Copulas checker-type approximations: application to quantiles estimation of aggregated variables

Andrés Cuberos, Esterina Masiello, Véronique Maume-Deschamps

COPULAS CHECKER-TYPE APPROXIMATIONS: APPLICATION TO QUANTILES ESTIMATION OF AGGREGATED VARIABLES

A. CUBEROS, E. MASIELLO, AND V. MAUME-DESCHAMPS

Abstract. Estimating quantiles of aggregated variables (mainly sums or weighted sums) is crucial in risk management for many application fields such as finance, insurance, environment... This question has been widely treated but new efficient methods are always welcome; especially if they apply in (relatively) high dimension. We propose an estimation procedure based on copula's approximations (*checkerboard copula, checkmin copula*). It allows to get rather good estimations from a (quite) small sample of the multivariate law and a full knowledge of the marginal laws. Estimations may be improved by including in the approximated copula some additional information (on the law of a sub-vector or on extreme probabilities). Our approach is illustrated by numerical examples.

1. Introduction

Consider a random vector $\mathbf{X} = (X_1, \dots, X_d)$ and a measurable function $\Psi : \mathbb{R}^d \to \mathbb{R}^+$, called the aggregation function. In the context of quantitative risk management $\mathbf{X}$ is known as a risk vector and generally represents the profit-losses of a portfolio at a given future date. $\Psi(\mathbf{X})$, the aggregated risk, represents its total future position. The main examples of aggregation functions are: the sum, max, weighted sums or a slightly more complex function that may include stop-loss reinsurance type function on each of the marginal distributions. In this paper, we will be essentially concerned with $\Psi = \sum$, which is the most commonly studied aggregation function. We are interested here in the estimation of p-quantiles, $0 < p < 1$, of $\Psi(\mathbf{X})$: $Q_p(\Psi(\mathbf{X})) = \inf\{x \in \mathbb{R}, F_\Psi(x) \geq p\}$, where we denote by $F_\Psi(x) = \mathbb{P}(\Psi(\mathbf{X}) \leq x)$ the distribution function of the aggregated random variable. In financial or insurance contexts, the Q_p's are called *Value at Risk* and denoted VaR_p. We will assume that the distributions $F_1, \dots, F_d$ of the marginal variables $X_1, \dots, X_d$ are known and that some information on the dependence between them is given. Usually this information is available via some observations of the joint distribution and also via expert opinion.

In practice, neither the marginals nor the dependence of the risk vector $\mathbf{X}$ are known. However, in many cases, the information available on the marginal distributions is much more important than the one on the dependence structure. For example, when some observations of the vector $\mathbf{X}$ are available, inferences one can do on the marginal distributions give better

Key words and phrases. risk aggregation, empirical copulas, checkerboard copula, checkmin copula, quantile estimation.

results than inferences one can do on the multivariate distribution. Also, samples available for marginal laws may be much larger than those available for the joint distribution. Moreover, on each marginal risk, some extra information, as for example expert opinion or prior information, may be available. These situations arise e.g. for environmental data or in insurance contexts. So, even if the assumption of the knowledge of marginal distributions may seem not realistic, by simplification we shall assume here that the marginal distributions F_i are known. However it may be well the case that these distributions are actually estimations based on data or extra marginal information.

When the marginals are known but the dependence is unknown, the re-arrangement algorithm (introduced in special cases in [25] and [24]) allows to obtain bounds on the distribution of $\Psi(\mathbf{X})$ ([23]) for $d \geq 30$. By improving the re-arrangement algorithm, bounds on the VaR are obtained in ([12]) in high dimensional ($d \geq 1000$) inhomogeneous portfolio. These bounds are usually too wide to be useful in practice. Cases in which some kind of dependence information is available lead to narrower bounds ([1, 2]) for the risk measure at hand. Bounds are also derived in ([4]) for dependence structures described by different copula models. A general mathematical framework which interpolates between marginal knowledge and full knowledge of the distribution function of $\mathbf{X}$ is considered in ([11]).

In this paper, we propose to use the check-min-erboard copula (as introduced in [17] in dimension 2 and developed in [20] in higher dimension) to merge the information given by a small sample of the distribution of $\mathbf{X}$ with the known marginal distributions. Related kinds of copulas approximation appeared already in [7] and have been studied in [15, 21, 16] for discrete variables. Durante et al. [10] presented *patchwork copulas* which give a general framework for studying piecewise approximations of copulas. We consider generalized versions of the checkerboard copula: the partition considered needs not to coïncide with the one given by the sample. Moreover, we introduce the checkerboard copula with information on the tail and with information on a sub-vector, to take into account some additional informations which may improve the quantile estimation (see Section 3). We introduce empirical versions of the above approximations that are proper copulas and allow to obtain efficient estimations of aggregated quantiles.

In Section 2, we recall basic definitions on copulas and we present the check-min-erboard copulas and their empirical version. We also prove some convergence results. In Section 3, we show how additional information (on the tail or on the law of a sub-vector) may be pushed into the check-min-erboard copula. A simulation study is presented in Section 4. Conclusions are provided in Section 5.

2. Copulas approximations

Let $\mathbf{F}$ be the distribution function of $\mathbf{X} = (X_1, \ldots, X_d)$, where $X_1, \ldots, X_d$ are assumed to be random variables living on an atomless probability space.

By Sklar's Theorem, there exists a copula distribution C on $[0,1]^d$ such that

$$\mathbf{F}(x_1,\ldots,x_d) = C(F_1(x_1),\ldots,F_d(x_d)),$$

where the F_i's are the distribution functions of the X_i's.
When the marginal random variables X_i are absolutely continuous this copula C is unique. We will assume that the marginals of $\mathbf{X}$ are absolutely continuous. Remark that if G is a distribution function on $[0,1]^d$, it is a copula if and only if the marginal distributions are uniform on $[0,1]$, that is, $G(x) = x_k$ for any $x \in [0,1]^d$ such that $x_i = 1$, $i \neq k$.
The aggregation function $\Psi : \mathbb{R}^d \to \mathbb{R}^+$ is considered to be the sum: $\Psi(\mathbf{X}) = \sum_{i=1}^{d} X_i$. As above, $F_\Psi(x) = \mathbb{P}(\Psi(\mathbf{X}) \leq x)$ will denote the distribution function of the aggregated random variable.
The check-min-erboard copula as introduced in [17] in dimension 2 and developed in [20] in higher dimension, is a flexible tool to approximate copulas. These copula's approximations are particular cases of patchwork copulas presented in [10]. In this section, we present the check-min-erboard copulas and their empirical version.

2.1. **Check-min-erboard Copulas.** As above, let $\mathbf{F}$ denote the cumulative distribution function (c.d.f.) of $\mathbf{X}$, C its copula function and F_i the c.d.f of X_i, $i = 1,\ldots,d$.
Let μ_C be the probability measure associated to C, i.e such that:

$$\mu_C\left(\prod_{i=1}^{d}[0,u_i]\right) = C(u_1,\ldots,u_d)$$

for any $u = (u_1,\ldots,u_d) \in [0,1]^d$.
By a μ-decomposition of a set $A \subset \mathbb{R}^d$ we mean a finite family of measurable sets $\{A_i \subset A\}$ such that

(1) $\mu(A_i \cap A_j) = 0$ whenever $i \neq j$
(2) $\sum_i \mu(A_i) = \mu(A)$.

Definition 1. A measure μ^* is a checkerboard approximation for a copula C if there exists a λ-decomposition $\mathcal{A} = \{(a_i, b_i)\}$ of I^d, the d-dimensional unit cube, made out of d-intervals such that for all i,

(1) μ^* is uniform on (a_i, b_i);
(2) $\mu^*(A) = \mu_C(A)$ for any $A \in \mathcal{A}$.

For $m \in \mathbb{N}$, let us consider the regular λ-decomposition of the unite cube $[0,1]^d$ denoted as $\mathcal{I}_m$ and consisting of m^d d-cubes with side length $1/m$:

$$I_{i,m} = \prod_{j=1}^{d}\left[\frac{i_j - 1}{m}, \frac{i_j}{m}\right],\ i = (i_1,\ldots,i_d),\ i_j \in \{1,\ldots,m\}.$$

μ_m^* is the checkerboard approximation associated to the regular decomposition $\mathcal{I}_m$.

We shall denote by C_m^* the checkerboard copula associated to the measure μ_m^*. The definition of the checkerboard copula may then be rewritten as:

$$C_m^*(x) = \sum_i m^d \mu(I_{i,m}) \lambda([0,x] \cap I_{i,m})$$

where $[0,x] = \prod_{i=1}^{d} [0,x_i]$, for $x = (x_1, \ldots, x_d) \in [0,1]^d$ and λ is the d-dimensional Lebesgue measure. From a probabilistic point of view, sampling with respect to C_m^* means choosing $i \in \{1, \ldots, m\}^d$ with probability $\mu(I_{i,m})$ and then sampling uniformly in $I_{i,m}$. This leads us to consider also the checkmin copula: instead of sampling uniformly in $I_{i,m}$, one may sample in $I_{i,m}$ with respect to the comonotonic copula. The checkmin copula is then given by:

$$C_m^\dagger(x) = \sum_i m\mu(I_{i,m}) \min\left(x_j - \frac{i_j - 1}{m}, \frac{1}{m}\right).$$

In what follows, C_m^o denotes either C_m^* or $C_m^\dagger$ and is called check-min-erboard copula.

In [20], it is proved that C_m^o is a copula that approximates C. The following proposition gives a more precise bound on the approximation of C by C_m^o by a factor 2, than the one presented in dimension 2 in [17], page 613, or than the result obtained in [10].

Proposition 2.1. *Let C_m^o be either the checkerboard copula or the checkmin copula defined above. We have:*

$$\sup_{x \in [0,1]^d} |C_m^o(x) - C(x)| \leq \frac{d}{2m}.$$

Proof. This is clear that for any $x \in [0,1]^d$ with $x = \frac{i}{m}$, $i \in \{1, \ldots, m\}^d$, $C_m^o(x) = C(x)$. We present the computations for C_m^*, the same computations give the result for $C_m^\dagger$.
For $a \in \{1, \ldots, m\}$ and $k \in \{1, \ldots, d\}$, we denote by B_a^{k+} and B_a^{k-} the (half)-strips:

$$B_a^{k+} = \left\{ x \in [0,1]^d,\ \frac{a}{m} - \frac{1}{2m} < x_k \leq \frac{a}{m} \right\} \text{ and}$$

$$B_a^{k-} = \left\{ x \in [0,1]^d,\ \frac{a}{m} < x_k \leq \frac{i_k}{m} - \frac{1}{a} \right\}.$$

If $x \in I_{i,m}$ with $i = (i_1, \ldots, i_d)$ then,

$$\begin{aligned}
|C_m^*(x) - C(x)| &\leq \sum_{k=1}^{d} |\mu_m^*(B_{i_k}^{k-}) - \lambda(B_{i_k}^{k-})| \mathbf{1}_{B_{i_k}^{k-}}(x) + \\
&\quad \sum_{k=1}^{d} |\mu_m^*(B_{i_k}^{k+}) - \lambda(B_{i_k}^{k+})| \mathbf{1}_{B_{i_k}^{k+}}(x) \\
&\leq \sum_{k=1}^{d} \min(\mu_m^*(B_{i_k}^{k-}), \lambda(B_{i_k}^{k-})) \mathbf{1}_{B_{i_k}^{k-}}(x) + \\
&\quad \sum_{k=1}^{d} \min(\mu_m^*(B_{i_k}^{k+}), \lambda(B_{i_k}^{k+})) \mathbf{1}_{B_{i_k}^{k+}}(x) \\
&= \frac{d}{2m}
\end{aligned}$$

since μ_m^* and λ are both associated to a copula,

$$\mu_m^*(B_{i_k}^{k-}) = \lambda(B_{i_k}^{k-}) = \mu_m^*(B_{i_k}^{k+}) = \lambda(B_{i_k}^{k+}) = \frac{1}{2m}.$$

The announced result follows. □

In what follows, we will define an empirical version of the check-min-erboard copula defined above, by using the empirical copula.

2.2. **Empirical check-min-erboard copulas.** The empirical copula, introduced by Deheuvels ([8]), may be used to estimate non parametrically the copula.

Definition 2. Let $\mathbf{X}^1, \ldots, \mathbf{X}^n$ be n independent copies of $\mathbf{X}$. Each of them writes $\mathbf{X}^j = (X_1^j, \ldots, X_d^j)$. Let $R_i^1, \ldots, R_i^n$, $i = 1, \ldots, d$ be their marginals ranks, i.e.,

$$R_i^j = \sum_{k=1}^{n} 1\{X_i^{(j)} \geq X_i^{(k)}\},\ i = 1, \ldots, d,\ j = 1, \ldots, n$$

where $X_i^{(1)} < \cdots < X_i^{(n)}$ are the order statistics associated to the ith coordinate sample $X_i^1, \ldots X_i^n$. The empirical copula $\widehat{C}_n$ of $\mathbf{X}^1, \ldots \mathbf{X}^n$ is defined as

$$\widehat{C}_n(u_1, \ldots, u_d) = \frac{1}{n} \sum_{k=1}^{n} 1 \left\{ \frac{1}{n} R_1^k \leq u_1, \ldots, \frac{1}{n} R_d^k \leq u_d \right\}.$$

It is well known (see [14] e.g.) that the empirical copula may be used to estimate C. Nevertheless, it is not a proper copula as its marginal laws are discrete. We shall use the empirical copula $\widehat{C}_n$ and the empirical probability measure $\widehat{\mu}$ associated to $\widehat{C}_n$, to define an empirical version of the checkerboard copulas introduced above. Note that in [8] an interpolation of the empirical copula in order to get an estimation of C which is a proper copula is also mentioned. As we shall see below, the main advantage of our approach is that it is very easy to simulate a sample from the empirical checkerboard copula.

Definition 3. Let $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ be n independent copies of $\mathbf{X}$. The empirical checkerboard copula (ECBC) $\widehat{C}_m^*$ is defined by

$$\widehat{C}_m^*(x) = \sum_i m^d \widehat{\mu}(I_{i,m}) \lambda([0,x] \cap I_{i,m}).$$

The empirical checkmin copula (ECMC) $\widehat{C}_m^\dagger$ is defined by

$$\widehat{C}_m^\dagger(x) = \sum_i \widehat{\mu}(I_{i,m}) m \min\left(x_j - \frac{i_j - 1}{m}, \frac{1}{m}\right).$$

In what follows, $\widehat{C}_m^o$ denotes either the ECBC or the ECMC and is called empirical check-min-erboard copula.

Remark. If $U^1, \ldots, U^n$ is an i.i.d sample distributed as C, it is known (see e.g. [14]) that the resulting empirical distribution function and $\widehat{C}_n$ defined above coincide. This remark justifies the use of the Donsker property below.

In the next proposition we show that $\widehat{C}_m^o$, defined above, is a copula whenever the integer m from the length size of the partition $\mathcal{I}_m$ divides n, the sample size.

Proposition 2.2. *The empirical chek-min-erboard copula $\widehat{C}_m^o$ defined on the regular partition $\mathcal{I}_m$ and based on an i.i.d sample of size n is a copula, if and only if m divides n.*

Proof. We give the details of the proof for C_m^*. The proof is the same for $C_m^\dagger$.
Suppose that $m \leq n$ and that m divides n. By definition the empirical checkerboard copula is a distribution function, we should simply check that the marginals are uniform. Without losing generality we show only that the projection on the first coordinate of the measure induced by $\widehat{C}_m^*$ is uniform, or equivalently that $\widehat{C}_m^*(x) = x_1$ for any $x \in [0,1]^d$ with $x_j = 1$ for $j \neq 1$.
For $\ell \in \{1, \ldots, m\}$, consider the strip B_ℓ^1:

$$B_\ell^1 = \left\{ x \in [0,1]^d,\ \frac{\ell - 1}{m} < x_1 \leq \frac{\ell}{m} \right\} = \left] \frac{\ell - 1}{m}, \frac{\ell}{m} \right] \times [0,1]^{d-1}.$$

The empirical copula is concentrated on n points of $[0,1]^d$ whose coordinates are of the form $\frac{j}{n}$, $j = 1, \ldots, n$. Moreover, there is exactly one mass on each strip B_j^1, $j = 1, \ldots, n$. So that if $k = n/m$, then the number of masses of $\widehat{C}_n$ on each strip B_ℓ^1, $\ell = 1, \ldots, m$ is exactly k, which means that $\widehat{\mu}(B_\ell^1) = \frac{k}{n} = \frac{1}{m}$. Let $x = (x_1, 1, \ldots, 1)$, $x_1 \in [0,1]$, $x \in B_\ell^1$ with

$\frac{\ell-1}{m} < x_1 \leq \frac{\ell}{m}$.

$$\begin{aligned}
\widehat{C}^*_m(x) &= \sum_{i\in\{1,\ldots,m\}^d} m^d \widehat{\mu}(I_{i,m})\lambda([0,x]\cap I_{i,m}) \\
&= \sum_{j<\ell} \widehat{\mu}(B^1_j) + \sum_{I_{i,m}\subset B^1_\ell} m^d \widehat{\mu}(I_{i,m}) \frac{(x_1 - \frac{\ell-1}{m})}{m^{d-1}} \\
&= \sum_{j<\ell} \widehat{\mu}(B^1_j) + m\left(x_1 - \frac{\ell-1}{m}\right)\widehat{\mu}(B^1_\ell) \\
&= \frac{\ell-1}{m} + \left(x_1 - \frac{\ell-1}{m}\right) = x_1,
\end{aligned}$$

which shows that the first marginal is uniform. On the other direction, it is easy to see that if m does not divide n then the number of masses of $\widehat{C}_n$ on each strip B^1_ℓ, $\ell = 1,\ldots,m$ is not the same and thus the uniform distribution for the margin is lost. □

We have proved that both the ECBC and ECMC are proper copulas provided that m divides n. The following result shows that they go to C at rate $\frac{1}{\sqrt{n}} + \frac{1}{m}$.

Proposition 2.3. *Let m divide n, we have:*

$$\sup_{t\in[0,1]} |\widehat{C}^o_m(t) - C(t)| \leq O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right) + \frac{d}{2\,m}.$$

Proof. Once more, we give the details for C^*_m, the proof is the same for $C^\dagger_m$.

$$|\widehat{C}^*_m(t) - C(t)| \leq |\widehat{C}^*_m(t) - C^*_m(t)| + |C^*_m(t) - C(t)|.$$

From Proposition 2.1, we know that

$$\sup_{t\in[0,1]^d} |C^*_m(t) - C(t)| \leq \frac{d}{2\,m}.$$

Furthermore,

$$|\widehat{C}^*_m(t) - C^*_m(t)| \leq \sum_{i\in\{1,\ldots,m\}^d} m^d\,\lambda([0,t]\cap I_{i,m})|\widehat{\mu}(I_{i,m}) - \mu(I_{i,m})|.$$

Let $\mathcal{I} = \{I_{i,m}, i \in \{1,\ldots,m\}^d, m \in \mathbb{N}\}$. By using Example 2.6.1 in [26], $\mathcal{I}$ is universally Donsker family. So that,

$$\sup_{i\in\{1,\ldots,m\}^d,\, m\in\mathbb{N}} |\widehat{\mu}(I_{i,m}) - \mu(I_{i,m})| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).$$

Hence the result. □

Since $\widehat{C}^o_m(t)$ converges to $C(t)$, it is natural to estimate $\mathbb{P}(\Psi(\mathbf{X}) \leq t) = F_\Psi(t)$ by $\mathbb{P}(\Psi(T^-(U^o_m)) \leq t) = F^o_m(t)$ where $U^o_m \rightsquigarrow \widehat{C}^o_m$ and $T^-(u_1,\ldots,u_d) = (F^{\leftarrow}_1(u_1),\ldots,F^{\leftarrow}_d(u_d))$. The following result is close to that of Mainik [18], where $\mathbb{P}(\Psi(\mathbf{X}) \leq t)$ is estimated by using the empirical copula and the empirical distribution functions of the marginal laws. For $t \in \mathbb{R}$, let $A_t = \{x \in \mathbb{R}^d,\ \sum_{i=1}^d x_i \leq t\}$. We shall assume the following regularity condition on C. This condition is satisfied for copulas with bounded density, as well as

Clayton copulas in dimension 2 and Gaussian copula in dimension 2, with $\rho > 0$ (see [18] for a more complete discussion on this condition). Simple symmetry considerations show that it will also be satisfied for survival Clayton copulas.

Assumption 2.1. For $t \in \mathbb{R}$, $B_t = \{u \in [0,1]^d,\ T^-(u) \in A_t\}$ and $\partial(B_t)$ its boundary, the regularity of F_j, $j = 1, \dots, d$, implies that $\partial(B_t)$ is a $d-1$ hyper-surface in $[0,1]^d$. For $\delta > 0$, $U_\delta(B_t)$ is the δ neighborhood of $\partial(B_t)$. The regularity condition on C is :

$$\mu(U_\delta(B_t)) = O(\delta). \tag{2.1}$$

Proposition 2.4. *Under Assumption 2.1*

$$\sup_{t\in\mathbb{R}} |F_\Psi(t) - F_m^o(t)| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{1}{m}\right).$$

Proof. We present the proof for the checkerboard copula. We have:

$$\begin{aligned}
|F_\Psi(t) - F_m^*(t)| &= \left| \sum_{i\in\{1,\dots,m\}^d} \widehat{\mu}(I_{i,m}) m^d \lambda(I_{i,m} \cap B_t) - \mu(I_{i,m} \cap B_t) \right| \\
&\leq \underbrace{\sum_{i\in\{1,\dots,m\}^d} |\widehat{\mu}(I_{i,m}) - \mu(I_{i,m})| m^d \lambda(I_{i,m} \cap B_t)}_{(1)} \\
&\quad + \underbrace{\sum_{i\in\{1,\dots,m\}^d} \mu(I_{i,m}) m^d \left|\lambda(I_{i,m} \cap B_t) - \mu(I_{i,m} \cap B_t)\right|}_{(2)}.
\end{aligned}$$

The term (1) is $O_{\mathbb{P}}(\frac{1}{\sqrt{n}})$ as in Proposition 2.3. The term (2) is bounded above by $\mu(U_{d_m}(B_t))$ with $d_m = \frac{\sqrt{d}}{m}$ the diameter of the $I_{i,m}$'s. The result follows. □

In Appendix A we describe how to simulate from a copula $\widehat{C}_m^o$. Figures 1 and 2 show simulations of ECBC and ECMC with different values for m. The size of the simulated sample from the ECBC and ECMC is $N = 10000$.

2.3. **Estimation procedure.** Assume the marginal laws are known and a (quite small) sample of size n of $\mathbf{X}$ is available.

(1) Get the n sample rank.
(2) Simulate a sample of size N from the copula $\widehat{C}_m^o$ for N large (with the procedure described in Appendix A):
$$(u_1^{(1)}, \dots, u_d^{(1)}), \dots, (u_1^{(N)}, \dots, u_d^{(N)})$$
(3) Get a sample of $\Psi(\mathbf{X}^o)$ using the marginals to transform the above sample:
$$\Psi\left(F_1^{\leftarrow}(u_1^{(1)}), \dots, F_d^{\leftarrow}(u_d^{(1)})\right), \dots, \Psi\left(F_1^{\leftarrow}(u_1^{(N)}), \dots, F_d^{\leftarrow}(u_d^{(N)})\right)$$

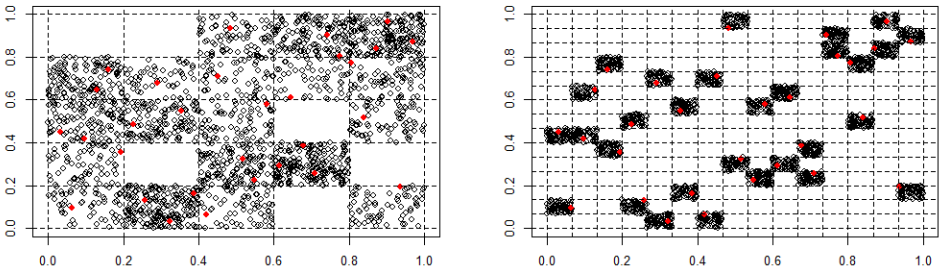

FIGURE 1. Simulation of empirical checkerboard copulas ($N = 10000$) based on the same sample of size $n = 30$, with $m = 5$ (left) and $m = 15$ (right). The red points are the sample rank points.

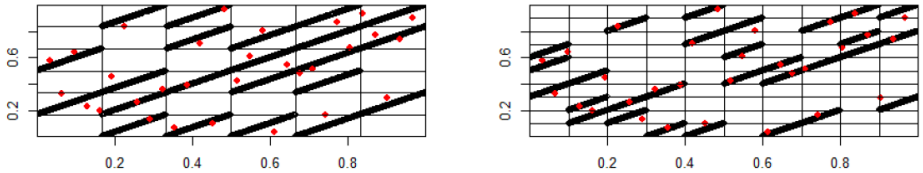

FIGURE 2. Simulation of empirical checkmin copulas ($N = 10000$) based on the same sample of size $n = 30$, with $m = 6$ (left) and $m = 10$ (right). The red points are the sample rank points.

(4) Estimate the distribution function F_Ψ^o of $\Psi(\mathbf{X}^o)$ empirically using the above sample. We will denote $\widehat{F}_\Psi^o$ the empirical distribution function from the sample above.

3. Checkerboard copulas with additional information

We define two kinds of *checkerboard copula with additional information* and then present their empirical version. First of all, we consider the case where the distribution of a sub-vector $\mathbf{X}^J = (X_i)_{i\in J}$, $J \subset \{1,\dots,d\}$, is known, $|J| = k < d$. Denote C^J the copula of $\mathbf{X}^J$. Let μ^J be the probability measure on $[0,1]^k$ associated to C^J. For $i = (i_1,\dots,i_d)$, let $x = (x_1,\dots,x_d) \in [0,1]^d$, $x^J = (x_j)_{j\in J}$, $x^{-J} = (x_j)_{j\notin J}$ and

$$I_{i,m}^J = \left\{ x \in [0,1]^d \ / \ x_j \in \left[\frac{i_j - 1}{m}, \frac{i_j}{m}\right] \ , \ j \in J \right\},$$

$$I_{i,m}^{-J} = \left\{ x \in [0,1]^d \ / \ x_j \in \left[\frac{i_j - 1}{m}, \frac{i_j}{m}\right] \ , \ j \notin J \right\}.$$

The checkerboard copula with information on $\mathbf{X}^J$ is defined below.

Definition 4. Consider the probability measure on $[0,1]^d$ defined by

$$\mu_m^J([0,x]) = \sum_{i\subset\{1,\dots,d\}} \frac{m^{d-k}}{\mu^J(I_{i,m}^J)}\mu(I_{i,m})\lambda([0,x^{-J}]\cap I_{i,m}^{-J})\mu^J([0,x^J]\cap I_{i,m}^J).$$

Let C_m^J, the checkerboard copula with additional information on $\mathbf{X}^J$, be defined by $C_m^J(x) = \mu_m^J([0,x])$.

Proposition 3.1. *C_m^J is a copula, it approximates C:*

$$\sup_{x\in[0,1]^d} |C_m^J(x) - C(x)| \leq \frac{d}{2m}.$$

If $\boldsymbol{X}^J$ and $\boldsymbol{X}^{-J}$ are independent then,

$$\sup_{x\in[0,1]^d} |C_m^J(x) - C(x)| \leq \frac{d-k}{2m}.$$

Proof. The definition of C_m^J ensures that it is a cumulative distribution function on $[0,1]^d$. The fact that C_m^J is a copula then follows from an easy computation to get that $C_m^J(x) = x_k$ whenever $x = (x_j)_{j=1,\dots,d}$, with $x_j = 1$ for $j \neq k$.
The rest of the proof is done as in that of Proposition 2.1. □

We may also add information on the tail and so define the following particular checkerboard copula.

Definition 5. For $p \in]0,1[$, let $E_p = \left(\prod_{i=1}^d [0,p]^d\right)^c$ and $\mathcal{E}_p$ the λ-decomposition of E_p consisting of the hyper rectangles $[a_1,b_1]\times\cdots\times[a_d,b_d]$ where $[a_i,b_i] = [0,p]$ or $[a_i,b_i] = [p,1]$ for all $i = 1,\dots,d$ with at least one of $[a_i,b_i] = [p,1]$. We assume that $\mu_C(A)$ is known for each $A \in \mathcal{E}_p$. Consider the λ-decomposition of the d-cube $[0,p]^d$ given by $\mathcal{J}_m$ consisting of the elements $J_{i,m} = p\cdot I_{m,i}^d$ for d-tuple $i = (i_1,\dots,i_d)$ in $\{0,\frac{1}{m},\dots,\frac{m-1}{m}\}^d$. Define $C_m^{\mathcal{E}_p}$ as the checkerboard copula associated to the λ-decomposition of the unit d-cube $\mathcal{J}_m \cup \mathcal{E}_p$, that is

$$C_m^{\mathcal{E}_p}(x) = \mu_C(E_p^c)\mu_m^*([0,x]/t)\mathbf{1}_{E_p^c}(x) + \sum_{E\in\mathcal{E}_p} \frac{\mu_C(E)}{\lambda(E)}\lambda([0,x]\cap E).$$

This is the checkerboard copula with extra information on the tail.

Empirical versions.

- The ECBC with information on a sub-vector $\mathbf{X}^J$ is defined by

$$\widehat{C}_m^J(x) = \sum_{i\subset\{1,\dots,d\}} \frac{m^{d-k}}{\mu^J(I_{i,m}^J)}\widehat{\mu}(I_{i,m})\lambda([0,x^{-J}]\cap I_{i,m}^{-J})\mu^J([0,x^J]\cap I_{i,m}^J).$$

- The ECBC with information on the tail is defined by:

$$\widehat{C}_m^{\mathcal{E}}(x) = \mu_C(E_p^c)\widehat{C}_m^*(x/t)\mathbf{1}_{E_p^c}(x) + \sum_{E\in\mathcal{E}_p} \frac{\mu_C(E)}{\lambda(E)}\lambda([0,x]\cap E).$$

In Figure 3 we present a simulation of the ECBC with information on the tail for different values of m.

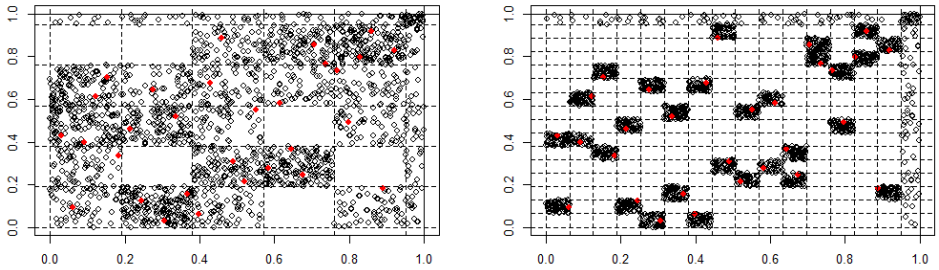

FIGURE 3. Simulation of empirical checkerboard copulas with information on the tail, with $n = 30$, $p = 0.95$ and $m = 5$ (left), $m = 15$ (right). The red points are the sample rank points

Remark. Checkmin copulas with additional information (on a subvector or on the tail) may be defined in the same way as checkerboard with additional information.

4. Numerical Application

In this section we use the estimator of the distribution function F_Ψ, as defined in Section 2.3, in order to estimate the quantiles $Q_p(S)$ for $S = X_1 + \cdots + X_d$ at different confidence levels $0 < p < 1$. We will consider the Pareto-Clayton model, defined in Section 4.1, because, in that case, the exact value of $Q_p(S)$ can be calculated, so that we may compare our simulation results to the exact one. A simulation study will also be presented for another model with gaussian copulas and lognormal marginal laws. In that case, the exact value of $Q_p(S)$ is not known so that we shall use huge sample (of size 10^7) to get a fine approximation of $Q_p(S)$ and then compare it with our estimation using the empirical check-min-erboard copulas based on relatively small samples. In Section 4.2, we shall see that our method is performant in high dimension ($d = 25,\ 50, 100$) with relatively small sample size. In Section 4.3, we provide some examples to show the impact of the information added in the checkerboard copula.

4.1. The Pareto - Clayton model.

We consider $\mathbf{X} = (X_1, \ldots, X_d)$ such that:

$$\mathbb{P}(X_1 > x_1, \ldots, X_d > x_d \,|\, \Lambda = \lambda) = \prod_{i=1}^{d} e^{-\lambda x_i},$$

that is, conditionally to the value of Λ the marginals of $\mathbf{X}$ are independent and exponentially distributed.

If Λ is Gamma distributed, then the X_i's are Pareto distributed with dependence given by a survival Clayton copula.

If Λ is Levy distributed, then the X_i's are Weibull distributed with a Gumbel survival copula.

These models have been studied by Oakes (1989) and Yeh (2007) [22, 27]. In the context of multivariate risk theory, they have been used e.g. in [19] and [6].

In what follows, we consider that $\Lambda \rightsquigarrow \Gamma(\alpha, \beta)$, so that the X_i's are Pareto

(α, β) distributed and the dependence structure is described by a survival Clayton copula with parameter $1/\alpha$. In [9], it is shown that, in this case, S follows the so-called Beta prime distribution:

$$F_S(x) = F_\beta\left(\frac{x}{1+x}\right)$$

where F_β is the distribution function of the Beta$(d\beta, \alpha)$ distribution. The inverse of F_S (or quantile function of S) can also be expressed in terms of the inverse of the Beta distribution

$$F_S^{\leftarrow}(p) = \frac{F_\beta^{\leftarrow}(p)}{1 - F_\beta^{\leftarrow}(p)}.$$

From these results (see also [5]), we may compute $Q_p(S)$.

4.2. **Simulation study in dimension** 25**,** 50 **and** 100**.** In this section we will consider several Pareto-Clayton models and several Gaussian-Lognormal models. Nevertheless, at the beginning we will presentat results in dimension 25, for one Pareto-Clayton model and then for one Gaussian-Lognormal model. One open question is the choice of the partition size m. Following Coeurjolly [3], we propose to keep the median of the estimations obtained for all the m divisor of n for each quantile. The study done in Section 4.2.1 indicates that this choice performs well, it is why in the other subsections, we just present the median result (and not the results for all m). We also perform simulations in dimension 50 and 100 (Section 4.2.2). Tables and boxplots showing the simulation results are postponed to Appendix B and Appendix C.

4.2.1. *Simulations in dimension* 25. We first perform simulations for a Pareto-Clayton model of parameters $\beta = 3$ and $\alpha = 1$. We vary the sample size: $n = 80$, $n = 200$. Recall that for the Pareto-Clayton model, the real value of the quantiles is known. Tables 4 and 5 give the results in terms of Relative Mean Square Error (RMSE) over a run of $s = 100$ different simulations of the initial sample. Let $\widehat{q}^k$ be the estimations of q on the kth simulation run, then

$$\text{RMSE} = \frac{\sqrt{\frac{1}{s}\sum_{k=1}^{s}(\widehat{q}^k - q)^2}}{q}.$$

The size of the checkerboard samples is $N = 10000$. The RMSE is computed for each value of m for the checkerboard (ECBC) and the chekmin (ECMC) approximations. Alternatively, the median choice of m is done for each of the 100 simulations and the RMSE is then computed for the corresponding estimations. The check-min-erboard estimations are compared with parametric ones: we adjust to the samples three copulas (survival Clayton, Gaussian, Clayton), simulate an aggregated sample of size 10000 with the estimated copula and the known marginals to estimate the aggregated quantile. Finally, we put in the table the result obtained with the empirical copula and the known margins.

In Tables 4 and 5, we remark that the checkmin approximation performs significantly better than the cherckerboard one. This may be explained by

the fact that the survival Calyton copula is closer to a comonotonic copula than to the independent copula for the upper tails. Tables 6 and 7 give the results for a Gaussian - lognormal model with correlation $\rho = 0.1$ for all pairs, the parameters of the lognormal margins are different for each of the 25 coordinates. It is expected in that case that the checkerboard copula will give better results than the checkmin. As already mentioned, in the case of a Gaussian - lognormal model, the exact value of the aggregated quantile is unknown. We estimate it with huge samples (of size 10^7) and use this estimate (called *near exact value*) to compute the RMSE's and compare the estimations.

Tables 6 and 7 show that, as expected, in this case, the checkerboard copula gives better results than the checkmin copula. This may be explained by the fact that for a correlation coefficient $\rho = 0.1$ for all pairs, the model is closer to the independence than to the comonotony. We now give results in higher dimension ($d = 50$, $d = 100$).

4.2.2. *Simulations in higher dimensions.* We now perform simulations in dimension 50 and 100 for samples of size $n = 400$. We consider the Pareto-Clayton model with parameters 2 and 1, and a Gaussian-lognormal model with correlations equal to 0.25, 0.5, 0.75 (one third of the coordinates of X for each correlation value). The boxplots corresponding to 100 iterations of the estimation algorithm are presented in Appendix C. In the case of the Pareto-Clayton model, the horizontal line is the real value.

Both Figures 4 and 5 show that the median ECMC estimation is performant for quantile levels 0.995 and 0.999. It is close in mean to the real VaR and much less dispersed than the empirical estimation. Of course the correct parametric model (here the survival Clayton copula) performs better but our approach will be interesting in cases where the parametric models are not well suited to data.

In the case of Gaussian-lognormal model, we consider the estimation with the gaussian copula as a reference. As above, Figures 6 and 7 show that the checkmin estimator is performant also for this model.

4.3. **Adding information (on the tail or on a sub-vector).** We consider a Pareto-Clayton model in dimension 2, with $\beta = 1$ and $\alpha = 2$. The multivariate sample is of size $n = 30$ and for each presented method we performed $N = 1000$ estimations of the p-quantile at different confidence levels. Table 1 presents the mean and the root mean square error of the $N = 1000$ estimations. The estimations were calculated using the ECBC with and without information on the tail on different λ-regular decompositions $\mathcal{I}_m$ for $m = 6, 15, 30$. The information on the tail is introduced on $\mathcal{E}_p$, for $p = 0.95, 0.99$, by giving to each E in $\mathcal{E}_p$ the measure $\mu_C(E)$ where C is the survival Clayton copula with parameter $1/2$. For comparison a direct estimation from the empirical distribution of S is given.

	Quantile 80%	Quantile 90%	Quantile 95%	Quantile 99%	Quantile 99.5%	Quantile 99.9%
Exact value	2.5	4.1	6.4	16.0	23.2	53.4
Empirical	2.5	4.0	6.1	12.2	13.2	14.0
	(26%)	(31%)	(39%)	(72%)	(70%)	(78%)
ECBC (m=6)						
No tail information	2.6	4.4	6.6	14.8	20.8	45.7
	(9%)	(8%)	(6%)	(8%)	(11%)	(15%)
Information on $\mathcal{E}_p$	2.6	4.4	6.4	14.2	22.7	49.5
p=0.99	(9%)	(8%)	(5%)	(11%)	(3%)	(8%)
Information on $\mathcal{E}_p$	2.7	4.1	6.1	15.6	21.8	46.8
p=0.95	(10%)	(5%)	(4%)	(3%)	(6%)	(13%)
ECBC (m=15)						
No tail information	2.5	4.2	6.8	15.5	21.5	46.4
	(12%)	(13%)	(11%)	(9%)	(10%)	(14%)
Information on $\mathcal{E}_p$	2.5	4.3	6.8	14.3	22.7	49.5
p=0.99	(12%)	(12%)	(12%)	(11%)	(3%)	(8%)
Information on $\mathcal{E}_p$	2.6	4.3	6.2	15.6	21.8	46.8
p=0.95	(11%)	(10%)	(4%)	(3%)	(6%)	(13%)
ECBC (m=30)						
No tail information	2.5	4.2	6.6	15.8	22.0	47.0
	(13%)	(15%)	(17%)	(13%)	(12%)	(14%)
Information on $\mathcal{E}_p$	2.5	4.2	6.7	14.3	22.7	49.5
p=0.99	(13%)	(16%)	(16%)	(11%)	(3%)	(8%)
Information on $\mathcal{E}_p$	2.6	4.4	6.2	15.6	21.8	46.8
p=0.95	(13%)	(11%)	(4%)	(3%)	(6%)	(13%)

TABLE 1. The mean and the RMSE in % of the exact value for 1000 estimations of the quantile for a Pareto-Clayton sum in dimension 2.

All the methods we proposed perform much better than the empirical estimation based on the multivariate sample alone. The estimations based on the ECBC with λ-decomposition $\mathcal{I}_m$, perform better when $m = 6$ and $m = 15$ than when $m = 30$. ECBC with $m = 6$ performs slightly better than ECBC with $m = 15$ for the estimation of the quantiles with confidence levels lower than 99.5% and slightly worst on the higher levels. When the information on the tail is introduced on $\mathcal{E}_p$ with $p = 0.95$ the estimation on the quantile with confidence level $> 0,95$ is significantly improved. When it is introduced on $\mathcal{E}_p$ with $p = 0.99$ the estimations improve on the higher confidence levels 99.5% and 99.9%.

Remark 1. It is known (see e.g. Proposition 2 in [12]) that the supremum of the aggregated VaR of level α over distributions $X = (X_1, X_2)$ with X_1 and X_2 having common distribution function F is:

$$2F^{-1}\left(\frac{1+\alpha}{2}\right).$$

For the above example (Pareto margins of parameters 1 and 2), we get the following bounds:

	Quantile 80%	Quantile 90%	Quantile 95%	Quantile 99%	Quantile 99.5%	Quantile 99.9%
Bound	4.32	6.94	10.65	26.284	38.00	87.44

These bounds have to be compared with the real values above. This comparison shows that the use of the bounds is not sharp for quantile estimations, even if they are very interesting to bound the uncertainty risk.

In order to assess the gain that the knowledge of the information on a sub-vector may give to the estimation, we performed here the following simulation study. Let $\mathbf{X} = (X_1, X_2, X_3)$ be the model where $X_1 = X_2 = Y/2$, and $X_3 \sim Y$ where Y is Pareto distributed with $\alpha = 2$. We assume that (Y, X_3) is a Pareto-Clayton model. That is, X_1 and X_2 are comonotonic (or fully dependent) and the dependence between X_1 and X_3 is given by a survival Clayton of parameter $1/2$. Clearly the distribution of the sum $S = X_1 + X_2 + X_3$ is equal to the distribution of the sum of the Pareto-Clayton model in dimension 2, with parameters $\alpha = 2$ and $\beta = 1$ and thus the exact value of the quantiles can be easily computed. We compare the results on the quantiles estimation using the ECBC method without and with information on the sub-vector (X_1, X_2) and $\lambda-$decompositions $\mathcal{I}_m$ for $m = 6, 15, 30$. As before the multivariate sample is of size $n = 30$ and for each method we performed $N = 1000$ estimations of the quantile at different confidence levels. The results are presented in Table 2.

It can be noticed that the RSME of the quantile estimation is lower when the information on (X_1, X_2) is introduced in the ECBC of dimension 3 and the gap is more important on higher confidence levels.

Remark. Simulating with respect to the empirical checkerboard copula with information on a sub-vector may be a difficult task because one has to simulate with respect to a given copula conditionally to belonging to a given set $I_{i,m}$. In the case of a comonotonic sub-vector, this becomes trivial because we only need to simulate one coordinate uniformly.

Simulation results with the same kind of model in dimension 6 are presented in Table 3. We assumed $\mathbf{X} = (X_1, \ldots, X_6)$ with $X_1 = X_2 = Y/2$ and X_3, X_4, X_5 and X_6 distributed as Y, a Pareto r.v. with parameter $\alpha = 2$. The copula of $\mathbf{X}$ is assumed to be a survival Clayton of parameter $1/2$. As above, the size of the multivariate sample is $n = 30$ and for each method we performed $N = 1000$ estimations of the quantile at different confidence levels.

Again, by introducing the information on the sub-vector (X_1, X_2) we get a smaller RMSE than in the case where no information is added. On the other hand, we also remark that by increasing the dimension (from $d = 3$ to

	Quantile 80%	Quantile 90%	Quantile 95%	Quantile 99%	Quantile 99.5%	Quantile 99.9%
Exact	2.5	4.1	6.4	16.0	23.2	53.4
ECBC (m=6)						
No information	2.7	4.6	6.6	14.0	19.1	40.7
	(13%)	(13%)	(7%)	(13%)	(18%)	(24%)
Information on	2.6	4.4	6.6	14.8	20.8	45.7
(X_1, X_2)	(9%)	(8%)	(6%)	(8%)	(11%)	(15%)
ECBC (m=10)						
No information	2.5	4.6	7.0	14.5	19.8	41.3
	(12%)	(13%)	(12%)	(11%)	(15%)	(23%)
Information on	2.5	4.3	6.7	15.2	21.2	46.1
(X_1, X_2)	(11%)	(9%)	(9%)	(8%)	(10%)	(15%)
ECBC (m=30)						
No information	2.5	4.2	6.8	15.9	21.4	43.3
	(14%)	(16%)	(19%)	(14%)	(14%)	(21%)
Information on	2.5	4.2	6.6	15.8	21.9	47.1
(X_1, X_2)	(13%)	(16%)	(17%)	(13%)	(13%)	(14%)

TABLE 2. The mean and the RMSE in % of the exact value for 1000 estimations of the quantiles in dimension 3, with or without using the knowledge of the comonotonic dependence between X_1 and X_2.

	Quantile 80%	Quantile 90%	Quantile 95%	Quantile 99%	Quantile 99.5%	Quantile 99.9%
Exact	6.1	9.8	14.9	36.4	52.4	120.1
ECBC (m=6)						
No information	7.2	11.0	15.0	28.2	37.3	74.3
	(19%)	(14%)	(7%)	(23%)	(29%)	(38%)
Information on	7.1	10.9	15.0	28.9	38.4	77.5
(X_1, X_2)	(17%)	(13%)	(7%)	(21%)	(27%)	(36%)
ECBC (m=10)						
No information	6.8	11.3	16.1	30.0	39.2	76.3
	(16%)	(18%)	(13%)	(19%)	(26%)	(37%)
Information on	6.7	11.1	16.0	30.6	40.3	79.7
(X_1, X_2)	(15%)	(16%)	(12%)	(17%)	(24%)	(34%)
ECBC (m=30)						
No information	6.3	10.4	16.3	33.7	43.4	81.3
	(15%)	(19%)	(20%)	(18%)	(22%)	(33%)
Information on	6.3	10.4	16.2	34.0	44.1	84.2
(X_1, X_2)	(15%)	(18%)	(19%)	(18%)	(20%)	(31%)

TABLE 3. The mean and the RMSE in % of the exact value for 1000 estimations of the quantiles in dimension 6.

$d = 6$) we get higher RMSE, for the same sample size, which is an expected behavior.

5. Conclusion

In this paper, we have constructed empirical check-min-erboard copulas with and without additional information on the joint law. We have used them to get efficient estimations of the quantiles of the sum when using a (relatively) small sample of the joint law and the knowledge of the marginal laws. Remark that a sample of size 200 in dimension 25 is small (and so does a sample of size 500 in dimension 100) and we get nevertheless rather good estimations. Our procedure provides a flexible tool for estimation of quantiles of sums from a *small* sample. This situation arises in many applications. In order to perform well, one has to figure out from the sample whether it is closer to an independent or comonotonic copula. Moreover, if one has partial information on the vector (copula of a sub-vector, information on the tail), it may be plugged in to improve the estimation. It is remarquable that the method remains quite efficient in high dimension. Of course, the interest of this non-parametric method would be for data for which a parametric estimation is not well suited.
We are aware that many theoretical and practical questions have to be studied further, among which the points below.

- The optimal choice of m with respect to the sample size n. The choice of the median value could allow avoid this problem. Simulations indicate that this may be a good choice when the data is not too close to independence. In that case, the checkmin copula should be preferred to the checkerboard one and the median estimation performs well. For data closer to independence, the checkerboard copula should be chosen and then it seems that the higher m the better the estimation. These claims would require theoretical support;
- The quantification of the impact of plugging additional information in the empirical checkerboard copula;
- Developing efficient algorithms to simulate with respect to the empirical checkerboard copula with information on a sub-vector, for other copulas than the comonotonic one;
- Developing efficient algorithms to simulate with respect to the empirical checkerboard copula with information on the tail in dimension larger than 2.

Acknowledgements
We are grateful to anonymous referees and to the editor for many interesting remarks who helped to improve the article. This work has been partially supported by the MultiRisk LEFE-MANU project and the research chair *Actuariat responsable* sponsored by Generali.

References

[1] Carole Bernard, Ludger Rüschendorf, and Steven Vanduffel. Value-at-Risk bounds with variance constraints. 2013.
[2] Carole Bernard and Steven Vanduffel. A new approach to assessing model risk in high dimensions. 2014.
[3] Jean-François Coeurjolly. Median-based estimation of the intensity of a spatial point process. *Annals of the Institute of Statistical Mathematics*, pages 1–29, 2016.

[4] Hélène Cossette, Marie-Pier Côté, Mélina Mailhot, and Etienne Marceau. A note on the computation of sharp numerical bounds for the distribution of the sum, product or ratio of dependent risks. *Journal of Multivariate Analysis*, 130:1–20, 2014.
[5] Andrés Cuberos, Esterina Masiello, and Véronique Maume-Deschamps. High level quantile approximations of sums of risks. *Dependence Modeling*, 3(1):141–158, 2015.
[6] Michel Dacorogna, Leila El Bahtouri, and Marie Kratz. Explicit diversification benefit for dependent risks. preprint, 2014.
[7] William F Darsow, Bao Nguyen, and Elwood T Olsen. Copulas and markov processes. *Illinois Journal of Mathematics*, 36(4):600–642, 1992.
[8] Paul Deheuvels. La fonction de dépendance empirique et ses propriétés. *Acad. Roy. Belg. Bull. Cl. Sci.*, 65(5):274 – 292, 1979.
[9] Satya D Dubey. Compound gamma, beta and F distributions. *Metrika*, 16(1):27–31, 1970.
[10] Fabrizio Durante, Juan Fernández-Sánchez, José Juan Quesada-Molina, and Ùbeda-Flores Manuel. Convergence results for patchwork copulas. *European Journal of Operational Research*, 247:525–531, 2015.
[11] Paul Embrechts and Giovanni Puccetti. Risk aggregation. *Copula Theory and Its Applications*, 2010.
[12] Paul Embrechts, Giovanni Puccetti, and Ludger Rüschendorf. Model uncertainty and var aggregation. *Journal of Banking & Finance*, 37(8):2750–2764, 2013.
[13] Paul Embrechts, Giovanni Puccetti, and Ludger Rüschendorf. Model uncertainty and var aggregation. *Journal of Banking & Finance*, 37(8):2750–2764, 2013.
[14] Jean-David Fermanian, Dragan Radulovic, and Marten Wegkamp. Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–960, 2004.
[15] Christian Genest and Johanna Nešlehová. A primer on copulas for count data. *Astin Bulletin*, 37(2):475–515, 2007.
[16] Christian Genest, Johanna G Nešlehová, and Bruno Rémillard. On the empirical multilinear copula process for count data. *Bernoulli*, 20(3):1344–1371, 2014.
[17] Xin Li, P Mikusiński, and Michael D Taylor. Strong approximation of copulas. *Journal of Mathematical Analysis and Applications*, 225(2):608–623, 1998.
[18] Georg Mainik. Risk aggregation with empirical margins: Latin hypercubes, empirical copulas, and convergence of sum distributions. *Journal of Multivariate Analysis*, 141:197–216, 2015.
[19] Véronique Maume-Deschamps, Didier Rullière, and Khalil Saïd. Impact of dependence on some multivariate risk indicators. 2015.
[20] Piotr Mikusinski and Michael D Taylor. Some approximations of n-copulas. *Metrika*, 72(3):385–414, 2010.
[21] Johanna Nešlehová. On rank correlation measures for non-continuous random variables. *Journal of Multivariate Analysis*, 98(3):544–567, 2007.
[22] David Oakes. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84(406):487–493, 1989.
[23] Giovanni Puccetti and Ludger Rüschendorf. Computation of sharp bounds on the distribution of a function of dependent risks. *Journal of Computational and Applied Mathematics*, 236(7):1833–1840, 2012.
[24] Ludger Rüschendorf. Random variables with maximum sums. *Advances in Applied Probability*, pages 623–632, 1982.
[25] Ludger Rüschendorf. Solution of a statistical optimization problem by rearrangement methods. *Metrika*, 30(1):55–61, 1983.
[26] AW Van der Vaart and JA Wellner. *Weak Convergence and Empirical Processes.* Springer, 1996.
[27] Hsiaw-Chan Yeh. The frailty and the Archimedean structure of the general multivariate Pareto distributions. *Bulletin Institute of Mathematics Acedemia Sinica*, 2(3):713–729, 2007.

Appendix A. Algorithm

Let us describe how to simulate from a checkerboard or checkermin copula with sample $\mathbf{x}_1 = (x_{11}, \dots, x_{1d}), \dots, \mathbf{x}_n = (x_{n1}, \dots, x_{nd})$ and partition $\mathcal{I}_m$: Using the rank marginals, transform the sample of the copula in the pseudo-sample $\mathbf{u}_1, \dots \mathbf{u}_n$, where

$$\mathbf{u}_i = (u_{i1}, \dots, u_{id}) = \left(\frac{R_{i1} - 1}{n}, \dots, \frac{R_{id} - 1}{n} \right),$$

where R_{ij} is the rank of x_{ij} amongst $(x_{1j}, \dots, x_{nj})$. Notice that the coordinates of $\mathbf{u}_i$ belong to the set $\{0, 1/n, \dots, (n-1)/n\}$.

(1) Choose randomly and uniformly one vector $\mathbf{u}_i$ from the pseudo-sample $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$.
(2) Let $\tilde{\mathbf{u}} = (\tilde{u}_1, \dots, \tilde{u}_d)$ be the vector with coordinates given by
$$\tilde{u}_j = \frac{\lfloor u_{ij} \cdot m \rfloor}{m}.$$
I.e, each coordinate $\tilde{u}_j$ in the vector $\tilde{\mathbf{u}}$ is the largest rational of the form k/m less than or equal to u_{ij}.
(3) Simulate an element $\mathbf{v} \in [0,1]^d$ according to an independent or comonotone copula, depending if we want to simulate from a checkerboard or a checkermin copula respectively.
(4) Then, the vector
$$\mathbf{z} = \tilde{\mathbf{u}} + \mathbf{v}/m$$
is a simulation from the required checkerboard or checkermin copula.

Appendix B. Tables

	Quantile 80%	Quantile 90%	Quantile 95%	Quantile 99%	Quantile 99.5%	Quantile 99.9%
Exact value	16.43	23.08	31.28	59.10	76.41	135.89
ECBC, $m = 2$	2%	16%	29%	51%	58%	68%
ECBC, $m = 4$	10%	5%	17%	43%	51%	64%
ECBC, $m = 5$	11%	4%	14%	40%	48%	63%
ECBC, $m = 8$	12%	6%	8%	33%	43%	59%
ECBC, $m = 10$	12%	8%	7%	30%	40%	58%
ECBC, $m = 20$	10%	9%	9%	21%	31%	52%
ECBC, $m = 40$	8%	9%	11%	18%	26%	48%
ECBC, $m = 80$	8%	9%	12%	23%	25%	44%
ECBC, median estimation	9%	5%	8%	31%	41%	59%
ECMC, $m = 2$	2%	7%	12%	18%	20%	24%
ECMC, $m = 4$	1%	2%	3%	5%	5%	13%
ECMC, $m = 5$	2%	3%	4%	6%	7%	13%
ECMC, $m = 8$	4%	3%	5%	10%	12%	16%
ECMC, $m = 10$	5%	3%	6%	11%	13%	19%
ECMC, $m = 20$	7%	5%	6%	14%	17%	23%
ECMC, $m = 40$	7%	6%	7%	15%	19%	27%
ECMC, $m = 80$	8%	7%	10%	16%	21%	32%
ECMC, median estimation	3%	3%	4%	9%	11%	15%
Gaussian copula	6%	3%	10%	27%	34%	48%
Survival Clayton copula	1%	2%	3%	5%	6%	12%
Clayton copula	5%	10%	23%	46%	54%	66%
Empirical copula	8%	9%	12%	23%	31%	56%

Table 4. RMSE in % of the exact value for the Pareto-Clayton model of parameters $\beta = 3$ and $\alpha = 1$, in dimension 25, for a sample size $n = 80$.

	Quantile 80%	Quantile 90%	Quantile 95%	Quantile 99%	Quantile 99.5%	Quantile 99.9%
Exact value	16.43	23.08	31.28	59.10	76.41	135.89
ECBC, $m = 2$	1%	16%	29%	51%	58%	68%
ECBC, $m = 4$	10%	4%	17%	42%	50%	64%
ECBC, $m = 5$	11%	3%	13%	39%	48%	63%
ECBC, $m = 8$	11%	6%	6%	31%	41%	59%
ECBC, $m = 10$	11%	7%	5%	27%	38%	56%
ECBC, $m = 20$	8%	8%	7%	17%	28%	50%
ECBC, $m = 25$	7%	8%	8%	15%	24%	47%
ECBC, $m = 40$	5%	6%	9%	12%	19%	43%
ECBC, $m = 50$	5%	6%	9%	12%	18%	40%
ECBC, $m = 100$	5%	6%	8%	15%	18%	35%
ECBC, $m = 200$	5%	6%	8%	16%	20%	32%
ECBC, median estimation	6%	5%	6%	18%	28%	50%
ECMC, $m = 2$	2%	7%	11%	18%	20%	25%
ECMC, $m = 4$	1%	2%	2%	4%	6%	11%
ECMC, $m = 5$	1%	2%	3%	5%	6%	12%
ECMC, $m = 8$	3%	2%	5%	9%	10%	14%
ECMC, $m = 10$	4%	2%	5%	9%	11%	15%
ECMC, $m = 20$	5%	4%	4%	11%	14%	19%
ECMC, $m = 25$	5%	4%	4%	11%	15%	20%
ECMC, $m = 40$	4%	5%	6%	10%	14%	22%
ECMC, $m = 50$	4%	5%	6%	10%	14%	23%
ECMC, $m = 100$	5%	5%	7%	11%	14%	23%
ECMC, $m = 200$	5%	6%	8%	13%	15%	28%
ECMC, median estimation	3%	3%	4%	8%	11%	14%
Gaussian copula	6%	2%	10%	27%	34%	48%
Survival Clayton copula	1%	1%	2%	4%	5%	10%
Clayton coupla	5%	10%	23%	46%	53%	65%
Empirical coupla	5%	6%	9%	17%	23%	42%

TABLE 5. RMSE in % of the exact value for the Pareto-Clayton model with parameters $\alpha = 3$ and $\beta = 1$, in dimension 25, for a sample size $n = 200$.

	Quantile 80%	Quantile 90%	Quantile 95%	Quantile 99%	Quantile 99.5%	Quantile 99.9%
Near exact value	93.94	111.65	129.81	176.99	200.82	270.14
ECBC, $m = 2$	4%	7%	10%	14%	15%	15%
ECBC, $m4$	2%	4%	7%	11%	12%	13%
ECBC, $m = 5$	2%	4%	6%	10%	11%	13%
ECBC, $m = 8$	2%	3%	5%	9%	10%	12%
ECBC, $m = 10$	2%	3%	4%	8%	10%	12%
ECBC, $m = 20$	2%	3%	4%	8%	9%	11%
ECBC, $m = 40$	3%	4%	4%	9%	9%	11%
ECBC, $m = 80$	3%	4%	5%	10%	11%	12%
ECBC, median estimation	2%	3%	5%	9%	10%	11%
ECMC, $m = 2$	3%	17%	34%	76%	95%	141%
ECMC, $m = 4$	2%	6%	14%	41%	53%	83%
ECMC, $m = 5$	2%	3%	11%	33%	44%	72%
ECMC, $m = 8$	3%	2%	5%	19%	27%	46%
ECMC, $m = 10$	2%	2%	3%	14%	21%	38%
ECMC, $m = 20$	2%	3%	3%	7%	10%	22%
ECMC, $m = 40$	2%	3%	4%	7%	8%	15%
ECMC, $m = 80$	3%	4%	5%	8%	10%	13%
ECMC, median estimation	2%	2%	4%	17%	24%	41%
Gaussian copula	1%	2%	2%	3%	4%	6%
Survival Clayton copula	2%	2%	3%	9%	12%	20%
Clayton copula	3%	7%	9%	13%	14%	14%
Empirical copula	5%	6%	9%	16%	22%	35%

TABLE 6. RMSE in % of the exact value for the Gaussian lognormal model with correlation $\rho = 0.1$ for all pairs, in dimension 25, for a sample size $n = 80$.

	Quantile 80%	Quantile 90%	Quantile 95%	Quantile 99%	Quantile 99.5%	Quantile 99.9%
Near exact value	93.94	111.70	129.88	176.26	201.28	270.35
ECBC, $m = 2$	3%	7%	10%	14%	14%	14%
ECBC, $m = 4$	1%	4%	6%	10%	12%	13%
ECBC, $m = 5$	1%	3%	5%	10%	11%	12%
ECBC, $m = 8$	1%	2%	4%	8%	9%	12%
ECBC, $m = 10$	1%	2%	3%	7%	8%	11%
ECBC, $m = 20$	2%	2%	3%	6%	7%	10%
ECBC, $m = 25$	2%	2%	3%	6%	7%	10%
ECBC, $m = 40$	2%	2%	3%	6%	7%	10%
ECBC, $m = 50$	2%	3%	3%	6%	8%	10%
ECBC, $m = 100$	2%	3%	4%	6%	9%	10%
ECBC, $m = 200$	2%	3%	4%	7%	9%	11%
ECBC, median estimation	1%	2%	3%	6%	8%	10%
ECMC, $m = 2$	3%	18%	34%	76%	97%	142%
ECMC, $m = 4$	2%	6%	15%	40%	53%	85%
ECMC, $m = 5$	2%	4%	11%	33%	43%	70%
ECMC, $m = 8$	2%	1%	5%	20%	28%	47%
ECMC, $m = 10$	2%	1%	3%	15%	21%	37%
ECMC, $m = 20$	2%	2%	2%	6%	10%	24%
ECMC, $m = 25$	2%	2%	3%	6%	9%	18%
ECMC, $m = 40$	2%	2%	3%	5%	6%	11%
ECMC, $m = 50$	2%	2%	3%	5%	6%	11%
ECMC, $m = 100$	2%	3%	4%	6%	7%	12%
ECMC, $m = 200$	2%	3%	4%	6%	8%	11%
ECMC, median estimation	1%	2%	3%	7%	11%	25%
Gaussian copula	1%	1%	2%	3%	3%	6%
Survival Calyton copula	2%	1%	3%	9%	13%	20%
Clayton copula	3%	6%	9%	13%	14%	15%
Empirical copula	3%	4%	6%	10%	13%	29%

TABLE 7. RMSE in % of the exact value for the Gaussian lognormal model with correlation $\rho = 0.1$ for all pairs, in dimension 25, for a sample size $n = 200$.

Appendix C. Boxplots

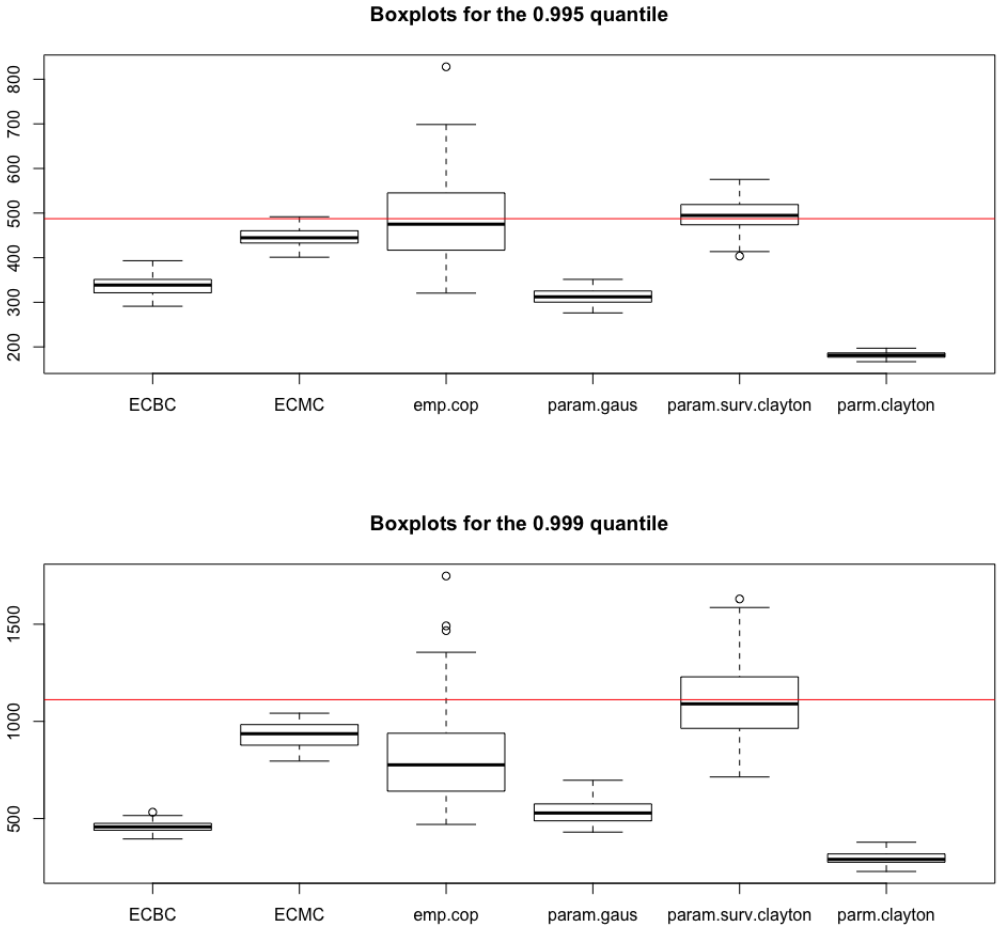

Figure 4. Pareto-Clayton model with parameters $\beta = 2$ and $\alpha = 1$, in dimension 50.

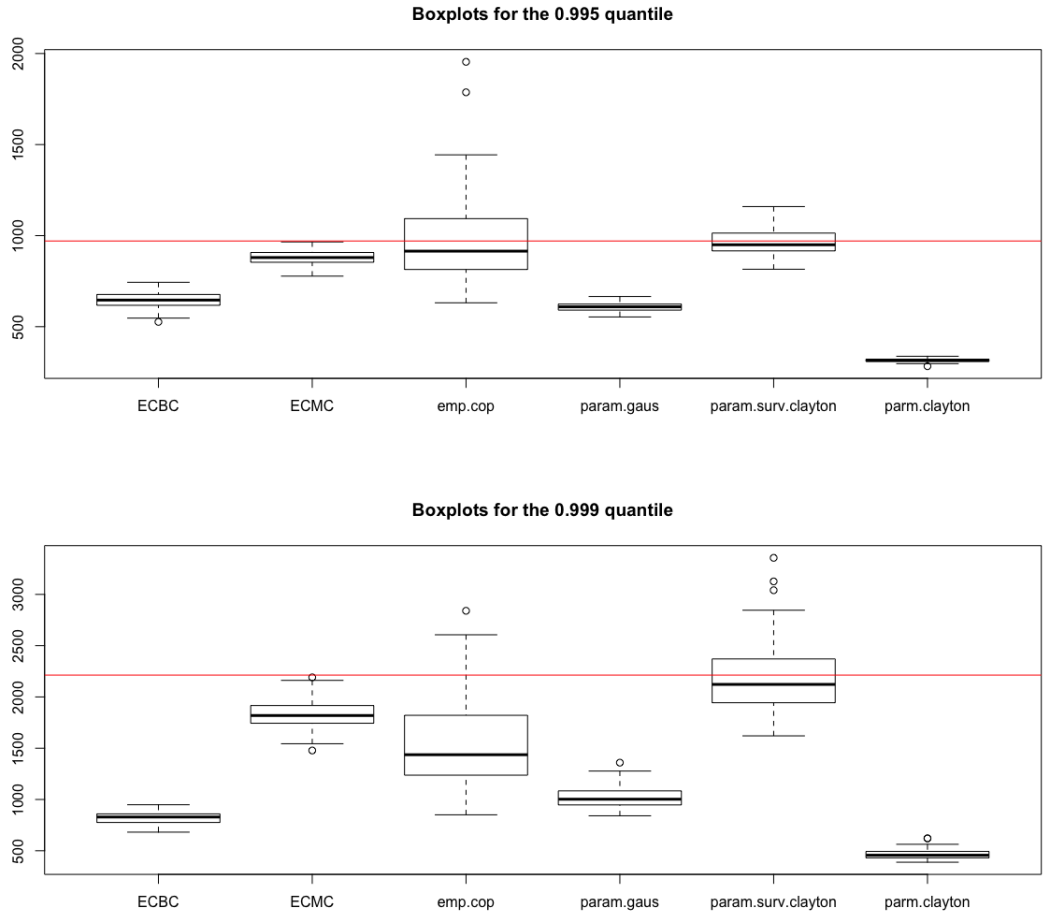

FIGURE 5. Pareto-Clayton model with parameters $\beta = 2$ and $\alpha = 1$, in dimension 100.

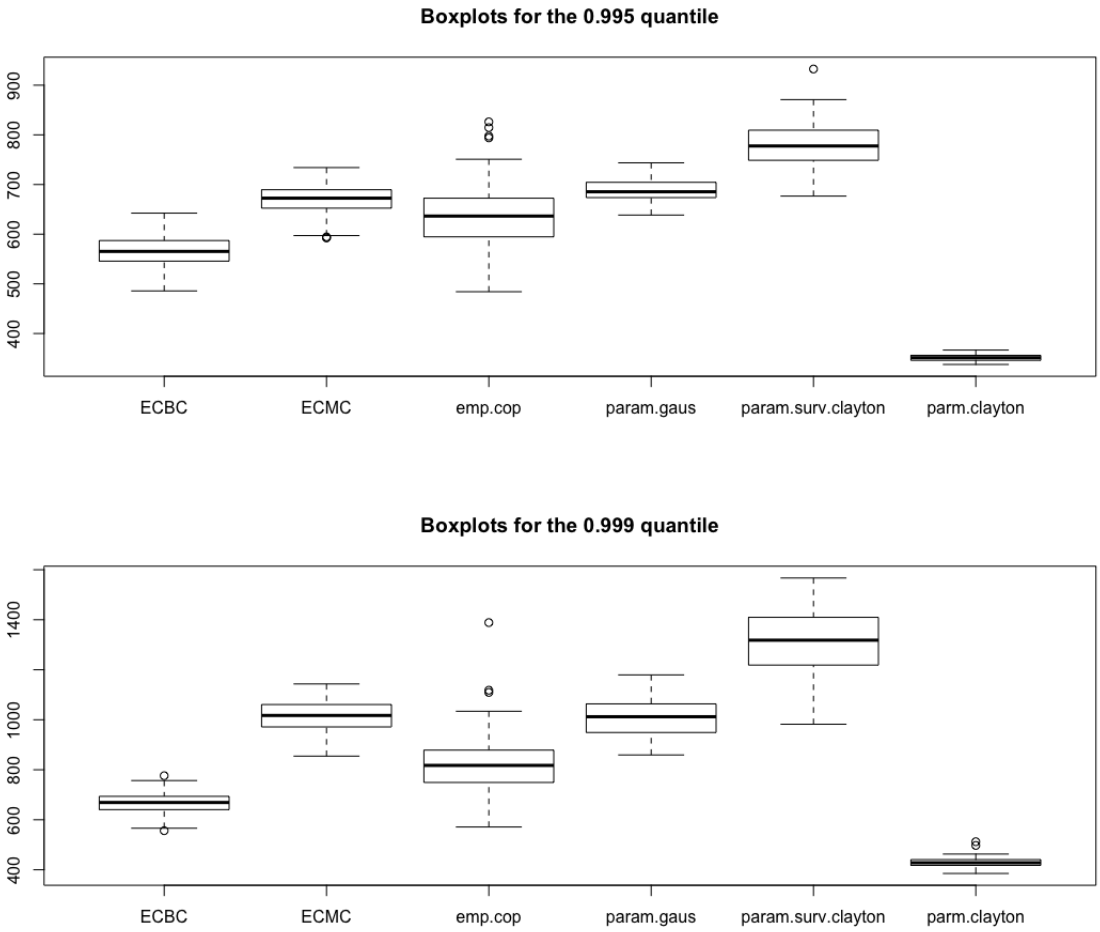

FIGURE 6. Gaussian-lognormal model with correlations $0.25, 0.5, 0.75$, in dimension 50.

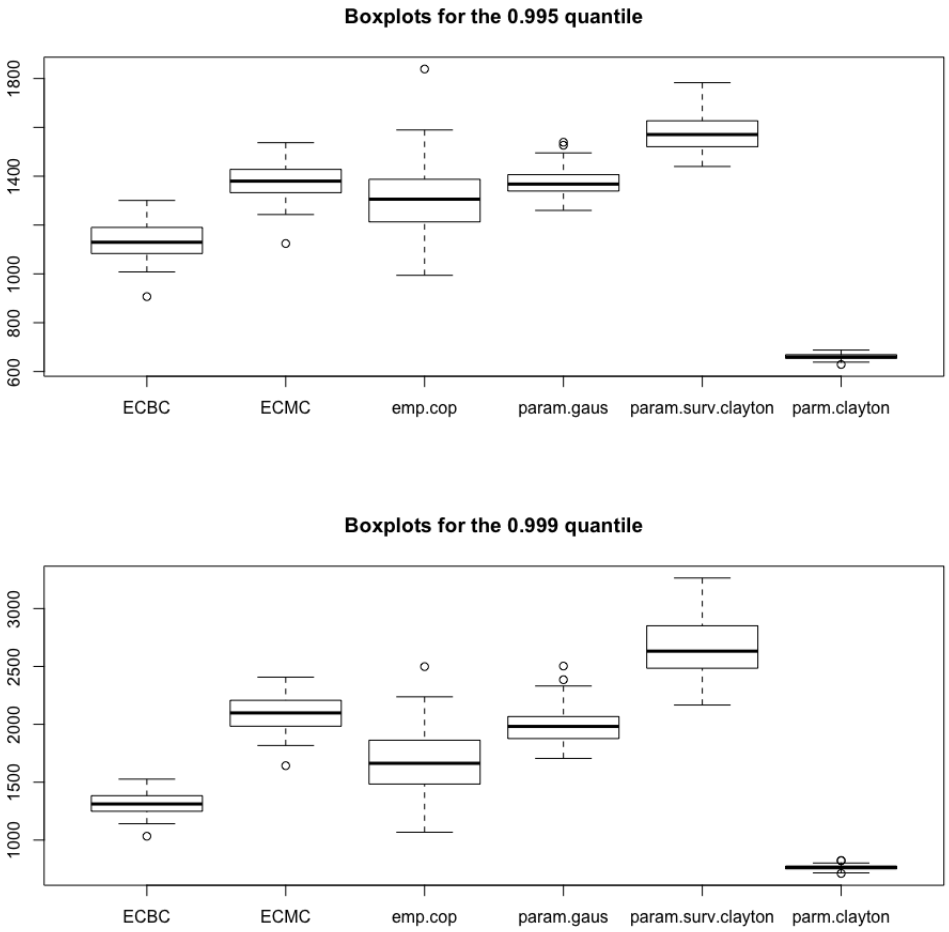

FIGURE 7. Gaussian-lognormal model with correlations $0.25, 0.5, 0.75$, in dimension 100.

SCOR SE, 5 AVENUE KLÉBER, 75795 PARIS CEDEX 16, FRANCE
E-mail address: acuberos@scor.com

UNIVERSITÉ DE LYON, UNIVERSITÉ LYON 1, INSTITUT CAMILLE JORDAN ICJ UMR 5208 CNRS
E-mail address: esterina.masiello@univ-lyon1.fr

UNIVERSITÉ DE LYON, UNIVERSITÉ LYON 1, INSTITUT CAMILLE JORDAN ICJ UMR 5208 CNRS
E-mail address: veronique.maume@univ-lyon1.fr