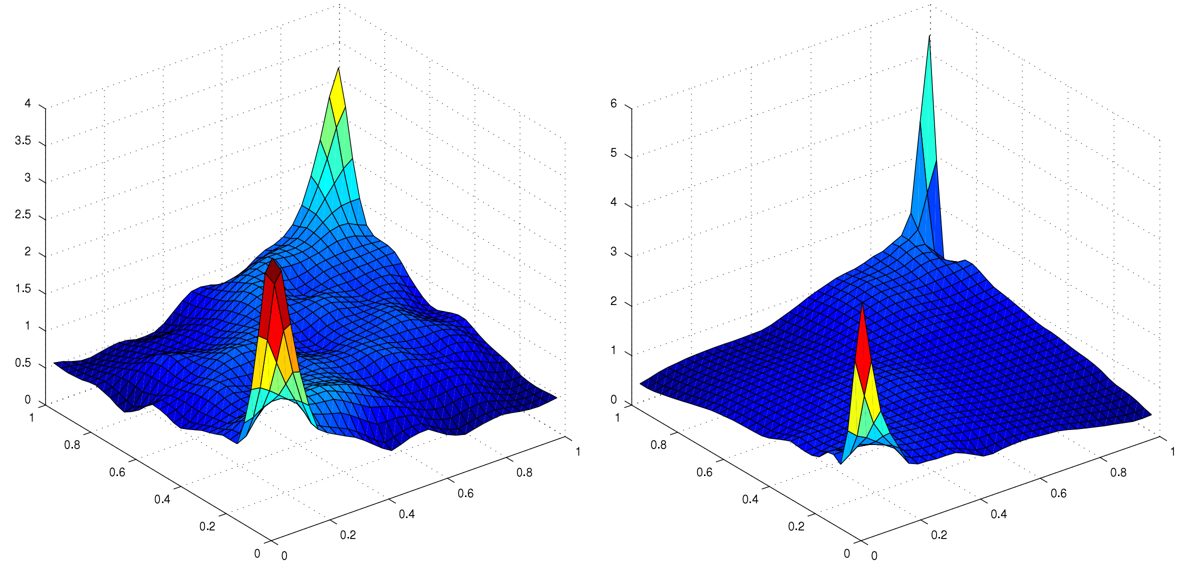


Fig. A.5. DowJones/Ftse100uk: Block Thresholding Method (left) and Local Thresholding Method (right).

Table A.8
DowJones/Ftse100uk: distances between the benchmarks and the parametric families.

		$\hat{\theta}_1$	E_1	$\hat{\theta}_2$	E_2	$\hat{\theta}_\infty$	E_∞
Gaussian	Block	0.30	0.0976	0.33	0.0202	0.20	0.4191
Gaussian	Local	0.26	0.0699	0.32	0.0234	0.11	0.2785
Student	Block	(0.28, 8)	0.0755	(0.29, 8)	0.0127	(0.18, 11)	0.3027
Student	Local	(0.17, 12)	0.0846	(0.17, 6)	0.0265	(0.12,20)	0.3748
Clayton	Block	0.40	0.1064	0.36	0.0318	0.26	0.4565
Clayton	Local	0.31	0.0978	0.33	0.0401	0.11	0.3465
Frank	Block	1.58	0.1094	1.88	0.0333	0.57	0.4366
Frank	Local	1.38	0.0687	1.73	0.0401	0.79	0.2762
Gumbel	Block	1.19	0.1081	1.17	0.0414	1.09	0.4427
Gumbel	Local	1.18	0.0782	1.18	0.0282	1.06	0.3866
All	Block	(0.28, 8)	Student 7.55%	(0.29, 8)	Student 1.15%	(0.18, 11)	Student 10.62%
All	Local	1.38	Frank 6.86%	0.32	Gaussian 2.12%	0.79	Frank 19.56%

Table A.9
Estimation of the parameter θ in a parametric family. For each line, we have generated a sample of size $n = 2000$ of copula specified in the first column and the parameter specified in the second one.This parameter is estimated by minimizing the empirical L_1, L_2, L_∞ errors between the parametric copulas and the non-parametric estimate. Each column specifies the estimated parameter and its standard error.

Copula	Parameter	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_\infty$
FGM	1	0.9240 (0.0609)	0.9029 (0.0604)	0.8690 (0.0794)
Normal	0	−0.0008 (0.0249)	−0.0011 (0.0224)	−0.0005 (0.0207)
Normal	0.5	0.4764 (0.0191)	0.4864 (0.0179)	0.4680 (0.0299)
Normal	0.9	0.8645 (0.0055)	0.8607 (0.0059)	0.8552 (0.0296)
Student	0.5	0.4988 (0.0438)	0.5066 (0.0292)	0.3612 (0.1011)
	1	1.9100 (0.2862)	1.9900 (0.0995)	2.0200 (0.1400)
Clayton	0.8	0.7038 (0.0467)	0.7352 (0.0503)	0.5597 (0.1150)
Clayton	6	3.8244 (0.1641)	2.1972 (0.0345)	2.0040 (0.0000)
Frank	4	4.0000 (0.0000)	4.0000 (0.0000)	4.0000 (0.0000)
Gumbel	8.3	5.0648 (0.1161)	5.0040 (0.0000)	5.0040 (0.0000)
Gumbel	1.25	1.2257 (0.0271)	1.2262 (0.0307)	0.1237 (0.0010)

Fix $\kappa > 0$ and take $\lambda = \sqrt{\frac{\kappa \log(n)}{n}}$. Using Lemmas 1 and 2, we get

$$\mathbb{P}(\mathcal{D}_0^c) \vee \mathbb{P}(\mathcal{D}_1^c) \leq d(K_0 \vee K_1) n^{-\delta} \tag{A.5}$$

for $2^j \leq (1/3)\, (2n/(\delta \log(n)))^{1/2}$. Since $\mu \geq 2K_3\, 2^{-j(d-m')/2}$, we apply Lemma 3 and we obtain

$$L_m \leq 2 \sum_{m'=0}^{m} \exp\left[-K_2 2^{-j} n \left(\mu^2 \wedge \mu 2^{1-j(d-m')/2}\right)\right] + d(K_0 \vee K_1) n^{-\delta} \leq K n^{-\delta}$$

for

$$\mu \geq \left(\frac{\delta}{K_2} \frac{2^j \log(n)}{n}\right)^{1/2} \vee \left(\frac{\delta}{K_2} \frac{2^{j(2+d-m')/2} \log(n)}{2n}\right). \tag{A.6}$$

Let us restrict ourselves to the case:

$$2^j \leq \left(\frac{n}{\log n}\right)^{1/d}.$$

Assuming that n and κ are large enough, the inequality (A.6) for μ is satisfied if, for any $m' = 0, \ldots m$

$$d \geq \frac{2m + m' - 1}{m + m'} \vee \frac{2m + d}{m + m' + 1}.$$

Since this condition is always satisfied by $d \geq 2$, we obtain the announced result.

References

[1] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.
[2] G. Kerkyacharian, D. Picard, Regression in random design and warped wavelets, Bernoulli 10 (2004) 1053–1105.
[3] D.L. Donoho, I. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, J. Amer. Assoc 90 (1995) 1200–1224.
[4] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard, Density estimation by wavelet thresholding, Ann. Statis. 24 (1996) 508–539.
[5] G. Kerkyacharian, D. Picard, K. Tribouley, L^p adaptive density estimation, Bernoulli 2 (1996) 229–247.
[6] C. Genest, E. Masiello, K. Tribouley, Estimating copula densities through wavelets, Insurance Math. Econom. 44 (2009) 170–181.
[7] A. Cohen, I. Daubechies, P. Vial, Wavelets on the interval and fast wavelet transforms, Appl. Comput. Harmon. Anal. 1 (1993) 54–81.
[8] Y. Meyer, Ondelettes et Opérateurs, Hermann, Paris, 1990.
[9] K. Tribouley, Practical estimation of multivariate densities using wavelet methods, Statist. Neerlandica 49 (1995) 41–62.
[10] G. Kerkyacharian, D. Picard, Density estimation in Besov spaces, Statist. Probab. Lett. 13 (1992) 15–24.
[11] G. Kerkyacharian, D. Picard, Thresholding algorithms, maxisets and well concentrated bases, Test 9 (2001) 283–344.
[12] I.A. Ibragimov, R.Z. Khasminski, Statistical Estimation, Springer-Verlag, New-York, 1981. Asymptotic theory, translated from the Russian by Samuel Kotz.
[13] A. Cohen, R. De Vore, G. Kerkyacharian, D. Picard, Maximal spaces with given rate of convergence for thresholding algorithms, Appl. Comput. Harmon. Anal. 11 (2001) 167–191.
[14] F. Autin, D. Picard, V. Rivoirard, Maxiset approach for Bayesian nonparametric estimation, Math. Methods Statist. 15 (4) (2006) 349–373.
[15] F. Autin, Maxiset for density estimation on $\mathbb{R}$, Math. Methods of Statist. 2 (2006) 123–145.
[16] F. Autin, E. Le Pennec, K. Tribouley, Thresholding methods to estimate the copula density, Preprint on the web (http://www.cmi.univ-mrs.fr/~autin/DONNEES/COPULAS), 2008.
[17] G. Gayraud, K. Tribouley, Good-fit-of test for the copula density, Submitted paper, 2008.