

*Quantification of the environmental structural risk with spoiling ties: is randomization worthwhile?*

**R. Pappadà, F. Durante & G. Salvadori**

# Quantification of the environmental structural risk with spoiling ties: is randomization worthwhile?

R. Pappadà<sup>1</sup> · F. Durante<sup>2</sup> · G. Salvadori<sup>3</sup>

**Abstract** Many recent works show that copulas turn out to be useful in a variety of different applications, especially in environmental sciences. Here the variables of interest are usually continuous, being times, lengths, weights, and so on. Unfortunately, the corresponding observations may suffer from (instrumental) adjustments and truncations, and eventually may show several repeated values (i.e., ties). In turn, on the one hand, a tricky issue of identifiability of the model arises, and, on the other hand, the assessment of the risk may be adversely affected. A possible remedy is to adopt suitable randomization procedures: here three different strategies are outlined. The goal of the work is to carry out a simulation study in order to evaluate the effects of the randomization of multivariate observations when ties are present. In particular, it is investigated whether, how, and to what extent, the randomization may change the estimation of the structural risk: for this purpose, a coastal

engineering example will be used, as archetypical of a broad class of models and problems in engineering applications. Practical advices and warnings about the use of randomization techniques are hence given.

**Keywords** Copula · Risk management · Randomization · Jittering · Structural risk

## 1 Introduction

Copulas have proved to be useful in a variety of different applications, especially in environmental sciences (see, e.g., Genest and Favre 2007; Salvadori et al. 2007; Genest and Nešlehová 2012a, b; AghaKouchak et al. 2013), where they contribute to quantify the risk in a suitable way. In fact, it is well known that the description of the joint probability law of a vector of random variables can be conveniently represented via Sklar’s theorem (Sklar 1959) as the composition of a copula and one-dimensional marginals governing the phenomenon of interest. In particular, both the copula and the marginal laws are chosen and fitted on a set of available data, considered as an i.i.d. sample from an unknown continuous joint distribution.

The emphasis on the adjective “continuous” is extremely important in the present context. In fact, if the marginals are continuous, then

- the observations assume (with probability 1) distinct values ranging in the support of the underlying distribution, and no *ties* (i.e., repeated observations) occur in the dataset;
- the copula associated with the underlying random vector can be uniquely determined.

---

✉ R. Pappadà  
rpappada@units.it

F. Durante  
fabrizio.durante@unibz.it

G. Salvadori  
gianfausto.salvadori@unisalento.it

<sup>1</sup> Department of Economics, Business, Mathematics and Statistics “Bruno de Finetti”, University of Trieste, 34127 Trieste, Italy

<sup>2</sup> Faculty of Economics and Management, Free University of Bozen-Bolzano, 39100 Bolzano, Italy

<sup>3</sup> Dipartimento di Matematica e Fisica “Ennio De Giorgi”, Università del Salento, 73100 Lecce, Italy

On the contrary, when the marginals are not continuous, the data typically contain ties, and a tricky issue of identifiability of the model arises. For an overview about possible problems using copulas with non-continuous data, see the excellent survey by Genest and Nešlehová (2007) (and also Marshall 1996).

In many applications of environmental sciences, however, the situation is somehow mixed. While it is not questionable (for physical reasons) that the random variables of interest can be viewed as continuous, the available measurements may suffer from adjustments and truncations (due, e.g., to instrumental limitations or sampling procedures), so that they may show several repeated observations. Now, as documented e.g. in Genest et al. (2011) and Bücher and Kojadinovic (2016), the presence of such repetitions may have a non-negligible impact on the rank-based inference of copulas: for instance, the performances of popular Goodness-of-Fit tests for copulas cannot be guaranteed anymore.

As stressed in Bücher and Kojadinovic (2016), when ties are present, the “copula-oriented” practitioner has (at least) two possibilities—excluding the limiting case of stopping any further statistical analysis: (i) discard the ties; (ii) randomize the data, by adding a suitable continuous noise to all observations. This paper focuses on the latter methodology which is known in literature as *randomization*, sometimes also called *jittering*. Note that the former case has been discussed in Genest et al. (2011). The goal of this note is to carry out a simulation study in order to evaluate the effects of the randomization of multivariate observations when ties are present. In particular, it will be investigated whether, how, and to what extent, the randomization may change the estimation of the structural risk, using a coastal engineering example representing an archetype of a broad class of models and problems in engineering practice. To this end, a practical illustration, involving a realistic simulation study tailored to a dataset previously investigated in other works, will be used. The results shown may provide practical advices and warnings about the adoption of randomization techniques.

## 2 The illustration

In the following, a coastal engineering application, related to the design of a rubble mound breakwater described in Salvadori et al. (2014) (see also Salvadori et al. 2015; Pappadà et al. 2016), is used. The target is to compute the quantiles associated with the weight  $W$  of a concrete cube element forming the breakwater structure, assuming that the environmental load is given by the pair of non-independent random variables  $(H, D)$ , where  $H$  represents the significant wave height (in meters), and  $D$  the sea storm

duration (in hours). For this purpose, a structural model  $\Psi$  is used, expressing  $W$  as a function of  $(H, D)$  by means of the formula

$$\begin{aligned} W &= \Psi(H, D) \\ &= \rho_S \cdot \left[ H \left( \frac{2\pi H}{g [4.597 \cdot H^{0.328}]^2} \right)^{0.1} \right]^3 / \\ &\quad \left[ \left( \frac{\rho_S}{\rho_W} - 1 \right) \cdot \left( 1 + \frac{6.7 \cdot N_d^{0.4}}{(3600D/[4.597 \cdot H^{0.328}])^{0.3}} \right) \right]^3. \end{aligned} \quad (1)$$

The values of the structural parameters  $g$ ,  $\rho_W$ ,  $\rho_S$ ,  $N_d$  in Eq. (1) are calibrated for the buoy of Alghero (Sardinia, Italy), previously investigated in Salvadori et al. (2015, Table 1).

Following a copula approach to the structural risk (see, for instance, Straub 2014), in order to estimate the quantiles associated with  $W$ , it is possible to proceed as follows.

- First, fit suitable univariate distribution functions for  $H$  and  $D$ , respectively.
- Then, fit a suitable copula for  $(H, D)$ .
- Finally, calculate the quantile of  $W$  either analytically (if possible), or by resampling from the joint distribution of  $(H, D)$  previously obtained via Sklar’s Theorem.

It is important to realize that, in principle, both  $H$  and  $D$  describe continuous phenomena (viz., a length and a time). Unfortunately, due to a limited (buoy) instrumental resolution and the particular sampling procedures adopted in coastal engineering practice, the available measurements may be a discretized version of the actual continuous values of these variables: for instance, this is the case of the sea storm data presented in Salvadori et al. (2014). Thus, ties may occur, and adversely affect the statistical analysis of the data both at the marginal level and at the copula level—see also De Michele et al. (2013, Sect. 3; Fig. 2) for another hydrological case study. Concerning this latter aspect, the situation is particularly problematic, since copula-based procedures generally require the possibility of uniquely determining the ranks of the observations.

A practical way to circumvent the problem could consist in adding random components to the coordinates of each observed pair  $(X_i, Y_i)$ —here, the pair  $(H, D)$ —by setting  $\tilde{X}_i = X_i + \Delta_X U_i$  and  $\tilde{Y}_i = Y_i + \Delta_Y V_i$ ,  $i = 1, \dots, N$ ,

$$(2)$$

where  $N$  is the available sample size,  $\Delta_X$  and  $\Delta_Y$  are the data resolutions, and  $U_1, \dots, U_N$  and  $V_1, \dots, V_N$  are suitable random samples from the uniform distribution on  $[0, 1]$ . Clearly,  $X$  (respectively,  $Y$ ) takes value on a discrete set, with points equi-spaced by a distance  $\Delta_X$  (respectively,

$\Delta_Y$ ). For instance, ( $\Delta_X = 1 \text{ m}^3/\text{s}, \Delta_Y = 1 \text{ day}$ ) in De Michele et al. (2013), and ( $\Delta_X = 10 \text{ cm}, \Delta_Y = 3 \text{ h}$ ) in Salvadori et al. (2014).

In the following, only the bivariate case (i.e.,  $d = 2$ ) will be investigated: however, the approach proposed can easily be generalized to the case  $d > 2$ . Here, three different randomization strategies will be considered, as outlined below.

- The independent randomization, which assumes that, for every index  $i$ ,  $U_i$  and  $V_i$  are independent, viz. are coupled via the product (independence) copula  $\Pi_2(u, v) = uv$ .
- The co-monotone randomization, which assumes that, for every index  $i$ ,  $U_i$  and  $V_i$  are coupled by the Fréchet–Hoeffding upper bound copula  $M_2(u, v) = \min\{u, v\}$ .
- The mixed randomization, which assumes that, for every index  $i$ ,  $U_i$  and  $V_i$  are coupled by a mixture  $C_\lambda$  of  $M_2$  and  $\Pi_2$ , viz.

$$C_\lambda(u, v) = \lambda M_2(u, v) + (1 - \lambda)\Pi_2(u, v), \quad (3)$$

with  $\lambda \in (0, 1)$ . Note that the mixing coefficient  $\lambda$  coincides with the Kendall’s  $\tau$  associated with the mixing copula  $C_\lambda$ . Here,  $\lambda$  is related to degree of association of the observations  $(X_i, Y_i)$ ’s, and can be computed via the available data simply by estimating the corresponding Kendall’s  $\tau$ —see below.

It is worth noting that  $C_\lambda$  is simply a member of the well known family of Fréchet copulas (Nelsen 2006). In the present case, such a mixture is used since it represents an “intermediate” case between the independence and the Co-monotone ones. Clearly, should it be appropriate, any other copula could be used instead.

*Remark 1* At a formal level, if ties are present, a pair of random variables is uniquely associated with a sub-copula, not with a copula (see, for instance, Durante and Sempi 2016). Thus, every type of randomization can be thought of as a specific way to extend a sub-copula to a copula (see, e.g., de Amo et al. 2012). In particular, the independent randomization is related to the multilinear extension of Genest et al. (2014) (see also Durante et al. 2015), while the co-monotone and mixed ones are associated with extensions that distribute the probability mass according to, respectively,  $M_2$  and  $C_\lambda$ .

Since the sample  $(\tilde{X}_i, \tilde{Y}_i)$ ’s is generated via a randomization process, it is crucial to perform a large number of independent randomizations in order to carry out a sensible statistical assessment. In particular, a suitable *distributional* analysis of the outcomes of the simulations should be carried out—see below, and Kojadinovic and Yan (2011) and Bücher and Kojadinovic (2016): in fact, averaging the results does not seem to mitigate the

inconveniences produced by the randomization, as shown in Genest and Nešlehová (2007) and Genest et al. (2011).

As mentioned in the Introduction, while it is arguable that the randomization procedures may provide indications for statistical inference for copulas (like parameter estimation, Goodness-of-Fit test, etc.), it is unclear whether, and to what extent, the risk quantification (in a given situation) can be affected by randomizing as well. In the present study, starting with the practical application described above, an attempt to shed light on this latter aspect is made.

In order to focus on those features that characterize the practical problem addressed here, the following assumptions are made.

- Both  $H$  and  $D$  follow the generalized Weibull distributions  $F_H$  and  $F_D$ , with the same parameters as those estimated in Salvadori et al. (2014, Table 1) for a specific dataset. Such a “realistic” setting gives the possibility to draw sensible conclusions about the structural risk from a practical perspective (at least concerning the archetypical case study considered here).
- The dependence structure of the pair  $(H, D)$  is modeled via different families of copulas with a specified parameter: namely, the Frank, the Gumbel, and the Clayton—these will be denoted by, respectively,  $C^{\text{Frk}}$ ,  $C^{\text{Gmb}}$ , and  $C^{\text{Cly}}$ . A number of practical reasons suggest the adoption of such families. On the one hand, they are often used in hydrological studies (see, e.g., many of the papers listed at the site [www.stahy.org](http://www.stahy.org), section /Topics/CopulaFunction). On the other hand, they are able to express different types of tail dependence (Salvadori et al. 2007, Sects. 3.4, 5.3): in particular, Frank copulas have both tail dependence coefficients equal to zero, whereas Gumbel ones only model non-zero upper tail dependence coefficient, and Clayton ones only show non-zero lower tail dependence coefficient. In addition, the Gumbel copula is Extreme Value, and the Clayton one is truncation invariant with respect to both variables (see, e.g., Nelsen 2006; Salvadori et al. 2007; Di Lascio et al. 2016).

### 3 The simulation study

The simulation study proposed in the following will reproduce the main features outlined above. For the sake of brevity and clarity, the procedure is explained assuming that the Frank copula is the “true” one (i.e., the one modeling the dependence structure of the data, from which suitable random samples are eventually extracted). The same steps can then be performed substituting,

respectively, the Gumbel and the Clayton families for the Frank one.

### 3.1 Discretization and randomization

In order to generate samples of discretized (truncated) observations, and to test the performance of the randomization strategies mentioned above, the following steps are carried out.

1. A set of  $N$  i.i.d. observations of the random pair  $(H, D)$  is generated, according to the *continuous* joint distribution function  $F^{\text{Frk}} = C^{\text{Frk}}(F_H, F_D)$  constructed via Sklar’s Theorem.
2. The simulated observations are truncated according to a predefined resolution. More specifically,
  - the measurements of  $H$  may have a basic resolution equal to  $\Delta_H \in \{0.01, 0.1, 0.5\}$  meters;
  - the measurements of  $D$  may have a basic resolution equal to  $\Delta_D \in \{0.5, 1, 3\}$  hours.

Note that the discretization levels given above roughly correspond to the actual ones for real buoy data. For instance, taking  $\Delta_H = 0.1$  and  $\Delta_D = 1$ , the (continuous) observation  $(H = \pi, D = \pi)$  would be turned into  $(3.1, 3)$ , i.e. it would be truncated. The resulting dataset generally presents several ties, and may conveniently reproduce some features observed in practice.

3. The randomization procedures previously described are applied to the discretized dataset, in order to carry out the “jittering”.

Given the randomized data, two questions will be considered.

- Is a Goodness-of-Fit test able to correctly identify the copula  $C^{\text{Frk}}$  that generates the sample? Viz., is it able to reject the assumption that the dependence structure of the data belongs to another family such as the Gumbel, the Clayton, etc.?
- If the dependence structure is correctly identified, does the copula estimated using the randomized data help to provide valuable information concerning the structural risk associated with the random variable  $W$  defined via Eq. (1)?

The answer to these questions will depend on the following three aspects.

1. The sample size  $N$ , which is set equal to 150 or 300—as in Kojadinovic et al. (2011).
2. The degree of association between  $H$  and  $D$ , which is expressed in terms of the Kendall’s  $\tau$ , and takes on values in  $\{0.25, 0.5, 0.75\}$ —as in Kojadinovic et al. (2011).

3. The jittering strategy, viz. the independent, co-monotone, or mixed randomization, where in the latter case the mixing coefficient  $\lambda$  is set equal to the sample Kendall’s  $\tau$  estimated by using the discretized pairs  $(H, D)$ ’s.

### 3.2 Model identification

The first analysis of the present study concerns the investigation of how the presence of ties may affect the performance of the rank-invariant procedures typically used for the identification of the copula of  $(H, D)$ . For  $i \in \{1, \dots, B\}$ , the following steps are repeated.

1. Simulate  $N$  pairs  $(H, D)$  from the model described above (based on Frank copulas), with three levels of dependence as given by  $\tau \in \{0.25, 0.5, 0.75\}$ .
2. Apply the truncation with resolutions  $(\Delta_H, \Delta_D) \in \{(0.01, 0.5), (0.1, 1), (0.5, 3)\}$ .
3. Randomize the discretized dataset (which typically contains ties) according to the three randomization strategies previously outlined.
4. Carry out a Crámer–von Mises Goodness-of-Fit test—here at a 5% significance level, using the multiplier variant proposed in Kojadinovic et al. (2011) and Kojadinovic and Yan (2011)—of the Null hypothesis  $\mathcal{H}_0$  that the copula of the randomized data belongs to a given family  $\mathcal{F}$  against the alternative that it does not belong to  $\mathcal{F}$ . Here, five families of dependence structures are considered: viz., the Frank, the Gumbel, the Clayton, the Normal, and the Student (with  $\nu = 4$  degrees of freedom).

The results are discussed in Sect. 4 and Tables 1, 2, 3, 4, 5 and 6.

In case the Goodness-of-Fit test does not reject the Null hypothesis that the copula belongs to the “true” family (at the given significance level), it could also be convenient to check how the parameter estimation varies with respect to the true value. The results are discussed in Sect. 4 and Figs. 1, 3 and 5.

### 3.3 Structural risk

As a further step, it is also useful to investigate how, and to what extent, the randomization procedures may affect the estimation of the structural risk, as represented by a suitable quantile of the random variable  $W$ : this task can be carried out via Monte Carlo simulations, by using the formula  $W = \psi(H, D)$  given by Eq. 1. The quantiles of order  $q = 0.9, 0.95, 0.99$  of  $W$ —the ones usually of interest in applications—with fixed marginals for  $H$  and  $D$ , are

**Table 1** “Frank” case:  $N = 150, B = 10,000$ 

$\Delta_H$	$\Delta_D$	$C^{\text{Frk}}$	$C^{\text{Cly}}$	$C^{\text{Gmb}}$	$C^{\text{Nrm}}$	$C^{t,4}$
Independent randomization						
$\tau = 0.25$						
0.01	0.50	0.05124	0.79127	0.38351	0.13594	0.26872
0.10	1.00	0.04945	0.79477	0.39801	0.14384	0.27942
0.50	3.00	0.07014	0.85656	0.37201	0.21383	0.34402
$\tau = 0.5$						
0.01	0.50	0.05024	0.99835	0.79057	0.42781	0.70698
0.10	1.00	0.05244	0.99865	0.78977	0.45780	0.73238
0.50	3.00	0.13344	0.99985	0.78817	0.68108	0.88006
$\tau = 0.75$						
0.01	0.50	0.04695	0.99995	0.94936	0.87446	0.95415
0.10	1.00	0.05314	0.99995	0.95485	0.90776	0.96545
0.50	3.00	0.48510	0.99995	0.96105	0.98595	0.99805
Mixed randomization						
$\tau = 0.25$						
0.01	0.50	0.05624	0.78357	0.40301	0.14954	0.28982
0.10	1.00	0.04865	0.78467	0.39181	0.14284	0.27142
0.50	3.00	0.06284	0.80867	0.36471	0.17653	0.28032
$\tau = 0.5$						
0.01	0.50	0.05134	0.99745	0.78987	0.43161	0.70758
0.10	1.00	0.05074	0.99705	0.79227	0.43841	0.70018
0.50	3.00	0.08604	0.99775	0.77617	0.53160	0.71798
$\tau = 0.75$						
0.01	0.50	0.04665	0.99995	0.95155	0.87826	0.95105
0.10	1.00	0.05044	0.99995	0.95225	0.87116	0.94086
0.50	3.00	0.26952	0.99995	0.96675	0.94256	0.97885
Co-monotone randomization						
$\tau = 0.25$						
0.01	0.50	0.05414	0.78937	0.38831	0.14044	0.27842
0.10	1.00	0.05454	0.77127	0.39981	0.13644	0.26362
0.50	3.00	0.11154	0.65968	0.51110	0.17803	0.23263
$\tau = 0.5$						
0.01	0.50	0.05104	0.99755	0.79537	0.43281	0.71168
0.10	1.00	0.05124	0.99655	0.79307	0.43211	0.68258
0.50	3.00	0.14354	0.98635	0.86526	0.50110	0.63939
$\tau = 0.75$						
0.01	0.50	0.04485	0.99995	0.94456	0.87906	0.95015
0.10	1.00	0.04625	0.99965	0.95305	0.86916	0.93116
0.50	3.00	0.37711	0.99995	0.98645	0.94596	0.97625

Probability of rejection (nominal level 5%) of the Null hypothesis that the copula belongs, respectively, to the Frank ( $C^{\text{Frk}}$ ), Clayton ( $C^{\text{Cly}}$ ), Gumbel ( $C^{\text{Gmb}}$ ), Normal ( $C^{\text{Nrm}}$ ), and Student- $t$  ( $C^{t,4}$ , with  $\nu = 4$  degrees of freedom) family, for a random sample generated from a Frank copula, with  $\tau = 0.25, 0.5, 0.75$  obtained from independent repetitions of the discretization and randomization procedures—see text

calculated, under different parameters as estimated from the “true” copula when properly identified. The results are discussed in Sect. 4 and Figs. 2, 4 and 6.

**Table 2** “Frank” case:  $N = 300, B = 10,000$ 

$\Delta_H$	$\Delta_D$	$C^{\text{Frk}}$	$C^{\text{Cly}}$	$C^{\text{Gmb}}$	$C^{\text{Nrm}}$	$C^{t,4}$
Independent randomization						
$\tau = 0.25$						
0.01	0.50	0.04465	0.96075	0.72028	0.25282	0.58729
0.10	1.00	0.04515	0.96545	0.72578	0.25852	0.58699
0.50	3.00	0.07664	0.98395	0.71198	0.40311	0.69888
$\tau = 0.50$						
0.01	0.50	0.04155	0.99995	0.98875	0.79937	0.96845
0.10	1.00	0.04305	0.99995	0.98985	0.82347	0.97525
0.50	3.00	0.19843	0.99995	0.99185	0.95615	0.99695
$\tau = 0.75$						
0.01	0.50	0.03545	0.99995	0.99995	0.99725	0.99965
0.10	1.00	0.04265	0.99995	0.99985	0.99835	0.99975
0.50	3.00	0.78647	0.99995	0.99985	0.99995	0.99995
Mixed randomization						
$\tau = 0.25$						
0.01	0.50	0.04675	0.96225	0.72508	0.25832	0.58509
0.10	1.00	0.04595	0.95625	0.71688	0.26052	0.58359
0.50	3.00	0.06414	0.96565	0.69228	0.32792	0.59529
$\tau = 0.50$						
0.01	0.50	0.04275	0.99995	0.98915	0.80207	0.96975
0.10	1.00	0.04675	0.99995	0.99055	0.80187	0.96585
0.50	3.00	0.10064	0.99995	0.98765	0.87986	0.96985
$\tau = 0.75$						
0.01	0.50	0.03345	0.99995	0.99995	0.99535	0.99995
0.10	1.00	0.03945	0.99995	0.99965	0.99645	0.99935
0.50	3.00	0.46760	0.99995	0.99995	0.99955	0.99985
Co-monotone randomization						
$\tau = 0.25$						
0.01	0.50	0.04485	0.96405	0.71368	0.25312	0.58819
0.10	1.00	0.04225	0.94796	0.71128	0.24493	0.55539
0.50	3.00	0.14684	0.87166	0.84107	0.31142	0.47770
$\tau = 0.50$						
0.01	0.50	0.04465	0.99995	0.98875	0.79737	0.96805
0.10	1.00	0.04775	0.99985	0.99025	0.78527	0.95415
0.50	3.00	0.24703	0.99995	0.99655	0.86776	0.93476
$\tau = 0.75$						
0.01	0.50	0.03445	0.99995	0.99985	0.99605	0.99985
0.10	1.00	0.04525	0.99995	0.99965	0.99475	0.99885
0.50	3.00	0.71128	0.99995	0.99995	0.99975	0.99995

Same as Table 1

## 4 Results

As already mentioned above, three families of copulas are used as “true” models to study the effects of the randomization strategies outlined in this work: viz., the Frank, the Gumbel, and the Clayton. Briefly, random samples of different sizes and degrees of association are simulated

**Table 3** “Gumbel” case:  $N = 150, B = 2500$ 

$\Delta_H$	$\Delta_D$	$C^{\text{Gmb}}$	$C^{\text{Frk}}$	$C^{\text{Cly}}$	$C^{\text{Nrm}}$	$C^{t,4}$
Independent randomization						
$\tau = 0.25$						
0.01	0.5	0.03699	0.3469	0.9166	0.1949	0.2497
0.1	1	0.03699	0.3353	0.9086	0.1909	0.2381
0.5	3	0.04138	0.3828	0.9498	0.2821	0.3405
$\tau = 0.5$						
0.01	0.5	0.03659	0.6803	0.9998	0.378	0.41
0.1	1	0.03938	0.6771	0.9998	0.4044	0.4692
0.5	3	0.1098	0.7859	0.9998	0.7123	0.7691
$\tau = 0.75$						
0.01	0.5	0.03259	0.883	0.9998	0.4996	0.504
0.1	1	0.05858	0.8886	0.9998	0.6539	0.6551
0.5	3	0.5672	0.9806	0.9998	0.9842	0.987
Mixed randomization						
$\tau = 0.25$						
0.01	0.5	0.03739	0.3373	0.913	0.2033	0.2533
0.1	1	0.03619	0.3445	0.909	0.1929	0.2437
0.5	3	0.04098	0.3457	0.919	0.2277	0.2769
$\tau = 0.5$						
0.01	0.5	0.03778	0.6851	0.9998	0.3884	0.4348
0.1	1	0.03739	0.6831	0.9998	0.3964	0.4392
0.5	3	0.05738	0.7395	0.9998	0.56	0.5684
$\tau = 0.75$						
0.01	0.5	0.03059	0.8914	0.9998	0.4744	0.4868
0.1	1	0.03778	0.897	0.9998	0.55	0.52
0.5	3	0.2497	0.9702	0.9998	0.9042	0.8858
Co-monotone randomization						
$\tau = 0.25$						
0.01	0.5	0.03619	0.3149	0.901	0.1849	0.2325
0.1	1	0.04098	0.3165	0.8998	0.1781	0.2181
0.5	3	0.07937	0.3501	0.8531	0.1741	0.1793
$\tau = 0.5$						
0.01	0.5	0.03459	0.6847	0.9998	0.3585	0.398
0.1	1	0.03898	0.6747	0.9998	0.3796	0.4028
0.5	3	0.08217	0.7715	0.9998	0.4532	0.4228
$\tau = 0.75$						
0.01	0.5	0.02779	0.8946	0.9998	0.4948	0.4972
0.1	1	0.04138	0.8874	0.9998	0.5344	0.498
0.5	3	0.2613	0.981	0.9998	0.8747	0.8455

Probability of rejection (nominal level 5%) of the Null hypothesis that the copula belongs, respectively, to the Gumbel ( $C^{\text{Gmb}}$ ), Frank ( $C^{\text{Frk}}$ ), Clayton ( $C^{\text{Cly}}$ ), Normal ( $C^{\text{Nrm}}$ ), and Student- $t$  ( $C^{t,4}$ , with  $v = 4$  degrees of freedom) family, for a random sample generated from a Gumbel copula, with  $\tau = 0.25, 0.5, 0.75$  obtained from independent repetitions of the discretization and randomization procedures—see text

from either the  $F^{\text{Frk}}$ , or the  $F^{\text{Gmb}}$ , or the  $F^{\text{Cly}}$  joint distributions: in turn, these latter represent the “true” bivariate model. Then, the data are discretized and randomized

**Table 4** “Gumbel” case:  $N = 300, B = 2500$ 

$\Delta_H$	$\Delta_D$	$C^{\text{Gmb}}$	$C^{\text{Frk}}$	$C^{\text{Cly}}$	$C^{\text{Nrm}}$	$C^{t,4}$
Independent randomization						
$\tau = 0.25$						
0.01	0.5	0.04218	0.5392	0.9906	0.3413	0.416
0.1	1	0.03739	0.5684	0.9958	0.3349	0.4288
0.5	3	0.04538	0.6188	0.9986	0.4808	0.6084
$\tau = 0.5$						
0.01	0.5	0.03898	0.9334	0.9998	0.6208	0.6779
0.1	1	0.05018	0.9262	0.9998	0.6683	0.7147
0.5	3	0.2181	0.9682	0.9998	0.9518	0.9718
$\tau = 0.75$						
0.01	0.5	0.03699	0.9966	0.9998	0.7347	0.7511
0.1	1	0.05218	0.993	0.9998	0.8966	0.8922
0.5	3	0.8858	0.9998	0.9998	0.9998	0.9998
Mixed randomization						
$\tau = 0.25$						
0.01	0.5	0.03778	0.5688	0.9938	0.3361	0.4308
0.1	1	0.04298	0.5524	0.9934	0.3329	0.4148
0.5	3	0.04778	0.5752	0.997	0.4064	0.4972
$\tau = 0.5$						
0.01	0.5	0.03379	0.9346	0.9998	0.6567	0.6883
0.1	1	0.03579	0.9342	0.9998	0.6539	0.6719
0.5	3	0.09776	0.9574	0.9998	0.8367	0.8423
$\tau = 0.75$						
0.01	0.5	0.03419	0.9966	0.9998	0.7223	0.7307
0.1	1	0.04338	0.9946	0.9998	0.7963	0.7487
0.5	3	0.4468	0.9998	0.9998	0.9922	0.991
Co-monotone randomization						
$\tau = 0.25$						
0.01	0.5	0.04138	0.5484	0.991	0.3361	0.4212
0.1	1	0.04338	0.532	0.9922	0.3081	0.3701
0.5	3	0.1034	0.5924	0.9806	0.2673	0.2749
$\tau = 0.5$						
0.01	0.5	0.03299	0.9326	0.9998	0.6224	0.6591
0.1	1	0.04418	0.9238	0.9998	0.61	0.6339
0.5	3	0.1246	0.9662	0.9998	0.7283	0.6555
$\tau = 0.75$						
0.01	0.5	0.03059	0.9946	0.9998	0.7347	0.7351
0.1	1	0.04178	0.9962	0.9998	0.7763	0.7183
0.5	3	0.48	0.9998	0.9998	0.9882	0.9794

Same as Table 3

according to the different strategies previously outlined, and several analyses are carried out. For the sake of clarity, the results will be discussed separately for each of these three copulas. Note that the full illustration of the outcomes requires six tables and 216 figures, which cannot all be

**Table 5** “Clayton” case:  $N = 150, B = 2500$ 

$\Delta_H$	$\Delta_D$	$C^{\text{Cly}}$	$C^{\text{Frk}}$	$C^{\text{Gmb}}$	$C^{\text{Nrm}}$	$C^{t,4}$
Independent randomization						
$\tau = 0.25$						
0.01	0.5	0.07377	0.4364	0.8283	0.3049	0.3856
0.1	1	0.111	0.4092	0.8195	0.3037	0.3984
0.5	3	0.4704	0.1573	0.7327	0.2661	0.3988
$\tau = 0.5$						
0.01	0.5	0.09696	0.9058	0.9994	0.8171	0.8443
0.1	1	0.2829	0.8595	0.9978	0.8335	0.883
0.5	3	0.9638	0.2689	0.991	0.7511	0.921
$\tau = 0.75$						
0.01	0.5	0.1673	0.955	0.9998	0.9802	0.9654
0.1	1	0.7407	0.8111	0.9998	0.9798	0.9822
0.5	3	0.9998	0.1365	0.9946	0.9842	0.9962
Mixed randomization						
$\tau = 0.25$						
0.01	0.5	0.07257	0.4336	0.8091	0.3029	0.38
0.1	1	0.09216	0.4176	0.8179	0.3137	0.3988
0.5	3	0.3836	0.1821	0.7307	0.2613	0.3609
$\tau = 0.5$						
0.01	0.5	0.09536	0.9146	0.9994	0.8215	0.8375
0.1	1	0.1881	0.8886	0.9982	0.8263	0.8651
0.5	3	0.8762	0.4188	0.9974	0.7499	0.8802
$\tau = 0.75$						
0.01	0.5	0.1549	0.9614	0.9998	0.9778	0.9578
0.1	1	0.4336	0.9138	0.9998	0.977	0.9726
0.5	3	0.997	0.2709	0.9994	0.975	0.993
Co-monotone randomization						
$\tau = 0.25$						
0.01	0.5	0.06657	0.452	0.8343	0.3201	0.4092
0.1	1	0.08337	0.44	0.8343	0.3249	0.3996
0.5	3	0.2225	0.4792	0.8671	0.4524	0.4912
$\tau = 0.5$						
0.01	0.5	0.08816	0.9082	0.9982	0.8303	0.8419
0.1	1	0.1317	0.9114	0.999	0.8331	0.8531
0.5	3	0.6711	0.8087	0.9998	0.8834	0.9326
$\tau = 0.75$						
0.01	0.5	0.1429	0.9558	0.9998	0.9818	0.9614
0.1	1	0.3257	0.945	0.9998	0.983	0.975
0.5	3	0.9842	0.6291	0.9994	0.991	0.9974

Probability of rejection (nominal level 5%) of the Null hypothesis that the copula belongs, respectively, to the Clayton ( $C^{\text{Cly}}$ ), Frank ( $C^{\text{Frk}}$ ), Gumbel ( $C^{\text{Gmb}}$ ), Normal ( $C^{\text{Nrm}}$ ), and Student- $t$  ( $C^{t,4}$ , with  $\nu = 4$  degrees of freedom) family, for a random sample generated from a Clayton copula, with  $\tau = 0.25, 0.5, 0.75$  obtained from independent repetitions of the discretization and randomization procedures—see text

shown in the present paper. In turn, only six tables and selected figures will be presented here: all the others are made available in the “Supplementary Material” file.

**Table 6** “Clayton” case:  $N = 300, B = 2500$ 

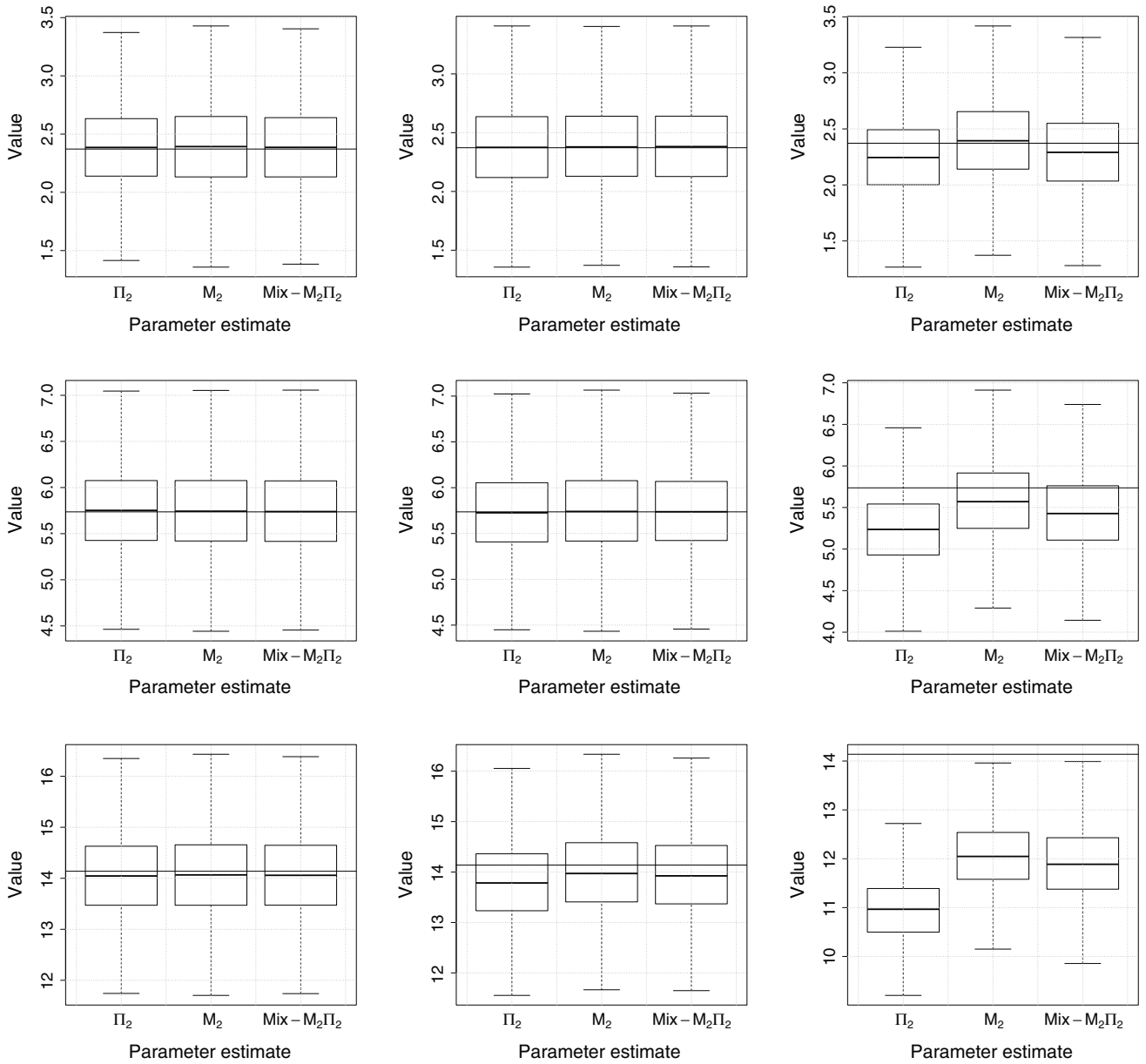
$\Delta_H$	$\Delta_D$	$C^{\text{Cly}}$	$C^{\text{Frk}}$	$C^{\text{Gmb}}$	$C^{\text{Nrm}}$	$C^{t,4}$
Independent randomization						
$\tau = 0.25$						
0.01	0.5	0.05978	0.7875	0.9874	0.6387	0.7567
0.1	1	0.111	0.7659	0.9874	0.6547	0.7839
0.5	3	0.7395	0.3389	0.9646	0.5596	0.7959
$\tau = 0.5$						
0.01	0.5	0.09456	0.9982	0.9998	0.9954	0.997
0.1	1	0.4408	0.9982	0.9998	0.9946	0.9978
0.5	3	0.9998	0.6208	0.9998	0.9894	0.9994
$\tau = 0.75$						
0.01	0.5	0.2061	0.9998	0.9998	0.9994	0.9994
0.1	1	0.9638	0.9986	0.9998	0.9998	0.9998
0.5	3	0.9998	0.4144	0.9998	0.9998	0.9998
Mixed randomization						
$\tau = 0.25$						
0.01	0.5	0.06897	0.7799	0.9882	0.6411	0.7439
0.1	1	0.1066	0.7603	0.9862	0.6275	0.7667
0.5	3	0.6124	0.3832	0.9646	0.5392	0.7331
$\tau = 0.5$						
0.01	0.5	0.101	0.999	0.9998	0.9934	0.9954
0.1	1	0.2721	0.9978	0.9998	0.995	0.9966
0.5	3	0.9822	0.8231	0.9998	0.9906	0.9966
$\tau = 0.75$						
0.01	0.5	0.1641	0.9998	0.9998	0.9998	0.9998
0.1	1	0.5924	0.9998	0.9998	0.9998	0.9998
0.5	3	0.9998	0.7571	0.9998	0.9998	0.9998
Co-monotone randomization						
$\tau = 0.25$						
0.01	0.5	0.06457	0.7875	0.9866	0.6319	0.7515
0.1	1	0.08297	0.7803	0.993	0.6467	0.7627
0.5	3	0.3221	0.7931	0.995	0.7775	0.8383
$\tau = 0.5$						
0.01	0.5	0.08617	0.9994	0.9998	0.9978	0.9982
0.1	1	0.1517	0.9978	0.9998	0.9962	0.997
0.5	3	0.8902	0.9886	0.9998	0.9978	0.9994
$\tau = 0.75$						
0.01	0.5	0.1525	0.9998	0.9998	0.9998	0.9998
0.1	1	0.4464	0.9998	0.9998	0.9998	0.9998
0.5	3	0.9998	0.9886	0.9998	0.9998	0.9998

Same as Table 5

#### 4.1 Frank case

Here the data have been sampled from the Frank family. Tables 1 and 2 report the probabilities of rejection of the Null hypothesis that the copula belongs to several families. For a sample size  $N = 150$ , in the case of weak/moderate





**Fig. 1** “Frank” case:  $N = 300$ ,  $B = 10,000$ . *Boxplots* of the copula parameter estimates: the *horizontal thick lines* indicate the “true” values—see text. From top to bottom, the rows correspond, respectively, to the values of the Kendall’s  $\tau = 0.25, 0.5, 0.75$ . From left to right, the columns correspond, respectively, to the following pairs of

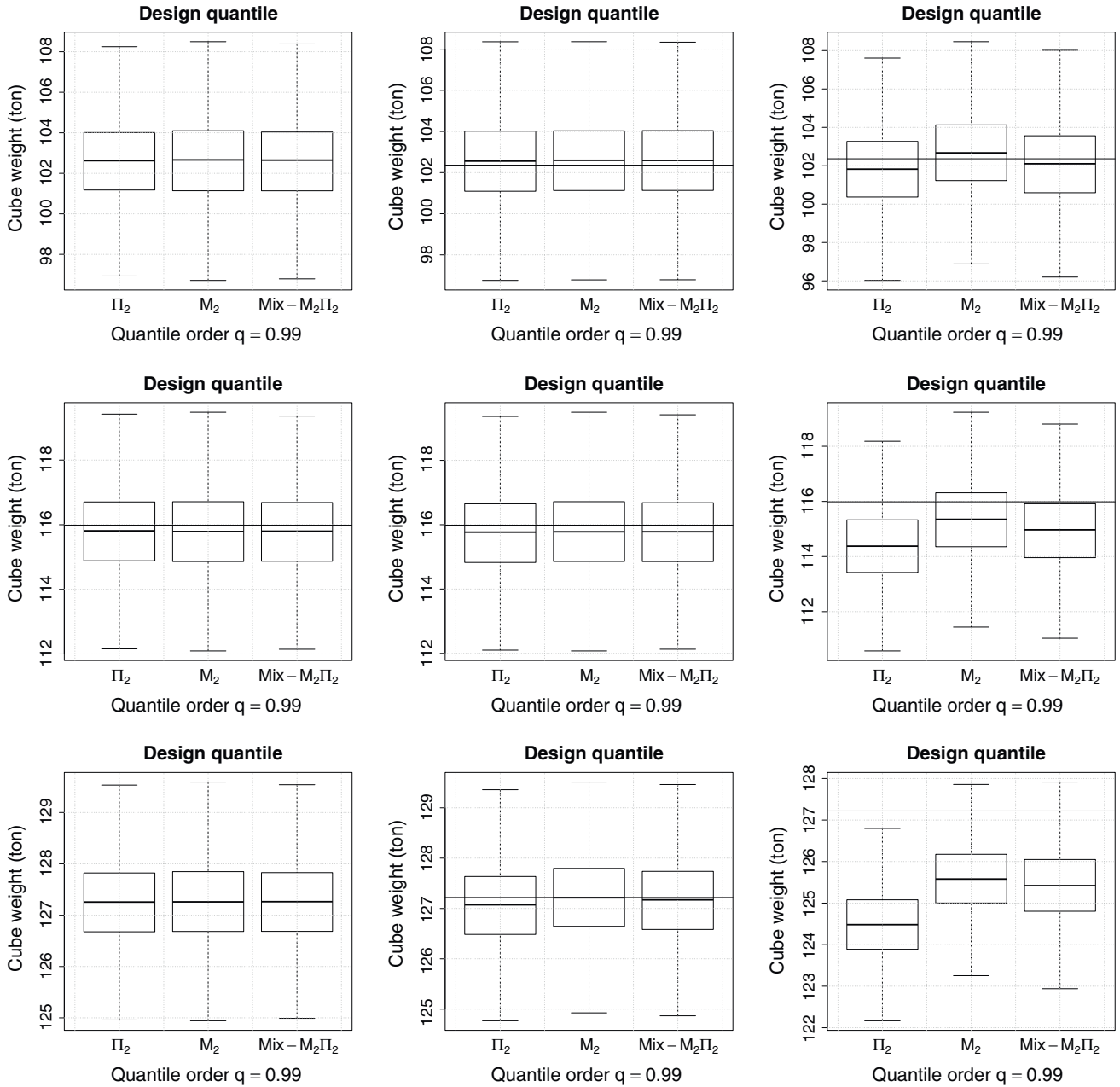
height and duration resolutions:  $(\Delta_H, \Delta_D) \in \{(0.01, 0.5), (0.1, 1), (0.5, 3)\}$ . The labels “ $\Pi_2$ ”, “ $M_2$ ”, and “ $Mix - M_2\Pi_2$ ” denote the use of an independent, a co-monotone, and a mixed randomization, respectively

discretization of the data, the Goodness-of-Fit test procedures tend to attain the nominal level (5%) for the Null hypothesis that the copula comes from the Frank family (in particular, the results are comparable with the ones obtained in Kojadinovic et al. (2011)). Instead, in the case of strong discretization, the test tends to reject the Null hypothesis too often: apparently, in this latter case, the mixed randomization seems to produce the least biased results. Furthermore, in testing the Null hypothesis that the data are extracted from another fixed copula family

different from the Frank one, the results seem overall quite reasonable. It should be noticed that, in case of weak dependence, the test does not seem to be fully able to distinguish between the Frank copula and the Normal one: the probabilities of rejection of the latter are sometimes of the order of 14%, which may not be thoroughly satisfactory. In all cases, the performance improves using larger samples of size  $N = 300$ . Concerning the estimate of the “true” parameter values, the results are illustrated in Fig. 1. In general, for a sample size  $N = 150$ , these values are

roughly identified (on average) in all cases and for all the randomization procedures, with the exception of a strong discretization and large dependence. In this latter case, the correct value of the parameter is generally underestimated and, consequently, a weaker degree of dependence is incorrectly perceived. Apparently, no improvements are achieved using larger samples of size  $N = 300$ .

The results concerning the estimate of the design quantiles are presented in Fig. 2. In general, there are no significant differences between the cases  $N = 150$  and  $N = 300$ . Overall, the approximations of the “true” values are quite reasonable: the fluctuations around the correct values are of the order of a few tons (or fractions of tons), representing small percentages of the cube weight, even in



**Fig. 2** “Frank” case:  $N = 300$ ,  $B = 10,000$ ,  $q = 0.99$ . Boxplots of the cube weight design quantiles estimates: the *horizontal thick lines* indicate the “true” values—see text. From top to bottom, the rows correspond, respectively, to the values of the Kendall’s  $\tau = 0.25, 0.5, 0.75$ . From left to right, the columns correspond,

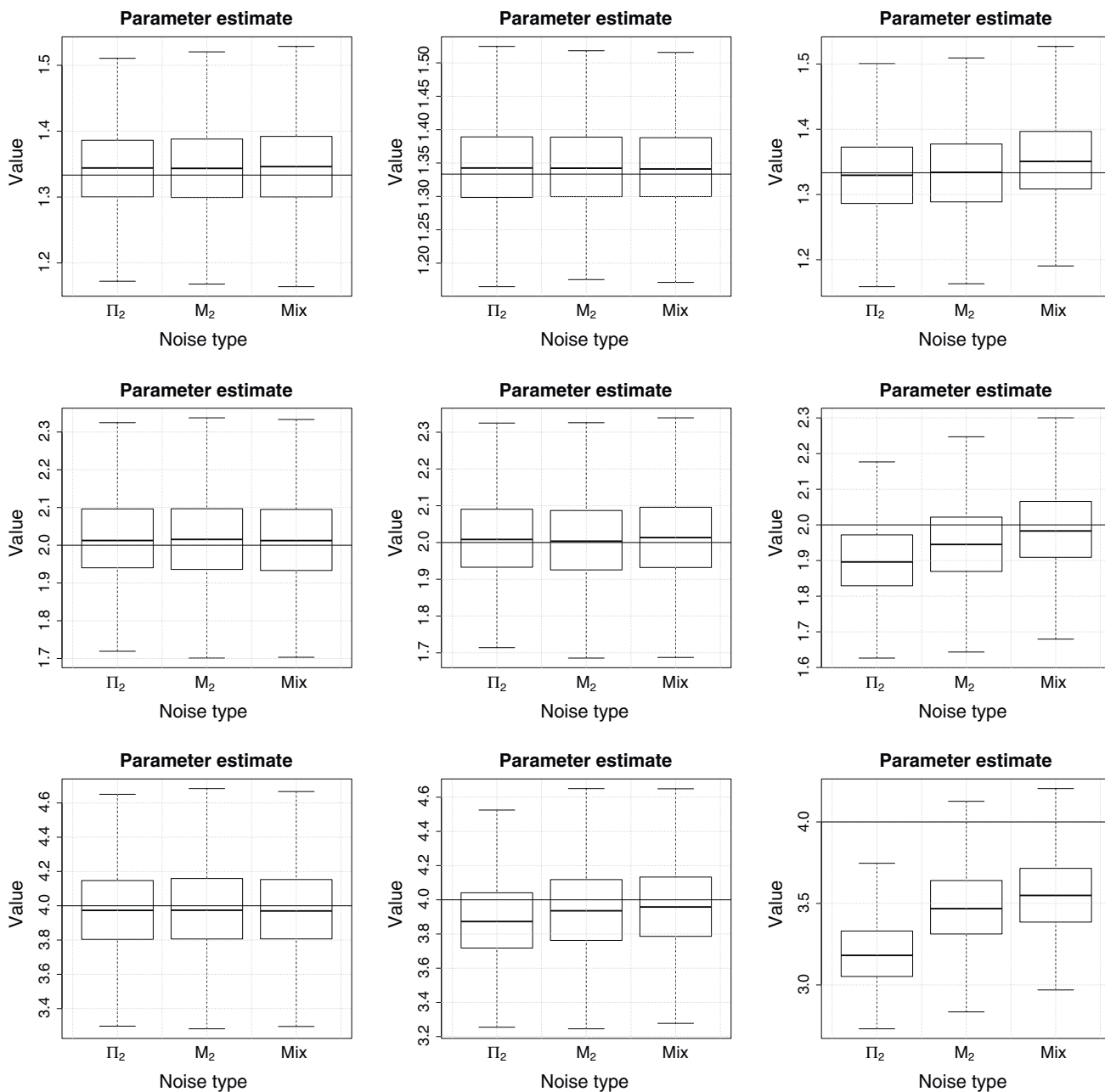
respectively, to the following pairs of height and duration resolutions:  $(\Delta_H, \Delta_D) \in \{(0.01, 0.5), (0.1, 1), (0.5, 3)\}$ . The labels “ $\Pi_2$ ”, “ $M_2$ ”, and “mix- $M_2\Pi_2$ ” denote the use of an independent, a co-monotone, and a mixed randomization, respectively

case strong discretization and large dependence are imposed.

## 4.2 Gumbel case

Here the data have been sampled from the Gumbel family. Tables 3 and 4 report the probabilities of rejection of the Null hypothesis that the copula belongs to several families.

For a sample size  $N = 150$ , in almost all cases the Goodness-of-Fit test procedures tend to attain the nominal level (5%) for the Null hypothesis that the copula comes from the Gumbel family, independently of the degree of discretization, the value of  $\tau$ , and the randomization strategy. The only exceptions are in case of strong discretization and one of the following situations: (i) for  $\tau = 0.5, 0.75$  and the independent procedure; (ii) for  $\tau = 0.75$  and the mixed



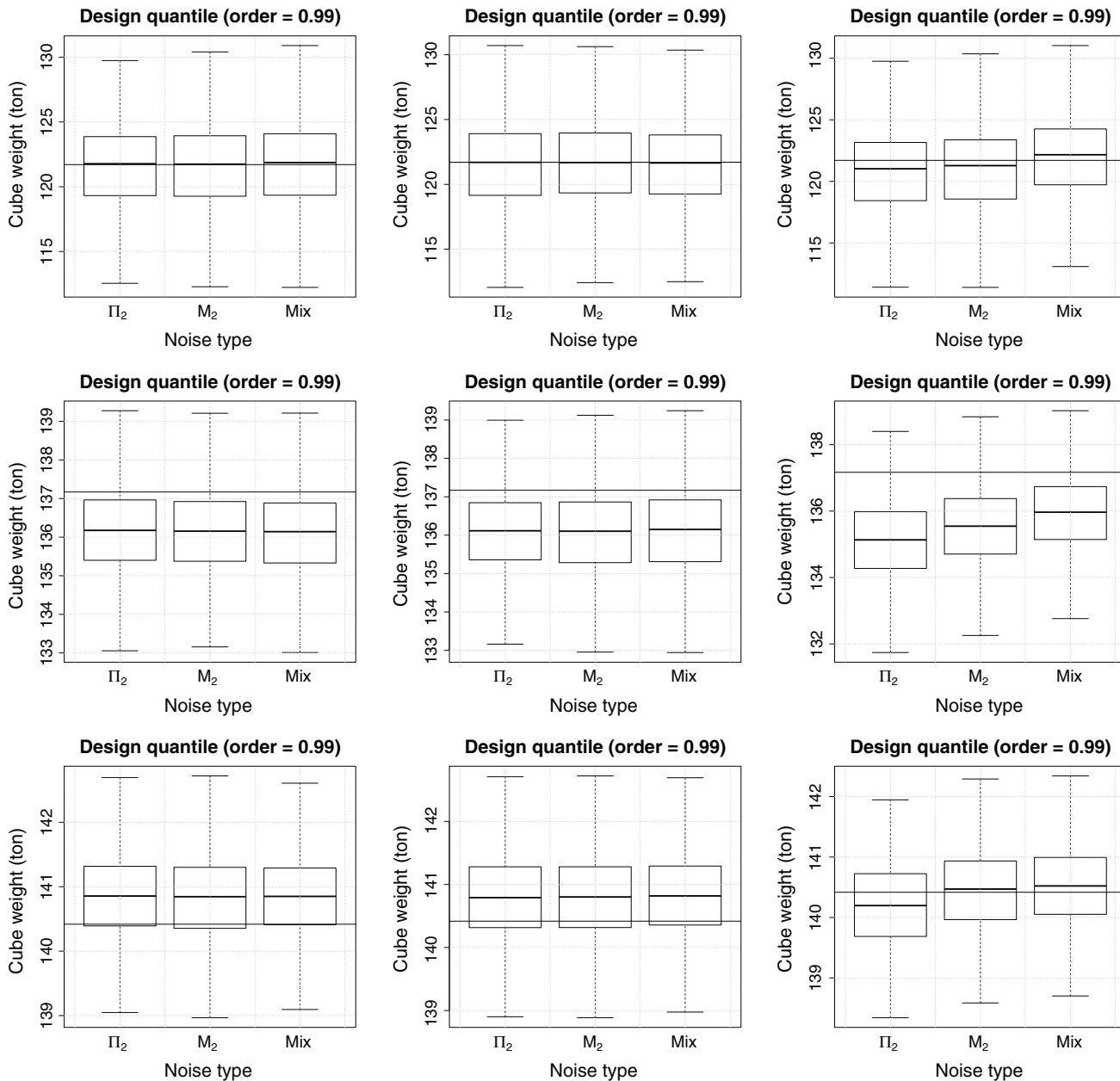
**Fig. 3** “Gumbel” case:  $N = 300$ ,  $B = 2500$ . Boxplots of the copula parameter estimates: the horizontal thick lines indicate the “true” values—see text. From top to bottom, the rows correspond, respectively, to the values of the Kendall’s  $\tau = 0.25, 0.5, 0.75$ . From left to right, the columns correspond, respectively, to the following pairs of

height and duration resolutions:  $(\Delta_H, \Delta_D) \in \{(0.01, 0.5), (0.1, 1), (0.5, 3)\}$ . The labels “ $\Pi_2$ ”, “ $M_2$ ”, and “mix” denote the use of an independent, a co-monotone, and a mixed randomization, respectively

procedure; (iii) for all  $\tau$ 's and the co-monotone procedure. Furthermore, in testing the Null hypothesis that the data are extracted from another fixed copula family different from the Gumbel one, the results seem overall quite reasonable. The performance is about the same using larger samples of size  $N = 300$ .

Concerning the estimate of the “true” parameter values, the results are illustrated in Fig. 3. In general, for a sample size  $N = 150$ , these values are roughly identified (on

average) in all cases and for all the randomization procedures, with the exception of a strong discretization and large dependence. In this latter case, the correct value of the parameter is generally underestimated and, consequently, a weaker degree of dependence is incorrectly perceived. Apparently, no significant improvements are achieved using larger samples of size  $N = 300$ , except that the variability of the estimates slightly decreases.



**Fig. 4** “Gumbel” case:  $N = 300$ ,  $B = 2500$ ,  $q = 0.99$ . Boxplots of the cube weight design quantiles estimates: the *horizontal thick lines* indicate the “true” values—see text. From top to bottom, the rows correspond, respectively, to the values of the Kendall’s  $\tau = 0.25, 0.5, 0.75$ . From left to right, the columns correspond,

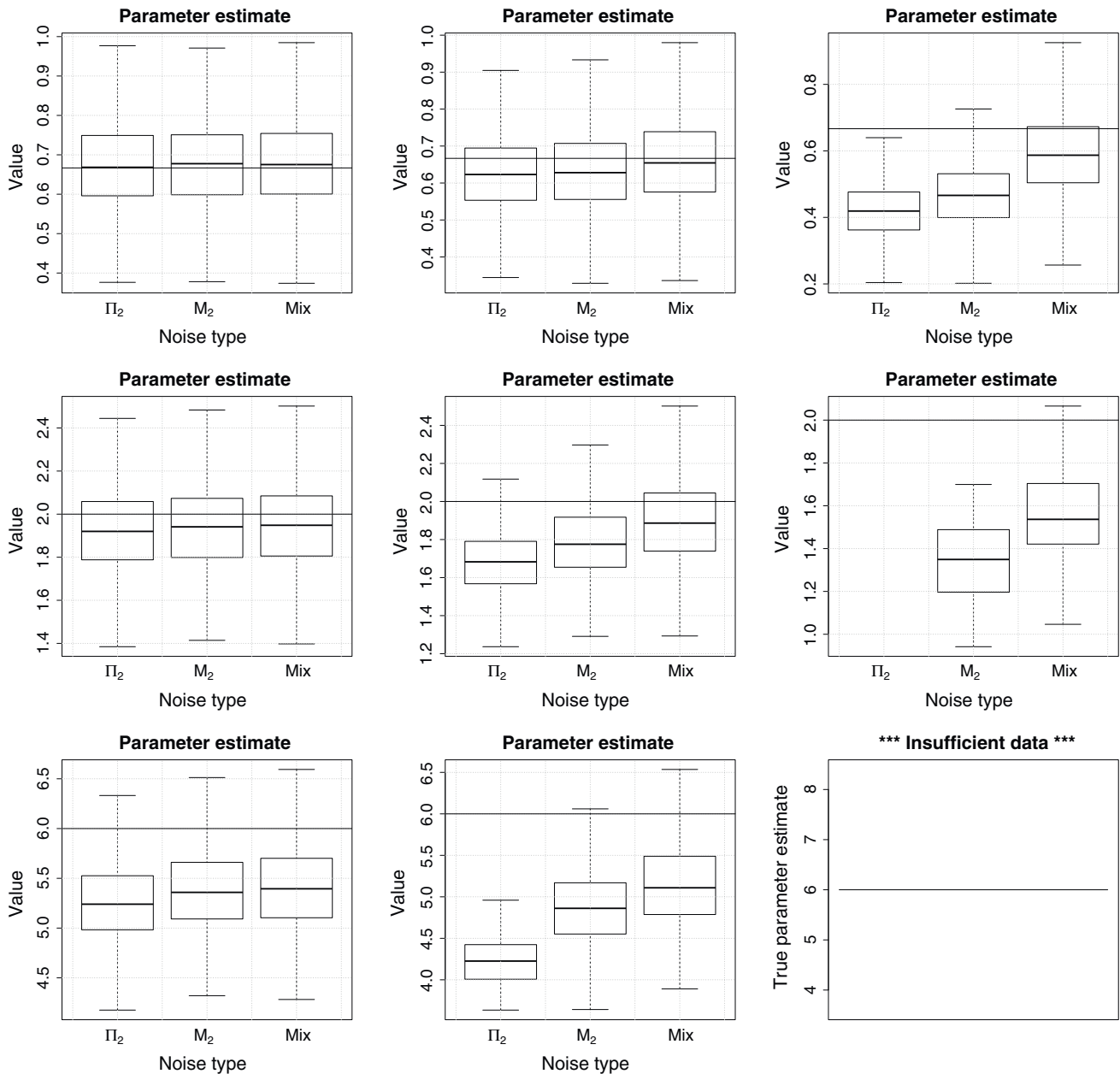
respectively, to the following pairs of height and duration resolutions:  $(\Delta_H, \Delta_D) \in \{(0.01, 0.5), (0.1, 1), (0.5, 3)\}$ . The labels “ $\Pi_2$ ”, “ $M_2$ ”, and “mix” denote the use of an independent, a co-monotone, and a mixed randomization, respectively

The results concerning the estimate of the design quantiles are presented in Fig. 4. In general, there are no significant differences between the cases  $N = 150$  and  $N = 300$ , and/or adopting different randomization strategies. Overall, the point approximations of the “true” values are quite reasonable: the fluctuations around the correct values are of the order of a few tons (or fractions of tons), representing small percentages of the cube weight, even in

case strong discretization and large dependence are imposed.

### 4.3 Clayton case

Here the data have been sampled from the Clayton family. Tables 5 and 6 report the probabilities of rejection of the Null hypothesis that the copula belongs to several families.



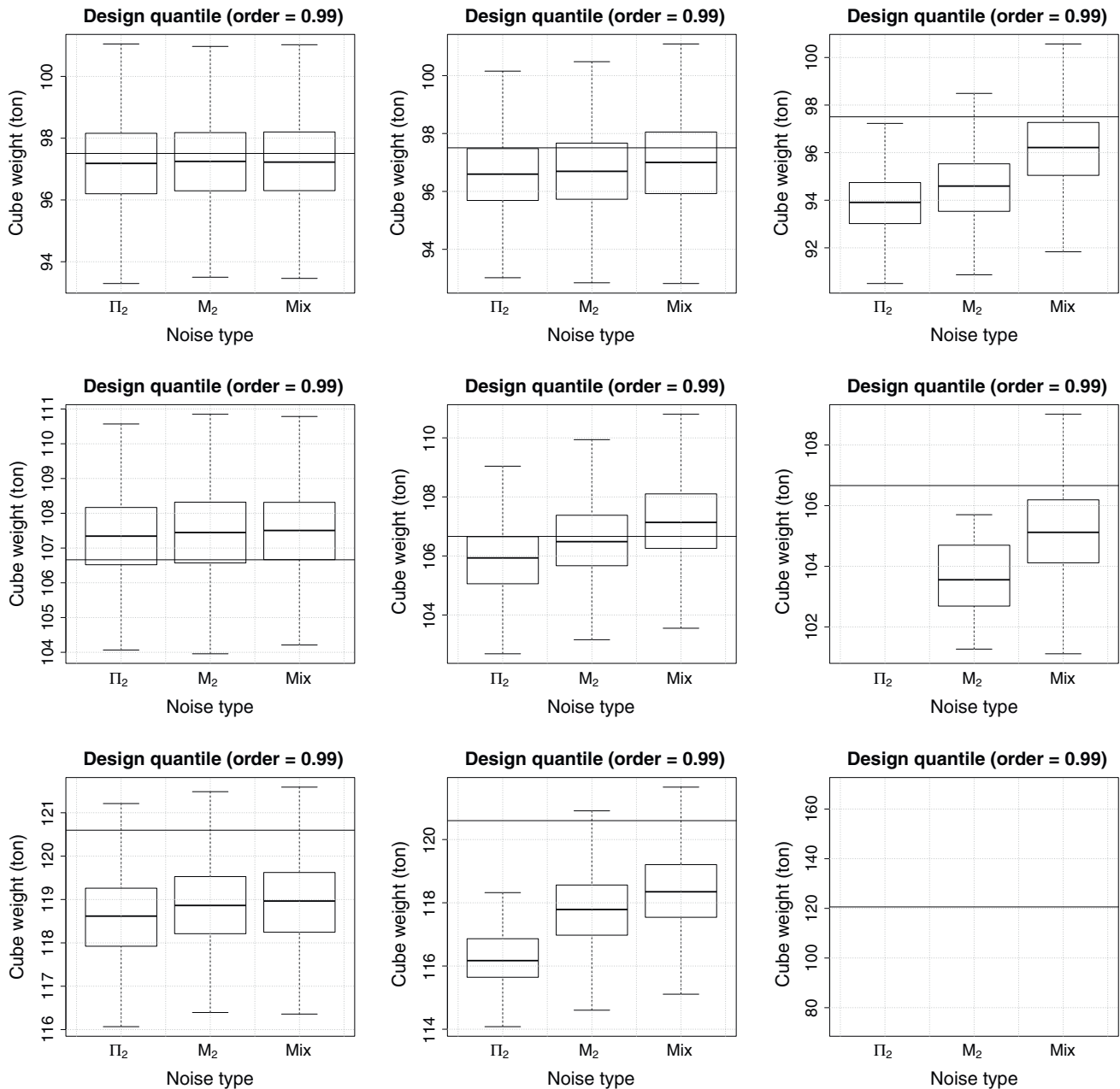
**Fig. 5** “Clayton” case:  $N = 300$ ,  $B = 2500$ . Boxplots of the copula parameter estimates: the horizontal thick lines indicate the “true” values—see text. From top to bottom, the rows correspond, respectively, to the values of the Kendall’s  $\tau = 0.25, 0.5, 0.75$ . From left to right, the columns correspond, respectively, to the following pairs of

height and duration resolutions:  $(\Delta_H, \Delta_D) \in \{(0.01, 0.5), (0.1, 1), (0.5, 3)\}$ . The labels “ $\Pi_2$ ”, “ $M_2$ ”, and “mix” denote the use of an independent, a co-monotone, and a mixed randomization, respectively

In almost all cases, the Goodness-of-Fit test procedures attain values (much) larger than the nominal level (5%) for the Null hypothesis that the copula comes from the Clayton family, independently of the degree of discretization, the value of  $\tau$ , the randomization strategy, and the sample size. In particular, the Clayton model is always rejected in case of a strong discretization. Furthermore, in testing the Null hypothesis that the data are extracted from another fixed

copula family different from the Clayton one, the results are reasonable.

Concerning the estimate of the “true” parameter values, the results are illustrated in Fig. 5. In general, these values are under-estimated (on average), independently of the degree of discretization, the value of  $\tau$ , and the randomization strategy: consequently, a weaker degree of dependence is incorrectly perceived. The only exceptions are as



**Fig. 6** “Clayton” case:  $N = 300$ ,  $B = 2500$ ,  $q = 0.99$ . Boxplots of the cube weight design quantiles estimates: the *horizontal thick lines* indicate the “true” values—see text. From top to bottom, the rows correspond, respectively, to the values of the Kendall’s  $\tau = 0.25, 0.5, 0.75$ . From left to right, the columns correspond,

respectively, to the following pairs of height and duration resolutions:  $(\Delta_H, \Delta_D) \in \{(0.01, 0.5), (0.1, 1), (0.5, 3)\}$ . The labels “ $\Pi_2$ ”, “ $M_2$ ”, and “mix” denote the use of an independent, a co-monotone, and a mixed randomization, respectively

follows: (i) for  $\tau = 0.25$  and weak/moderate discretization; (ii) for  $\tau = 0.5$  and weak discretization. Apparently, no improvements are achieved using larger samples of size  $N = 300$ . It is worth noting that some boxplots are incomplete, since some data are missing: namely, for the combinations  $\{N = 150, \tau = 0.75, (\Delta_H, \Delta_D) = (0.5, 3)\}$ ,  $\{N = 300, \tau = 0.5, (\Delta_H, \Delta_D) = (0.5, 3)\}$ , and  $\{N = 300, \tau = 0.75, (\Delta_H, \Delta_D) = (0.5, 3)\}$ . In these cases, the Null hypothesis that the copula comes from the Clayton family is always rejected over all the  $B$  tests (at a 5% level): in turn, no meaningful estimates of the Clayton copula parameter can be computed, as well as no corresponding design quantiles.

The results concerning the estimate of the design quantiles are presented in Fig. 6. In general, the behavior is the same as the one of the estimates of the copula parameter: viz., an overall under-estimate of the “true” values, with only a few exceptions.

## 5 Discussion and conclusions

The results presented in Sect. 4 may be read under two alternative and complementary perspectives, one of more theoretical nature (1), and the other of more practical relevance (2), as discussed below.

1. On the one hand, apparently, the randomization procedures may not help in identifying the “true” model once the data are made available in a discretized form, e.g. due to (instrumental) adjustments and truncations: in fact, the examples presented above show that in some cases the jittering may provide a valuable tool (i.e., the Gumbel case), whereas in some other cases it may not work at all (i.e., the Clayton case), or only partially (i.e., the Frank case). Roughly speaking, in some cases the features of a copula family which are usually detected by the Goodness-of-Fit test—for instance, the Crámer–von Mises one used in this experiment—are only weakly affected by the discretization (e.g., in the Gumbel case), whereas some other families may be strongly affected (e.g., in the Clayton case). As a consequence, since the underlying “true” copula is unknown, in general it is not possible to trust the randomization procedure as a tool for identifying the original copula in discretized samples.
2. On the other hand, at least concerning the setting adopted in this experiment, apparently, the point estimates of the design values seem to be only weakly affected and spoiled by the discretization mechanism. Clearly, this may represent a good news for practitioners, who generally are interested in the estimation of approximate design values for assessing the risk.

However, in the present framework, the variability of the estimates of the design quantiles introduced by the discretization/randomization procedures is about of the same order as the actual one between the “true” design values computed using the three different copula models.

As a summary, from the partial (yet realistic) simulation results presented above, it seems sensible to conclude that, concerning the identification problem, the randomization procedures outlined in this work may be of little help when discretized data are made available. However, and especially if the level of discretization of the data is weak or moderate, apparently the same procedures may yield fair approximations to the structural risk. In case a strong discretization is at play, underestimates of the true risk have to be expected.

This paper may be of particular interest for coastal engineers, since Weibull marginals like the ones used in this work are widely adopted in practical applications for modeling the main variables playing a significant role in sea storms. Clearly, the results presented above should not be over-generalized: for instance, by using heavy-tailed marginals like Generalized Pareto or Generalized Extreme Value laws.

**Acknowledgements** Helpful discussions with C. Genest (McGill University, Montréal (Québec), Canada), I. Kojadinovic (Université de Pau et des Pays de l’Adour, Pau, France), and C. Sempi (Università del Salento, Lecce, Italy) are gratefully acknowledged. [RP] The support of the Department of Economics, Business, Mathematics and Statistics “Bruno De Finetti” (University of Trieste, Italy), via the project FRA, is acknowledged. [FD] The support of Faculty of Economics and Management, Free University of Bozen-Bolzano, via the project “COCCO”, is acknowledged. [GS] The support of the CRM-CANSSI (Université de Montréal, Montréal (Québec), Canada), where the work originated, is gratefully acknowledged. The support of the CMCC [Centro Euro-Mediterraneo sui Cambiamenti Climatici, Lecce (Italy)] is acknowledged.

## References

- AghaKouchak A, Easterling D, Hsu K, Schubert S, Sorooshian S (eds) (2013) *Extremes in a changing climate*. Springer, Houten
- Bücher A, Kojadinovic I (2016) An overview of nonparametric tests of extreme-value dependence and of some related statistical procedures. In: Dey D, Yan J (eds) *Extreme value modeling and risk analysis: methods and applications*. CRC Press, Boca Raton, pp 377–398
- de Amo E, Díaz-Carrillo M, Fernández-Sánchez J (2012) Characterization of all copulas associated with non-continuous random variables. *Fuzzy Sets Syst* 191:103–112
- De Michele C, Salvadori G, Vezzoli R, Pecora S (2013) Multivariate assessment of droughts: frequency analysis and dynamic return period. *Water Resour Res* 49(10):6985–6994. doi:10.1002/wrcr.20551
- Di Lascio FML, Durante F, Jaworski P (2016) Truncation invariant copulas and a testing procedure. *J Stat Comput Simul* 86:2362–2378. doi:10.1080/00949655.2015.1110820

- Durante F, Sempì C (2016) Principles of copula theory. CRC/Chapman & Hall, Boca Raton
- Durante F, Fernández-Sánchez J, Quesada-Molina JJ, Úbeda-Flores M (2015) Convergence results for patchwork copulas. *Eur J Oper Res* 247(2):525–531
- Genest C, Favre AC (2007) Everything you always wanted to know about copula modeling but were afraid to ask. *J Hydrol Eng* 12(4):347–368
- Genest C, Nešlehová J (2007) A primer on copulas for count data. *Astin Bull* 37(2):475–515
- Genest C, Nešlehová J (2012a) Copula modeling for extremes. In: El-Shaarawi AH, Piegorsch WW (eds) *Encyclopedia of environmental metrics*, 2nd edn. Wiley, Chichester, pp 530–541
- Genest C, Nešlehová J (2012b) Copulas and copula models. In: El-Shaarawi AH, Piegorsch WW (eds) *Encyclopedia of environmental metrics*, 2nd edn. Wiley, Chichester, pp 541–553
- Genest C, Nešlehová J, Ruppert M (2011) Discussion: statistical models and methods for dependence in insurance data. *J Korean Stat Soc* 40(2):141–148
- Genest C, Nešlehová J, Rémillard B (2014) On the empirical multilinear copula process for count data. *Bernoulli* 20(3):1344–1371
- Kojadinovic I, Yan J (2011) A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Stat Comput* 21:17–30
- Kojadinovic I, Yan J, Holmes M (2011) Fast large-sample goodness-of-fit tests for copulas. *Stat Sin* 21:841–871
- Marshall AW (1996) Copulas, marginals, and joint distributions. In: *Distributions with fixed marginals and related topics* (Seattle, WA, 1993) (IMS lecture notes monograph series), vol 28. Institute of Mathematical Statistics, Hayward, pp 213–222
- Nelsen RB (2006) An introduction to copulas, 2nd edn. Springer, New York
- Pappadà R, Perrone E, Durante F, Salvadori G (2016) Spin-off extreme value and Archimedean copulas for estimating the bivariate structural risk. *Stoch Environ Res Risk Assess* 30:327–342. doi:10.1007/s00477-015-1103-8
- Salvadori G, De Michele C, Kottegoda N, Rosso R (2007) *Extremes in nature. An approach using copulas, water science and technology library*. Springer, Dordrecht
- Salvadori G, Tomasicchio GR, D’Alessandro F (2014) Practical guidelines for multivariate analysis and design in coastal and offshore engineering. *Coast Eng* 88:1–14. doi:10.1016/j.coastaleng.2014.01.011
- Salvadori G, Durante F, Tomasicchio GR, D’Alessandro F (2015) Practical guidelines for the multivariate assessment of the structural risk in coastal and off-shore engineering. *Coast Eng* 95:77–83. doi:10.1016/j.coastaleng.2014.09.007
- Sklar A (1959) Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ Inst Stat Univ Paris* 8:229–231
- Straub D (2014) Engineering risk assessment. In: Klüppelberg C, Straub D, Welpé IM (eds) *Risk—a multidisciplinary introduction*. Springer, Cham