

# The checkerboard copula and dependence concepts

Liyuan Lin\*      Ruodu Wang<sup>†</sup>      Ruixun Zhang<sup>‡</sup>      Chaoyi Zhao<sup>§</sup>

31st January 2025

## Abstract

We study the problem of choosing the copula when the marginal distributions of a random vector are not all continuous. Inspired by four motivating examples including simulation from copulas, stress scenarios, co-risk measures, and dependence measures, we propose to use the checkerboard copula, that is, intuitively, the unique copula with a distribution that is as uniform as possible within regions of flexibility. We show that the checkerboard copula has the largest Shannon entropy, which means that it carries the least information among all possible copulas for a given random vector. Furthermore, the checkerboard copula preserves the dependence information of the original random vector, leading to two applications in the context of diversification penalty and impact portfolios. The numerical and empirical results illustrate the benefits of using the checkerboard copula in the calculation of co-risk measures.

**Keywords:** Orthant dependence; positive and negative association; Shannon entropy; Co-VaR; simulation

## 1 Introduction

The copula theory has been actively studied over the past few decades with many applications in statistics, finance, engineering, and the natural sciences; for an introduction, see the monographs of [Nelsen \(2006\)](#) and [Joe \(2014\)](#).

It is well known through Sklar's theorem ([Nelsen \(2006, Theorem 2.10.9\)](#)) that the copula of a random vector is unique if and only if it has continuous marginal distributions. However, when the marginals are non-continuous, the uniqueness of the copula no longer holds. Discrete marginals

---

\*Department of Statistics and Actuarial Science, University of Waterloo, Canada. [1891lin@uwaterloo.ca](mailto:1891lin@uwaterloo.ca)

<sup>†</sup>Department of Statistics and Actuarial Science, University of Waterloo, Canada. [wang@uwaterloo.ca](mailto:wang@uwaterloo.ca)

<sup>‡</sup>School of Mathematical Sciences, Center for Statistical Science, and National Engineering Laboratory for Big Data Analysis and Applications, Peking University, China. [zhangruixun@pku.edu.cn](mailto:zhangruixun@pku.edu.cn)

<sup>§</sup>School of Mathematical Sciences, Peking University, China; Sloan School of Management and Laboratory for Financial Engineering, MIT, United States. [zhaochaoyi@pku.edu.cn](mailto:zhaochaoyi@pku.edu.cn)

are common in empirical studies, as the collected data is often discrete. Several works, including [Marshall \(1996\)](#), [Carley \(2002\)](#), [Perrone et al. \(2019\)](#), and [Geenens \(2020\)](#), discuss the dependence structure of discrete data through copulas. [Genest and Nešlehová \(2007\)](#) discussed difficulties in identifying copulas for discrete distributions. The purpose of this paper is to understand whether it is possible to identify a canonical copula for a random vector in some sense if it does not have continuous marginal distributions.

To answer this question, we seek inspiration from four applications. Let  $\mathbf{X} = (X_1, \dots, X_d)$  be a  $d$ -dimensional random vector with  $d \geq 2$ , which may have non-unique copulas. Denote by  $\mathcal{C}_{\mathbf{X}}$  the set of all copulas of  $\mathbf{X}$ . For a random variable  $X$ , its probability integral transform  $U$  is a uniform random variable on  $[0, 1]$  satisfying  $F^{-1}(U) = X$  almost surely (a.s.), where  $F$  is the distribution function of  $X$  and  $F^{-1}$  is the quantile function of  $X$ . Let  $(U_1, \dots, U_d)$  be any vector of probability integral transforms of  $X_1, \dots, X_d$  with a joint distribution  $C$ ; certainly,  $C$  is a copula of  $\mathbf{X}$ . All random variables live in an atomless probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ .

1. **Simulating from the copula of  $\mathbf{X}$ .** One of the most popular applications of copulas in finance is to model default correlation, as famously done by [Li \(2000\)](#); see [McNeil et al. \(2015\)](#) for discussions. In such applications, one needs to simulate from the copula of  $\mathbf{X}$ , where  $\mathbf{X}$  may have non-continuous marginal distributions (e.g., losses from default events). Assume that we can simulate  $\mathbf{X}$ , and we also have knowledge of all marginal distributions of  $\mathbf{X}$ . How can we find a reasonable copula  $C \in \mathcal{C}_{\mathbf{X}}$  to simulate from, that is determined only by  $\mathbf{X}$  but not by any particular modeling choices (such as the Gaussian copula)?
2. **Stressing the distribution of  $\mathbf{X}$ .** In sensitivity analysis and risk management, it is often necessary to stress, or distort, the distribution of  $\mathbf{X}$  to obtain post-stress distributions. In the stressing mechanisms studied by [Millosovich et al. \(2024\)](#), one needs to find a stressed probability measure  $Q_1$  by using  $dQ_1/d\mathbb{P} = g(U_1)$  for a non-negative increasing function  $g$  with  $\int_0^1 g(u)du = 1$ , such as  $g(u) = 2u$ . The simple interpretation of  $Q_1$  is to gradually increase the weight of realizations  $\omega \in \Omega$  at which  $X_1$  is large. Similarly, one can simultaneously stress all components of  $\mathbf{X}$  by considering a measure  $Q$  such that  $dQ/d\mathbb{P} = (1/d) \sum_{i=1}^d g_i(U_i)$  or  $dQ/d\mathbb{P} = c \prod_{i=1}^d g_i(U_i)$  with a normalizing constant  $c > 0$  ( $c = 1$  if  $U_1, \dots, U_d$  are independent), where  $g_i$  are non-negative increasing functions with  $\int_0^1 g_i(u)du = 1$ . If we are only interested in the post-stress distribution  $\hat{F}_1^{Q_1}$  of  $X_1$  under  $Q_1$ , the choice of the copula  $C \in \mathcal{C}_{\mathbf{X}}$  is irrelevant. However, the choice of the copula  $C \in \mathcal{C}_{\mathbf{X}}$  matters for the distribution  $\hat{F}_i^Q$  of  $X_i$  under  $Q$ , as well as for the distribution  $\hat{F}_i^{Q_1}$  of  $X_i$  under  $Q_1$ .

3. **Computing a co-risk measure.** Co-risk measures (e.g., [Adrian and Brunnermeier \(2016\)](#)) are calculated for the conditional distribution of a random variable  $X_2$  given some event related to  $X_1$ . A classic example is the Marginal Expected Shortfall (Marginal ES) at level  $p \in (0, 1)$ , which is defined as, assuming that  $X_1$  is continuously distributed,

$$\rho(X_2|X_1) := \mathbb{E}[X_2|X_1 > F_1^{-1}(p)] = \mathbb{E}[X_2|U_1 > p]. \quad (1)$$

Generally,  $\rho$  is the mean of  $X_2$  given a (not necessarily unique)  $p$ -tail event of  $X_1$  in the sense of [Wang and Zitikis \(2021\)](#). This risk measure  $\rho$  does not depend on the choice of  $C \in \mathcal{C}_{\mathbf{X}}$  if  $X_1$  is continuously distributed ( $p$ -tail event is unique a.s.); however, it may depend on  $C \in \mathcal{C}_{\mathbf{X}}$  if  $X_1$  has some points of mass. Other co-risk measures, such as CoVaR ([Adrian and Brunnermeier, 2016](#)), also face the same issue.

4. **Maintaining dependence measures.** Kendall's  $\tau$  is a dependence measure defined based on concordance. For a bivariate random vector  $(X, Y)$ , its Kendall's  $\tau$  is defined as

$$\tau(X, Y) = \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0),$$

where  $(X_1, Y_1)$  and  $(X_2, Y_2)$  are two independent copies of  $(X, Y)$ . If  $(X, Y)$  has continuous marginals, we have further

$$\tau(X, Y) = 4 \int_{[0,1]^2} C(\mathbf{u}) dC(\mathbf{u}) - 1, \quad (2)$$

where  $C$  is the copula for  $(X, Y)$ . If  $(X, Y)$  has non-continuous marginals, (2) does not hold for every  $C \in \mathcal{C}_{(X,Y)}$ . It is natural to ask which copula  $C \in \mathcal{C}_{(X,Y)}$  conveys the property in the case of continuous marginals. We can also consider similar applications for other concordance-based dependence measures such as Spearman's  $\rho$ .

All of the above contexts point to the question of choosing a good copula  $C \in \mathcal{C}_{\mathbf{X}}$ . [Nešlehová \(2004\)](#) discusses similar applications for the choice of copulas. The main idea of [Nešlehová \(2004\)](#) is to extend the subcopula, which is the part of the copula that uniquely determined by joint distribution, to capture the dependence of the original random vector analogous to the case with continuous marginals. In this paper, we address this problem from the view of probability integral transformation. We first offer a new characterization of all copulas of a given random vector by constructing all probability integral transformations in Section 2 in Theorem 1. In Section 3, we give some intuitive and heuristic arguments for the questions above, leading to the proposal of

using the checkerboard copula, that is, the unique copula of  $\mathbf{X}$  that is as uniform as possible in regions where the copulas of  $\mathbf{X}$  are not uniquely determined, formally defined in Definition 1. The checkerboard copula is the same as the standard extension proposed by Nešlehová (2004), which has been shown to preserve quadrant dependence, tail dependence, and weak convergence results of the joint distribution. Although the arguments in Section 3 are heuristic, the use of the checkerboard copula indeed has a theoretical justification, which we present in Section 4. The checkerboard copula has the maximum Shannon entropy among all possible copulas of  $\mathbf{X}$ , as shown in Theorem 2. In Section 5, we show in Theorem 3 that the checkerboard copula preserves various dependence concepts that are satisfied by  $\mathbf{X}$ . This result is intuitive, but the proof requires serious technical analysis. We discuss two applications of our results in diversification penalty and induced order statistics in Section 6. Section 7 uses numerical and empirical experiments to demonstrate that the checkerboard copula is a convenient and natural choice that can produce reliable results. Section 8 concludes the paper.

## 2 Copulas for a discrete random vector

Let  $d \geq 2$  be an integer and  $[d] = \{1, \dots, d\}$ . All inequalities are interpreted component-wise when applied to vectors. All random variables live in an atomless probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $\mathbf{X} = (X_1, \dots, X_d)$  be a  $d$ -dimensional random vector,  $F_1, \dots, F_d$  be the marginal distributions of  $\mathbf{X}$ , and  $\text{Ran}(F_i)$  be the range of  $F_i$  for  $i \in [d]$ . By Sklar's theorem, the copula of  $\mathbf{X}$  is uniquely determined on  $\text{Ran}(F_1) \times \dots \times \text{Ran}(F_d)$  but undetermined in other regions. Therefore, when the marginal distribution  $F_i$  is not continuous for some  $i \in [d]$ , the copula of  $\mathbf{X}$  may not be unique. In this section, we give a concrete representation for any copulas of  $\mathbf{X}$ .

We start with the observation that, if a random variable  $X$  is continuously distributed, the random variable

$$U_X := F_X(X)$$

will be uniformly distributed over  $[0, 1]$ , where  $F_X$  is the cumulative distribution function of  $X$ . More generally, regardless of whether  $X$  is continuously distributed, we can define its probability integral transform

$$U_X := F_X(X-) + V_X(F_X(X) - F_X(X-)), \tag{3}$$

where  $F_X(x-) = \lim_{y \uparrow x} F_X(x) = \mathbb{P}(X < x)$  for  $x \in \mathbb{R}$  and  $V_X \sim \text{U}[0, 1]$  is independent of  $X$ ,

assumed to exist.<sup>1</sup> The probability integral transform  $U_X$  satisfies  $U_X \sim \text{U}[0, 1]$  and  $F_X^{-1}(U_X) = X$  a.s. (see e.g., [Rüschendorf \(2013, Proposition 1.3\)](#)). Therefore, the probability integral transform (3) converts any random variable  $X$  to a  $\text{U}[0, 1]$  distributed random variable  $U_X$  using  $V_X$ .

We extend this idea to the case of a random vector  $\mathbf{X}$ . Let  $\mathbf{V} = (V_1, \dots, V_d)$  be a random vector with  $\text{U}[0, 1]$  marginals such that  $V_i$  is independent of  $X_i$  for each  $i \in [d]$ . Denote the set of such  $\mathbf{V}$  by  $\mathcal{V}_{\mathbf{X}}$ . Similar to (3), let us define the probability integral transform for  $\mathbf{X} = (X_1, \dots, X_d)$ :

$$U_i := F_i(X_i-) + V_i(F_i(X_i) - F_i(X_i-)), \quad i \in [d]. \quad (4)$$

It immediately follows that  $U_i \sim \text{U}[0, 1]$  and  $F_i^{-1}(U_i) = X_i$  a.s.. Therefore,  $\mathbf{U} = (U_1, \dots, U_d)$  is a random vector with uniform marginals. This transformation comes from randomized hypothesis tests ([Ferguson, 1967, Section 5.3](#)) and has been applied in various contexts; see, e.g., [Moore and Spruill \(1975\)](#), [Nešlehová \(2007\)](#), [Rüschendorf \(1981, 2009, 2013\)](#), and [Faugeras \(2017\)](#).

Let  $C_{\mathbf{X}}^{\mathbf{V}}$  be the copula of  $\mathbf{U}$ . Because  $F_i^{-1}(U_i) = X_i$  a.s. for each  $i \in [d]$ , we have

$$C_{\mathbf{X}}^{\mathbf{V}}(F_1(x_1), \dots, F_d(x_d)) = \mathbb{P}(U_1 \leq F_1(x_1), \dots, U_d \leq F_d(x_d)) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d)$$

for any  $(x_1, \dots, x_d) \in \mathbb{R}^d$ . Hence,  $C_{\mathbf{X}}^{\mathbf{V}}$  is a copula of  $\mathbf{X}$ .

According to (4), the copula  $C_{\mathbf{X}}^{\mathbf{V}}$  is determined by the joint distribution of  $(\mathbf{X}, \mathbf{V})$ . In particular, the copula  $C_{\mathbf{X}}^{\mathbf{V}}$  does not depend on the choice of  $V_i$  for  $i$  such that  $X_i$  is continuously distributed because, for these  $i$ ,  $U_i$  in (4) is a.s. equal to  $F_i(X_i)$ . While for  $i$  such that  $X_i$  is discrete,  $V_i$  does have an impact on the copula  $C_{\mathbf{X}}^{\mathbf{V}}$ .

In general, the choice of  $\mathbf{V} \in \mathcal{V}_{\mathbf{X}}$  for constructing the copula  $C_{\mathbf{X}}^{\mathbf{V}}$  may not be unique. This is because  $\mathcal{V}_{\mathbf{X}}$  allows two types of dependence that might be present in the construction of  $\mathbf{V}$ : First, the components of  $\mathbf{V}$  may be mutually dependent. Second,  $V_i$  may depend on  $X_j$  for  $i \neq j$ . Naturally, a different choice of  $\mathbf{V} \in \mathcal{V}_{\mathbf{X}}$  often leads to a different copula  $C_{\mathbf{X}}^{\mathbf{V}}$ ; see the following example.

**Example 1.** Assume that  $d = 2$ ,  $X_1$  is a constant, and  $X_2$  is continuously distributed. It is well known that any copula is a copula of  $\mathbf{X}$  in this case. For instance, by choosing  $V_1$  to be independent of  $X_2$ ,  $C_{\mathbf{X}}^{\mathbf{V}}$  is the independence copula, and by choosing  $V_1 = F_2(X_2)$ ,  $C_{\mathbf{X}}^{\mathbf{V}}$  is the comonotonic copula.

The following result says that all copulas of  $X$  can be realized by some  $C_{\mathbf{X}}^{\mathbf{V}}$ . Hence, (4) gives a stochastic representation for any copula of  $\mathbf{X}$ . The representation is quite intuitive, but we did

---

<sup>1</sup>This assumption is safe as we are interested in distributional properties, and we can extend the probability space to include such independent  $V_X$ , if necessary.

not find it in the literature, so we provide a self-contained proof.

**Theorem 1.** *Let  $\mathbf{X}$  be a random vector such that there exists a continuously distributed random variable independent of  $\mathbf{X}$ . A copula  $C$  is a copula of  $\mathbf{X}$  if and only if  $C = C_{\mathbf{X}}^{\mathbf{V}}$  for some  $\mathbf{V} \in \mathcal{V}_{\mathbf{X}}$ .*

*Proof.* We have seen that  $C_{\mathbf{X}}^{\mathbf{V}}$  is a copula of  $\mathbf{X}$ . It suffices to show the “only if” statement. Let  $C$  be a copula of  $\mathbf{X}$ , take  $\mathbf{U}' = (U'_1, \dots, U'_d) \sim C$ , and write  $\mathbf{X}' = (F_1^{-1}(U'_1), \dots, F_d^{-1}(U'_d))$ . Because  $C$  is a copula of  $\mathbf{X}$ , for  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ , we have

$$\begin{aligned} \mathbb{P}(\mathbf{X} \leq \mathbf{x}) &= C(F_1(x_1), \dots, F_d(x_d)) = \mathbb{P}(U'_1 \leq F_1(x_1), \dots, U'_d \leq F_d(x_d)) \\ &= \mathbb{P}(F_1^{-1}(U'_1) \leq x_1, \dots, F_d^{-1}(U'_d) \leq x_d) = \mathbb{P}(\mathbf{X}' \leq \mathbf{x}). \end{aligned}$$

Hence,  $\mathbf{X} \stackrel{d}{=} \mathbf{X}'$ . Take  $\mathbf{U}^* = (U_1^*, \dots, U_d^*)$  such that  $(\mathbf{X}, \mathbf{U}^*) \stackrel{d}{=} (\mathbf{X}', \mathbf{U}')$ , and we then have  $\mathbf{X} = (F_1^{-1}(U_1^*), \dots, F_d^{-1}(U_d^*))$  a.s. Furthermore, take  $V' \sim U[0, 1]$  which is independent of  $(\mathbf{X}, \mathbf{U}^*)$ . The existence of  $\mathbf{U}^*$  and  $V'$  is guaranteed by the assumption of the existence of a continuously distributed random variable independent of  $\mathbf{X}$ . For  $i \in [d]$ , let  $\mathbf{V} = (V_1, \dots, V_d)$  be given by

$$V_i = \frac{U_i^* - F_i(X_{i-})}{F_i(X_i) - F_i(X_{i-})} \mathbf{1}_{\{F_i(X_i) > F_i(X_{i-})\}} + V' \mathbf{1}_{\{F_i(X_i) = F_i(X_{i-})\}}.$$

Fix  $i \in [d]$  below. Let  $D_i$  be the set of discontinuity points of  $F_i$ . Note that for  $x \in D_i$ , we have

$$\mathbb{P}(U_i^* \in [F_i(x-), F_i(x)] | X_i = x) = 1 \quad \text{and} \quad \mathbb{P}(U_i^* \in [F_i(x-), F_i(x)] | X_i \neq x) = 0.$$

Because  $U_i^*$  is uniformly distributed over  $[0, 1]$ ,  $U_i^*$  is uniform on  $[F_i(x-), F_i(x)]$  conditional on  $X_i = x \in D_i$ . Thus,

$$\mathbb{P}(U_i^* \leq u | X_i = x) = \frac{u - F_i(x-)}{F_i(x) - F_i(x-)}, \quad u \in [F_i^{-1}(x-), F_i^{-1}(x)].$$

Therefore, for  $u \in [0, 1]$ ,

$$\begin{aligned} \mathbb{P}(V_i \leq u | X_i) &= \mathbb{P}\left(\frac{U_i^* - F_i(X_{i-})}{F_i(X_i) - F_i(X_{i-})} \leq u \mid X_i\right) \mathbf{1}_{\{X_i \in D_i\}} + \mathbb{P}(V' \leq u) \mathbf{1}_{\{X_i \notin D_i\}} \\ &= \mathbb{P}(U_i^* \leq u(F_i(X_i) - F_i(X_{i-})) + F_i(X_{i-}) | X_i) \mathbf{1}_{\{X_i \in D_i\}} + u \mathbf{1}_{\{X_i \notin D_i\}} \\ &= u \mathbf{1}_{\{X_i \in D_i\}} + u \mathbf{1}_{\{X_i \notin D_i\}} = u. \end{aligned}$$

Hence,  $V_i$  follows  $U[0, 1]$  and is independent of  $X_i$ . Note that, by the construction,  $U_i^*$ ,  $V_i$ , and  $X_i$  satisfy  $U_i^* = F_i(X_{i-}) + V_i(F_i(X_i) - F_i(X_{i-}))$  a.s., and hence  $\mathbf{U}^* \sim C_{\mathbf{X}}^{\mathbf{V}}$ . This shows  $C = C_{\mathbf{X}}^{\mathbf{V}}$ .  $\square$

Theorem 1 implies  $\mathcal{C}_{\mathbf{X}} = \{C_{\mathbf{X}}^{\mathbf{V}} : \mathbf{V} \in \mathcal{V}_{\mathbf{X}}\}$ , providing a characterization of copulas in a stochastic form. Note that  $\mathcal{C}_{\mathbf{X}}$  is a singleton if and only if all marginal distributions of  $\mathbf{X}$ ,  $F_1, \dots, F_d$ , are continuous functions. The “if” direction of Theorem 1 in the case  $d = 2$  is shown by Nešlehová (2007, Proposition 4). The characterization of copulas in an analytical form is provided by de Amo et al. (2017).

### 3 Motivating arguments for the checkerboard copula

Theorem 1 gives the entire class of copulas for  $\mathbf{X}$ . We now consider which  $\mathbf{V} \in \mathcal{V}_{\mathbf{X}}$  can answer the four motivating questions in Section 1, which all point to the same unique choice of  $\mathbf{V} \in \mathcal{V}_{\mathbf{X}}$ .

1. **Simulating from the copula of  $\mathbf{X}$ .** A natural approach to simulating from the copula of  $\mathbf{X}$  with some atoms in the marginal distributions is by first simulating a pair of  $(\mathbf{X}, \mathbf{V})$ , and then applying the probability integral transform using (4). Theorem 1 shows that all copulas of  $\mathbf{X}$  can be simulated this way. For this purpose, the simplest and most natural choice of  $\mathbf{V}$  is  $\mathbf{V} \sim U([0, 1]^d)$  which is independent of  $\mathbf{X}$ . In fact, we could not think of an argument against the use of this particular  $\mathbf{V}$  in the context of simulation.
2. **Stressing the distribution of  $\mathbf{X}$ .** To understand how the choice of  $\mathbf{V}$  affects the stressed distribution of  $X_2$ , we look at the simple example in Example 1 with  $g(u) = 2u$ . Choosing  $V_1$  independent of  $X_2$  would lead to  $\hat{F}_2^{Q_1} = F_2$ , whereas choosing  $V_1 = F_2(X_2)$  would lead to  $\hat{F}_2^{Q_1} = (F_2)^2$ . Because we are interested in the effect of stressing  $X_1$  on  $X_2$ , and  $X_1$  is a constant in this example, it is natural to choose a  $V_1$  that affects the distribution of  $X_2$  minimally, which is achieved when  $V_1$  is independent of  $X_2$ . Translating this argument into the general  $d$ -dimensional setting suggests choosing  $\mathbf{V} \sim U([0, 1]^d)$  independent of  $\mathbf{X}$ .
3. **Computing a co-risk measure.** To understand how the choice of  $\mathbf{V}$  affects the value of the co-risk measure, we again look at Example 1. We have  $\rho(X_2|X_1) = \mathbb{E}[X_2]$  if  $V_1$  is independent of  $X_2$ , and  $\rho(X_2|X_1) = \text{ES}_p(X_2)$  if  $V_1 = F_2(X_2)$ , where  $\text{ES}_p(X_2) = \mathbb{E}[X_2|U_2 > p]$  is the Expected Shortfall of  $X_2$  at level  $p$ . The interpretation of  $\rho$  as the mean of  $X_2$  on a tail event of  $X_1$  suggests that it is natural to choose  $V_1$  independent of  $X_2$ , because  $X_1$  is a constant and its tail event should not affect  $X_2$ .
4. **Maintaining dependence measures.** It has been shown in Denuit and Lambert (2005), Nešlehová (2007), and Genest and Nešlehová (2007) that (2) holds for the copula introduced by

$\mathbf{V} \sim \text{U}([0, 1]^2)$ . This property leads to extensions of dependence measures in the multivariate case based on this copula; see, e.g., [Mesfioui and Quesy \(2010\)](#) and [Genest et al. \(2013\)](#).

In all the considerations above,  $\mathbf{V} \sim \text{U}([0, 1]^d)$  independent of  $\mathbf{X}$  appears to be a good choice. Let us denote this by  $\mathbf{V}_{\mathbf{X}}^{\perp}$  and the corresponding copula by  $C_{\mathbf{X}}^{\perp}$ , where  $\perp$  reflects that independence is used twice to construct  $\mathbf{V}$  (within components of  $\mathbf{V}$  and between  $\mathbf{V}$  and  $\mathbf{X}$ ). From the four motivating examples above, the choice of the particular copula  $C_{\mathbf{X}}^{\perp}$  is natural and has several unique features. This choice has been known as the checkerboard copula.

**Definition 1.** The copula  $C_{\mathbf{X}}^{\perp}$  is called the *checkerboard copula* of  $\mathbf{X}$ .

The copula  $C_{\mathbf{X}}^{\perp}$  is also called the multilinear extension copula of  $\mathbf{X}$ ; see [Genest et al. \(2017\)](#) for its properties, its empirical process, and a history. One notable property is that  $X_1, \dots, X_d$  are independent if and only if  $C_{\mathbf{X}}^{\perp}$  is the independence copula.

The rest of the paper focuses on the properties and applications of the checkerboard copula.

## 4 Entropy maximization

Given the natural choice of  $C_{\mathbf{X}}^{\perp}$  in the applications in Section 3, it should have some unique properties within the class  $\mathcal{C}_{\mathbf{X}}$ . The applications seem to suggest that  $C_{\mathbf{X}}^{\perp}$  relies less on external information compared to other choices of  $\mathbf{V}$ . Such consideration is typically studied via entropy. Indeed, as argued by [Jaynes \(1957\)](#), the maximum-entropy distribution should be the only unbiased choice given available information. If a copula  $C$  has a density function  $c$ , then its Shannon (differential) entropy is defined as

$$H(C) = - \int_{[0,1]^d} c(\mathbf{u}) \log c(\mathbf{u}) d\mathbf{u}.$$

One problem with the above formulation is that a copula  $C$  often does not have a density. We set  $H(C) = -\infty$  if  $C$  does not have a density, which is intuitive and can be seen as a limiting case; see [Koliander et al. \(2016\)](#) for a discussion on the definition of entropy for singular distributions.

*Remark 1.* By definition,  $H(C) = -D_{\text{KL}}(P_C \| P_L)$ , where  $D_{\text{KL}}(P_C \| P_L)$  is the KL divergence between the probability measure  $P_C$  with distribution function  $C$  and the Lebesgue measure  $P_L$  on  $[0, 1]^d$ . Since the KL divergence quantifies the similarity between  $P_C$  and  $P_L$  via the likelihood ratio  $dP_C/dP_L$ , being singular is the extreme form of non-similarity in terms of likelihood ratio. Therefore, it is natural to set  $D_{\text{KL}}(P_C \| P_L)$  as  $\infty$  whenever  $P_C$  is not absolutely continuous with respect to  $P_L$ . This is a standard approach in the literature on KL divergence and differential entropy;



see e.g., [Csiszár \(1975\)](#) and [Cover and Thomas \(1991\)](#). Hence, to keep the same intuition and consistency with the KL divergence, we set  $H(C) = -\infty$  when  $C$  does not have density.

However, even the checkerboard copula  $C_{\mathbf{X}}^{\perp}$  may not have a density if the distribution of  $\mathbf{X}$  has some singular continuous part. This issue may be solved by considering other measures of information, but for now, let us stick to the Shannon entropy, which is the most popular notion in information theory. We would like to compare  $H(C_{\mathbf{X}}^{\perp})$  with  $H(C)$  for  $C \in \mathcal{C}_{\mathbf{X}}$ , or equivalently,  $H(C_{\mathbf{X}}^{\mathbf{V}})$  for other choices of  $\mathbf{V} \in \mathcal{V}_{\mathbf{X}}$ . The main result of this section is to show that  $H(C_{\mathbf{X}}^{\perp})$  has the largest entropy among all other choices.

**Theorem 2.** *For  $C \in \mathcal{C}_{\mathbf{X}}$ , we have  $H(C_{\mathbf{X}}^{\perp}) \geq H(C)$ .*

The proof of [Theorem 2](#) essentially boils down to showing the following lemma, which states that the density of the checkerboard copula can be expressed as the conditional expectation for the density of other possible copulas in  $\mathcal{C}_{\mathbf{X}}$ . From this lemma and Jensen's inequality, [Theorem 2](#) follows.

**Lemma 1.** *For  $C \in \mathcal{C}_{\mathbf{X}}$ , if the density  $c$  of  $C$  exists, then the density  $c^{\perp}$  of  $C_{\mathbf{X}}^{\perp}$  exists. Moreover,  $c^{\perp}(\mathbf{U}) = \mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}]$ , where  $\mathbf{U} = (U_1, \dots, U_d) \sim \mathbf{U}([0, 1]^d)$ ,  $\hat{\mathbf{X}} = (F_1^{-1}(U_1), \dots, F_d^{-1}(U_d))$ , and  $F_1, \dots, F_d$  are the marginals of  $\mathbf{X}$ .*

*Proof.* Since  $\mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}]$  is  $\sigma(\hat{\mathbf{X}})$ -measurable, there exists a function  $f : \mathbb{R}^d \rightarrow [0, 1]$  such that  $f(\hat{\mathbf{X}}) = \mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}]$  (in the almost sure sense). Let  $c^{\perp}$  be a function  $[0, 1]^d \rightarrow [0, 1]$  defined as  $c^{\perp}(\mathbf{u}) = f(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))$  for any  $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$ . We claim that  $c^{\perp}$  is the density of  $C_{\mathbf{X}}^{\perp}$ . This claim implies that  $c^{\perp}$  exists and  $c^{\perp}(\mathbf{U}) = \mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}]$ .

To prove this claim, let  $\mathbf{U}^c \sim C$ ,  $\mathbf{U}^{\perp} \sim C_{\mathbf{X}}^{\perp}$ , and  $R = \times_{i=1}^d \text{Ran}(F_i)$ . We first show that  $\int_A c^{\perp}(\mathbf{u})d\mathbf{u} = \mathbb{P}(\mathbf{U}^{\perp} \in A)$  for the following two types of the set  $A$ .

- (i) Let  $A = \times_{i=1}^d [0, a_i]$  with  $\mathbf{a} = (a_1, \dots, a_d) \in R$ . We have  $\mathbf{1}_{\{\mathbf{U} \leq \mathbf{a}\}} = \mathbf{1}_{\{\hat{\mathbf{X}} \leq (F_1^{-1}(a_1), \dots, F_d^{-1}(a_d))\}}$ . Therefore,

$$\begin{aligned} \int_A c^{\perp}(\mathbf{u})d\mathbf{u} &= \int_{\times_{i=1}^d [0, a_i]} f(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) du_1 \cdots du_d \\ &= \mathbb{E} \left[ f(\hat{\mathbf{X}}) \mathbf{1}_{\{\mathbf{U} \leq \mathbf{a}\}} \right] \\ &= \mathbb{E} \left[ \mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}] \mathbf{1}_{\{\hat{\mathbf{X}} \leq (F_1^{-1}(a_1), \dots, F_d^{-1}(a_d))\}} \right] \\ &= \mathbb{E} [c(\mathbf{U}) \mathbf{1}_{\{\mathbf{U} \leq \mathbf{a}\}}] = \mathbb{P}(\mathbf{U}^c \leq \mathbf{a}) = \mathbb{P}(\mathbf{U}^{\perp} \leq \mathbf{a}), \end{aligned}$$

where the last equality holds because

$$\mathbb{P}(\mathbf{U}^c \leq \mathbf{a}) = \mathbb{P}(\mathbf{X} \leq (F_1^{-1}(a_1), \dots, F_d^{-1}(a_d))) = \mathbb{P}(\mathbf{U}^\perp \leq \mathbf{a}).$$

This further implies that  $\int_A c^\perp(\mathbf{u})d\mathbf{u} = \mathbb{P}(\mathbf{U}^\perp \in A)$  for any  $A = \times_{i=1}^d A_i$  such that  $A_i \in \{[0, a_i] : a_i \in \text{Ran}(F_i)\} \cup \{(F_i(x_i-), F_i(x_i)) : x_i \text{ is a discontinuity point of } F_i\}$  for  $i \in [d]$ .

- (ii) Let  $A = (\times_{i=1}^k [0, a_i]) \times (\times_{j=k+1}^d (s_j, t_j])$  with  $k \in \{0, 1, \dots, d\}$  such that  $a_i \in \text{Ran}(F_i)$  for  $i \in [k]$  and  $(s_j, t_j] \cap \text{Ran}(F_j) = \emptyset$  for  $j \in [d] \setminus [k]$ . For  $j \in [d] \setminus [k]$ , denote by  $x_j = F_j^{-1}(s_j)$ , and thus  $(s_j, t_j] \subseteq (F_j(x_j-), F_j(x_j))$ . By the definition of  $c^\perp$ , for fixed  $u_i \in [0, a_i]$  and  $i \in [k]$ ,  $c^\perp(u_1, \dots, u_k, v_{k+1}, \dots, v_d)$  is a constant for all  $(v_{k+1}, \dots, v_d) \in \times_{j=k+1}^d (F_j(x_j-), F_j(x_j))$ . Therefore, let  $B = (\times_{i=1}^k [0, a_i]) \times (\times_{j=k+1}^d (F_j(x_j-), F_j(x_j)))$ , we have

$$\int_A c^\perp(\mathbf{u})d\mathbf{u} = \left( \prod_{j=k+1}^d \frac{t_j - s_j}{F_j(x_j) - F_j(x_j-)} \right) \int_B c^\perp(\mathbf{u})d\mathbf{u}.$$

Let  $\mathbf{V} = (V_1, \dots, V_d) \sim \text{U}([0, 1]^d)$  be independent of  $\mathbf{X}$ , and for  $j \in [d] \setminus [k]$ , denote by  $s'_j = (s_j - F_j(x_j-))/(F_j(x_j) - F_j(x_j-))$  and  $t'_j = (t_j - F_j(x_j-))/(F_j(x_j) - F_j(x_j-))$ . Hence,

$$\prod_{j=k+1}^d \frac{t_j - s_j}{F_j(x_j) - F_j(x_j-)} = \mathbb{P}(V_j \in (s'_j, t'_j] \text{ for all } j \in [d] \setminus [k]).$$

In addition, by (i), we can get

$$\int_B c^\perp(\mathbf{u})d\mathbf{u} = \mathbb{P}(\mathbf{U}^\perp \in B) = \mathbb{P}(X_i \leq F^{-1}(a_i), X_j = x_j \text{ for all } i \in [k], j \in [d] \setminus [k]).$$

Therefore,

$$\begin{aligned} \int_A c^\perp(\mathbf{u})d\mathbf{u} &= \mathbb{P} \left( \bigcap_{j \in [d] \setminus [k]} \{V_j \in (s'_j, t'_j]\} \right) \mathbb{P} \left( \bigcap_{i \in [k], j \in [d] \setminus [k]} \{X_i \leq F^{-1}(a_i), X_j = x_j\} \right) \\ &= \mathbb{P} \left( \bigcap_{i \in [k], j \in [d] \setminus [k]} \{X_i \leq F^{-1}(a_i), X_j = x_j, V_i \in [0, 1], V_j \in (s'_j, t'_j]\} \right) \\ &= \mathbb{P} \left( \mathbf{U}^\perp \in \left( \times_{i=1}^k [0, a_i] \right) \times \left( \times_{j=k+1}^d (s_j, t_j] \right) \right) = \mathbb{P}(\mathbf{U}^\perp \in A). \end{aligned}$$

By the same argument, we have  $\int_A c^\perp(\mathbf{u})d\mathbf{u} = \mathbb{P}(\mathbf{U}^\perp \in A)$  for any  $A = \times_{i=1}^d A_i$  such that

$A_i \in \{[0, a_i] : a_i \in \text{Ran}(F_i)\} \cup \{(s_i, t_i] : (s_i, t_i] \cap \text{Ran}(F_i) = \emptyset\}$  for  $i \in [d]$ .

For any  $\mathbf{a} = (a_1, \dots, a_d) \in [0, 1]^d$ , the region  $\times_{i=1}^d [0, a_i]$  can always be represented by an at most countable disjoint union of regions studied in (i) and (ii). Hence, we can obtain

$$\int_{\times_{i=1}^d [0, a_i]} c^\perp(\mathbf{u}) d\mathbf{u} = \mathbb{P}(\mathbf{U}^\perp \leq \mathbf{a}).$$

This proves our claim that  $c^\perp$  is the density of  $C_{\mathbf{X}}^\perp$ .  $\square$

*Proof of Theorem 2.* If  $H(C) = -\infty$ , there is nothing to show. Hence, it suffices to consider the case that  $C$  has a density, which we denote by  $c$ . By Lemma 1, we have  $c^\perp(\mathbf{U}) = \mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}]$  where  $c^\perp$  is the density of  $C_{\mathbf{X}}^\perp$ ,  $\mathbf{U} = (U_1, \dots, U_d) \sim \mathbf{U}([0, 1]^d)$ , and  $\hat{\mathbf{X}} = (F_1^{-1}(U_1), \dots, F_d^{-1}(U_d))$  with  $F_1, \dots, F_d$  as the marginals of  $\mathbf{X}$ . Define a function  $g(x) = x \log x$  for  $x \in (0, \infty)$ . It is clear that  $g$  is convex. By the fact that  $\mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}] = c^\perp(\mathbf{U})$  and Jensen's inequality, we have

$$H(C_{\mathbf{X}}^\perp) = -\mathbb{E}[g(c^\perp(\mathbf{U}))] = -\mathbb{E}[g(\mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}])] \geq -\mathbb{E}[\mathbb{E}[g(c(\mathbf{U}))|\hat{\mathbf{X}}]] = -\mathbb{E}[g(c(\mathbf{U}))] = H(C).$$

Thus,  $H(C_{\mathbf{X}}^\perp) \geq H(C)$  for all  $C \in \mathcal{C}_{\mathbf{X}}$ .  $\square$

Theorem 2 demonstrates that the entropy of  $C_{\mathbf{X}}$  cannot be greater than  $C_{\mathbf{X}}^\perp$ . This result is related to those of Piantadosi et al. (2012) and Kuzmenko et al. (2020), and the difference is that, in Theorem 2, we fix a joint (possibly discrete) distribution and seek to find the copula consistent with this distribution that maximizes the entropy, which is the checkerboard copula. In contrast, Piantadosi et al. (2012) and Kuzmenko et al. (2020) do not fix a joint distribution. Instead, they search for a checkerboard copula that maximizes the entropy subject to matching either a correlation coefficient or the distribution of the sum of random variables. Therefore, the problems they address are different from our Theorem 2.

When  $C_{\mathbf{X}}$  and  $C_{\mathbf{X}}^\perp$  contain singular components, by definition,  $H(C_{\mathbf{X}}) = H(C_{\mathbf{X}}^\perp) = -\infty$ . In this case, the next proposition shows that, the entropy for the absolutely continuous part of  $C_{\mathbf{X}}^\perp$  is still greater than that for the absolutely continuous part of  $C_{\mathbf{X}}$ .

**Proposition 1.** *Assume  $C_{\mathbf{X}} \in \mathcal{C}_{\mathbf{X}}$  such that  $C_{\mathbf{X}} = \lambda G_{\mathbf{A}} + (1 - \lambda)G_{\mathbf{S}}$ , where  $\lambda \in [0, 1]$ ,  $G_{\mathbf{A}}$  is an absolutely continuous distribution function, and  $G_{\mathbf{S}}$  is a singular distribution function. Then, there exists an absolutely continuous distribution function  $G_{\mathbf{A}}^\perp$  and a distribution function  $G_{\mathbf{S}}^\perp$  such that  $C_{\mathbf{X}}^\perp = \lambda G_{\mathbf{A}}^\perp + (1 - \lambda)G_{\mathbf{S}}^\perp$  and  $H(G_{\mathbf{A}}^\perp) \geq H(G_{\mathbf{A}})$ .*

*Proof.* Let  $F_1, \dots, F_d$  be the marginal distributions of  $\mathbf{X}$ . For  $x \in [0, 1]$ , let  $l_i(x) = \sup\{y : y \in \text{Ran}(F_i), y \leq x\}$  and  $u_i(x) = \inf\{y : y \in \text{Ran}(F_i), y \geq x\}$ . Define two distribution functions  $G_A^\perp$  and  $G_S^\perp$ , which are linear interpolations of  $G_A$  and  $G_S$  from  $\times_{i=1}^d \text{Ran}(F_i)$  to  $[0, 1]^d$ : For  $\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$ ,

$$G_A^\perp(\mathbf{x}) = \begin{cases} G_A(\mathbf{x}), & \mathbf{x} \in \times_{i=1}^d \text{Ran}(F_i), \\ \sum_{\substack{y_i \in \{l_i(x_i), u_i(x_i)\} \\ i \in [d]}} \prod_{j=1}^d \beta_j(x_j, y_j) G_A(y_1, \dots, y_d), & \mathbf{x} \in [0, 1]^d \setminus \times_{i=1}^d \text{Ran}(F_i), \end{cases}$$

and

$$G_S^\perp(\mathbf{x}) = \begin{cases} G_S(\mathbf{x}), & \mathbf{x} \in \times_{i=1}^d \text{Ran}(F_i), \\ \sum_{\substack{y_i \in \{l_i(x_i), u_i(x_i)\} \\ i \in [d]}} \prod_{j=1}^d \beta_j(x_j, y_j) G_S(y_1, \dots, y_d), & \mathbf{x} \in [0, 1]^d \setminus \times_{i=1}^d \text{Ran}(F_i), \end{cases}$$

where

$$\beta_i(x, y) = \frac{u_i(x) - x}{u_i(x) - l_i(x)} \mathbb{1}_{\{y=l_i(x)\}} + \frac{x - l_i(x)}{u_i(x) - l_i(x)} (1 - \mathbb{1}_{\{y=l_i(x)\}})$$

for  $x, y \in [0, 1]$  and  $i \in [d]$  with the convention  $0/0 = 1$ . Note that  $\beta_i(x, y)$  is linear in  $x$  on each segment of  $[0, 1] \setminus \text{Ran}(F_i)$ . It is clear that  $G_A^\perp$  is continuous.

Let  $g_A$  be the density of  $G_A$  and  $g$  be the derivative of  $G_A^\perp$ , respectively. For  $\mathbf{x} \in \times_{i=1}^d \text{Ran}(F_i)$ , we have

$$g(\mathbf{x}) = \frac{\partial^d G_A^\perp(\mathbf{x})}{\partial x_1 \cdots \partial x_d} = \frac{\partial^d G_A(\mathbf{x})}{\partial x_1 \cdots \partial x_d} = g_A(\mathbf{x}).$$

For  $\mathbf{x} = (x_1, \dots, x_{d-1}, x_d)$  such that  $x_i \in \text{Ran}(F_i)$  for  $i \in [d-1]$  and  $x_d \notin \text{Ran}(F_d)$ , we have

$$G_A^\perp(\mathbf{x}) = \frac{u_d(x_d) - x_d}{u_d(x_d) - l_d(x_d)} G_A(x_1, \dots, x_{d-1}, l_d(x_d)) + \frac{x_d - l_d(x_d)}{u_d(x_d) - l_d(x_d)} G_A(x_1, \dots, x_{d-1}, u_d(x_d)).$$

Hence,

$$\begin{aligned} g(\mathbf{x}) &= \frac{\partial^d G_A^\perp(\mathbf{x})}{\partial x_1 \cdots \partial x_d} \\ &= \frac{\partial^{d-1} [G_A(x_1, \dots, x_{d-1}, u_d(x_d)) - G_A(x_1, \dots, x_{d-1}, l_d(x_d))]}{\partial x_1 \cdots \partial x_{d-1}} = \int_{l_d(x_d)}^{u_d(x_d)} g_A(x_1, \dots, x_{d-1}, y) dy. \end{aligned}$$

Let  $k \in [d]$ . Similarly, for  $\mathbf{x} = (x_1, \dots, x_d)$  such that  $x_i \in \text{Ran}(F_i)$  for  $i \in [k-1]$  (with  $[0] = \emptyset$ ) and

$x_i \notin \text{Ran}(F_i)$  for  $i \in [d] \setminus [k-1]$ , we have

$$g(\mathbf{x}) = \int_{l_k(x_k)}^{u_k(x_k)} \cdots \int_{l_d(x_d)}^{u_d(x_d)} g_A(x_1, \dots, x_{k-1}, y_k, \dots, y_d) dy_k \cdots dy_d.$$

Note that  $u_i(x_i) = F_i(F_i^{-1}(x_i))$  and  $l_i(x_i) = F_i(F_i^{-1}(x_i)-)$  for  $i \in [d]$ . From the discussion above, we can see that  $g(\mathbf{x})$  is a constant in  $\times_{i=1}^d [F_i(y_i-), F_i(y_i)]$  for all  $\mathbf{y} = (y_1, \dots, y_d) \in \mathbb{R}^d$ . Let  $\mathbf{U} = (U_1, \dots, U_d) \sim \text{U}([0, 1]^d)$  and  $\hat{\mathbf{X}} = (F_1^{-1}(U_1), \dots, F_d^{-1}(U_d))$ . Thus,  $g(\mathbf{U}) = \mathbb{E}[g_A(\mathbf{U})|\hat{\mathbf{X}}]$ . Hence  $g$  is Lebesgue integrable and  $G_A^\perp$  is an absolutely continuous distribution function with density function  $g_A^\perp = g$ .

Next, we show that  $C_{\hat{\mathbf{X}}}^\perp = \lambda G_A^\perp + (1 - \lambda)G_S^\perp$  and  $H(G_A^\perp) \geq H(G_A)$ . For  $\mathbf{x} \in \times_{i=1}^d \text{Ran}(F_i)$ ,

$$\lambda G_A^\perp(\mathbf{x}) + (1 - \lambda)G_S^\perp(\mathbf{x}) = \lambda G_A(\mathbf{x}) + (1 - \lambda)G_S(\mathbf{x}) = C_{\mathbf{X}}(\mathbf{x}) = C_{\hat{\mathbf{X}}}^\perp(\mathbf{x}).$$

For  $\mathbf{x} \in [0, 1]^d \setminus \times_{i=1}^d \text{Ran}(F_i)$ ,

$$\begin{aligned} \lambda G_A^\perp(\mathbf{x}) + (1 - \lambda)G_S^\perp(\mathbf{x}) &= \sum_{\substack{y_i \in \{l_i(x_i), u_i(x_i)\} \\ i \in [d]}} \prod_{j=1}^d \beta(x_j, y_j) (\lambda G_A(y_1, \dots, y_d) + (1 - \lambda)G_S(y_1, \dots, y_d)) \\ &= \sum_{\substack{y_i \in \{l_i(x_i), u_i(x_i)\} \\ i \in [d]}} \prod_{j=1}^d \beta(x_j, y_j) C_{\mathbf{X}}(y_1, \dots, y_d) = C_{\hat{\mathbf{X}}}^\perp(\mathbf{x}). \end{aligned}$$

Let  $h(x) = -x \log x$  for  $x \in (0, \infty)$ . It is clear that  $h$  is a concave function. By Jensen's inequality, we have

$$H(G_A^\perp) = \mathbb{E}[h(g_A^\perp(\mathbf{U}))] = \mathbb{E}[h(\mathbb{E}[g_A(\mathbf{U})|\hat{\mathbf{X}}])] \geq \mathbb{E}[\mathbb{E}[h(g_A(\mathbf{U}))|\hat{\mathbf{X}}]] = \mathbb{E}[h(g_A(\mathbf{U}))] = H(G_A).$$

This completes the proof. □

## 5 Checkerboard copula and dependence concepts

In this section, we study how the checkerboard copula preserves dependence concepts. This question is motivated by a problem raised in the context of diversification in [Chen et al. \(2024b\)](#), which we describe in [Section 6.1](#).

## 5.1 Dependence concepts

We first define several notions of positive dependence, introduced and studied by [Lehmann \(1966\)](#), [Esary et al. \(1967\)](#), and [Benjamini and Yekutieli \(2001\)](#), and the corresponding notions of negative dependence, introduced and studied by [Lehmann \(1966\)](#), [Alam and Saxena \(1981\)](#), [Block et al. \(1982, 1985\)](#), [Joag-Dev and Proschan \(1983\)](#), and [Chen et al. \(2024a\)](#).

In what follows, for  $i \in [d]$  and an  $d$ -dimensional random vector  $\mathbf{X} = (X_1, \dots, X_d)$ , write  $\mathbf{X}_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_d)$ , and for  $A, B \subseteq [d]$ , write  $\mathbf{X}_A = (X_k)_{k \in A}$  and  $\mathbf{X}_B = (X_k)_{k \in B}$ . A set  $S \subseteq \mathbb{R}^d$  is *decreasing* if  $\mathbf{x} \in S$  implies  $\mathbf{y} \in S$  for all  $\mathbf{y} \leq \mathbf{x}$ .

**Definition 2.** A random vector  $\mathbf{X}$  is

- (i) (a) *positively associated* (PA) if for every pair of subsets  $A, B$  of  $[d]$  and any functions  $f$  and  $g$  both increasing or decreasing coordinatewise, provided the covariance below exists,

$$\text{Cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)) \geq 0;$$

- (b) *negatively associated* (NA) if for every pair of disjoint subsets  $A, B$  of  $[d]$  and any functions  $f$  and  $g$  both increasing or decreasing coordinatewise, provided the covariance below exists,

$$\text{Cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)) \leq 0;$$

- (ii) (a) *positively regression dependent* (PRD) if for every  $i \in [d]$ , the random variable  $\mathbb{E}[g(\mathbf{X}_{-i})|X_i]$  is an increasing function of  $X_i$  for any coordinatewise increasing function  $g$  such that the conditional expectation exists;

- (b) *negatively regression dependent* (NRD) if for every  $i \in [d]$ , the random variable  $\mathbb{E}[g(\mathbf{X}_{-i})|X_i]$  is a decreasing function of  $X_i$  for any coordinatewise increasing function  $g$  such that the conditional expectation exists;

- (iii) (a) *weakly positively associated* (WPA) if for any  $i \in [d]$ , decreasing set  $S \subseteq \mathbb{R}^{d-1}$ , and  $x \in \mathbb{R}$  with  $\mathbb{P}(X_i \leq x) > 0$ ,

$$\mathbb{P}(\mathbf{X}_{-i} \in S \mid X_i \leq x) \geq \mathbb{P}(\mathbf{X}_{-i} \in S);$$

- (b) *weakly negatively associated* (WNA) if for any  $i \in [d]$ , decreasing set  $S \subseteq \mathbb{R}^{d-1}$ , and  $x \in \mathbb{R}$  with  $\mathbb{P}(X_i \leq x) > 0$ ,

$$\mathbb{P}(\mathbf{X}_{-i} \in S \mid X_i \leq x) \leq \mathbb{P}(\mathbf{X}_{-i} \in S);$$

- (iv) (a) *positively orthant dependent* (POD) if for all  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ ,  $\mathbb{P}(\mathbf{X} \leq \mathbf{x}) \geq \prod_{i=1}^d \mathbb{P}(X_i \leq x_i)$  and  $\mathbb{P}(\mathbf{X} > \mathbf{x}) \geq \prod_{i=1}^d \mathbb{P}(X_i > x_i)$ ;
- (b) *negatively orthant dependent* (NOD) if for all  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ ,  $\mathbb{P}(\mathbf{X} \leq \mathbf{x}) \leq \prod_{i=1}^d \mathbb{P}(X_i \leq x_i)$  and  $\mathbb{P}(\mathbf{X} > \mathbf{x}) \leq \prod_{i=1}^d \mathbb{P}(X_i > x_i)$ .

Moreover, we say that a distribution or a copula is PA, PRD, WPA, POD, NA, NRD, WNA, or NOD if the corresponding random vector is.

Note that the definition of PA does not require  $A$  and  $B$  to be disjoint, whereas the definition of NA requires this.

The relationship between the above notions is summarized below (see e.g., [Chen et al. \(2024a\)](#)).

$$\text{PA} \implies \text{WPA}; \quad \text{PRD} \implies \text{WPA}; \quad \text{WPA} \implies \text{POD};$$

$$\text{NA} \implies \text{WNA}; \quad \text{NRD} \implies \text{WNA}; \quad \text{WNA} \implies \text{NOD}.$$

Within the class of multivariate normal distributions, the four concepts of positive dependence are equivalent, and each is equivalent to having nonnegative bivariate correlation coefficients; similarly, the four concepts of negative dependence are equivalent, and each is equivalent to having nonpositive bivariate correlation coefficients.

In the sequel, we use  $\mathfrak{D}$  to represent one of the following: PA, PRD, WPA, POD, NA, NRD, WNA, or NOD. Our question is whether these properties are properties purely based on copulas. It turns out that the checkerboard copula can help answer this question.

## 5.2 The checkerboard copula preserves dependence

We first present a self-consistency property of those negative dependence concepts in the spirit of [Joag-Dev and Proschan \(1983, Property P<sub>6</sub>\)](#) for NA.

**Lemma 2.** *If  $f_1, \dots, f_d$  are increasing functions and  $\mathbf{X}$  satisfies  $\mathfrak{D}$ , then  $(f_1(X_1), \dots, f_d(X_d))$  also satisfies  $\mathfrak{D}$ .*

*Proof.* We only show the result for the concepts of negative dependence, as the case of positive dependence is similar.

The self-consistency properties of NA and NOD are shown in [Joag-Dev and Proschan \(1983, Property P<sub>6</sub>\)](#) and [Lehmann \(1966, Lemma 1\)](#), respectively. We will show the properties for NRD and WNA. Let  $\mathbf{Y} = (f_1(X_1), \dots, f_d(X_d))$ .

1. Assume  $\mathbf{X}$  is NRD. Fix  $i \in [d]$ . Let  $g$  be a coordinatewise increasing function and  $g' = g \circ (f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_d)$ . As a result, we have  $g'$  is a coordinatewise increasing function and  $g(\mathbf{Y}_{-i}) = g'(\mathbf{X}_{-i})$ . For any  $y \in \mathbb{R}$ , let  $A_y = \{x : f_i(x) = y\}$ . We have  $\{Y_i = y\} = \{X_i \in A_y\}$ . Therefore,  $\mathbb{E}[g(\mathbf{Y}_{-i})|Y_i = y] = \mathbb{E}[g'(\mathbf{X}_{-i})|X_i \in A_y]$ . Assume  $y_1 < y_2$ . For any  $x_1 \in A_{y_1}$  and  $x_2 \in A_{y_2}$ , we have  $x_1 \leq x_2$ ; hence,  $\mathbb{E}[g'(\mathbf{X}_{-i})|X_i = x_1] \geq \mathbb{E}[g'(\mathbf{X}_{-i})|X_i = x_2]$ . Thus,

$$\begin{aligned} \mathbb{E}[g'(\mathbf{X}_{-i})|X_i \in A_{y_1}] &= \mathbb{E}[\mathbb{E}[g'(\mathbf{X}_{-i})|X_i]|X_i \in A_{y_1}] \\ &\geq \mathbb{E}[\mathbb{E}[g'(\mathbf{X}_{-i})|X_i]|X_i \in A_{y_2}] = \mathbb{E}[g'(\mathbf{X}_{-i})|X_i \in A_{y_2}], \end{aligned}$$

which implies that  $\mathbb{E}[g(\mathbf{Y}_{-i})|Y_i = y_1] \geq \mathbb{E}[g(\mathbf{Y}_{-i})|Y_i = y_2]$ ; hence  $\mathbf{Y}$  is NRD.

2. Assume  $\mathbf{X}$  is WNA. For  $i \in [d]$ , let  $S \subseteq \mathbb{R}^{d-1}$  be a decreasing set, and

$$S_i^f = \{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d) : (f_1(x_1), \dots, f_{i-1}(x_{i-1}), f_{i+1}(x_{i+1}), \dots, f_d(x_d)) \in S\}.$$

It is clear that  $\{\mathbf{Y}_{-i} \in S\} = \{\mathbf{X}_{-i} \in S_i^f\}$ . For any  $\mathbf{x}_1 \leq \mathbf{x}_2$  and  $\mathbf{x}_2 \in S_i^f$ , we have  $f_k(x_{1,k}) \leq f_k(x_{2,k})$  for all  $k \in [d] \setminus \{i\}$ . Furthermore, because  $S$  is decreasing, we have  $\mathbf{x}_1 \in S_i^f$ , which implies  $S_i^f$  is a decreasing set. For any  $y \in \mathbb{R}$  with  $\mathbb{P}(Y_i \leq y) > 0$ , let  $x = \sup\{t \in \mathbb{R} : f_i(t) \leq y\}$ . If  $f_i(x) \leq y$ , we have  $\{Y_i \leq y\} = \{X_i \leq x\}$  and  $\mathbb{P}(X_i \leq x) > 0$ . Therefore,

$$\mathbb{P}(\mathbf{Y}_{-i} \in S|Y_i \leq y) = \mathbb{P}(\mathbf{X}_{-i} \in S_i^f|X_i \leq x) \leq \mathbb{P}(\mathbf{X}_{-i} \in S_i^f) = \mathbb{P}(\mathbf{Y}_{-i} \in S),$$

which implies that  $\mathbf{Y}$  is WNA. If  $f_i(x) > y$ , we have  $\{Y_i \leq y\} = \{X_i < x\}$  and  $\mathbb{P}(X_i < x) > 0$ . Therefore,

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_{-i} \in S, Y_i \leq y) &= \mathbb{P}(\mathbf{X}_{-i} \in S_i^f, X_i < x) \\ &= \lim_{t \uparrow x} \mathbb{P}(\mathbf{X}_{-i} \in S_i^f, X_i \leq t) \\ &\leq \lim_{t \uparrow x} \mathbb{P}(\mathbf{X}_{-i} \in S_i^f) \mathbb{P}(X_i \leq t) \\ &= \mathbb{P}(\mathbf{X}_{-i} \in S_i^f) \lim_{t \uparrow x} \mathbb{P}(X_i \leq t) \\ &= \mathbb{P}(\mathbf{X}_{-i} \in S_i^f) \mathbb{P}(X_i < x) = \mathbb{P}(\mathbf{Y}_{-i} \in S) \mathbb{P}(Y_i \leq y), \end{aligned}$$

which implies that  $\mathbb{P}(\mathbf{Y}_{-i} \in S|Y_i \leq y) \leq \mathbb{P}(\mathbf{Y}_{-i} \in S)$  and  $\mathbf{Y}$  is WNA.  $\square$

The following theorem demonstrates that the checkerboard copula of  $\mathbf{X}$  preserves the dependence information of  $\mathbf{X}$ .



**Theorem 3.** *A random vector  $\mathbf{X}$  satisfies  $\mathfrak{D}$  if and only if it has a copula that satisfies  $\mathfrak{D}$ . Moreover, the copula can be chosen as the checkerboard copula  $C_{\mathbf{X}}^{\perp}$ .*

*Proof.* The “if” part follows from Lemma 2 because, for  $\mathbf{U} = (U_1, \dots, U_d)$  following the copula of  $\mathbf{X}$  that satisfies  $\mathfrak{D}$ , we have  $(X_1, \dots, X_d) = (F_1^{-1}(U_1), \dots, F_d^{-1}(U_d))$  and  $F_i^{-1}$  is increasing for all  $i \in [d]$ .

Now we show the “only if” part. Let  $\mathbf{U} = (U_1, \dots, U_d)$  be the random vector given by (4) with  $\mathbf{V} = (V_1, \dots, V_d) \sim U([0, 1]^d)$  independent of  $\mathbf{X}$ . Hence, we have  $\mathbf{U} \sim C_{\mathbf{X}}^{\perp}$  and  $C_{\mathbf{X}}^{\perp}$  is a copula of  $\mathbf{X}$ . Note that, for any  $i \in [d]$ , given  $V_i$ , we have that  $U_i$  is an increasing function of  $X_i$ . Hence, by Lemma 2,  $\mathbf{X}$  satisfies  $\mathfrak{D}$  implies that  $\mathbf{U}|\mathbf{V}$  also satisfies  $\mathfrak{D}$ .

Assume  $\mathbf{X}$  is NA. For any given pair of disjoint subsets  $A, B$  of  $[d]$  and any given functions  $f$  and  $g$  both increasing or decreasing coordinatewise, we have

$$\begin{aligned} \text{Cov}(f(\mathbf{U}_A), g(\mathbf{U}_B)) &= \mathbb{E}[\text{Cov}(f(\mathbf{U}_A), g(\mathbf{U}_B)|\mathbf{V})] + \text{Cov}(\mathbb{E}[f(\mathbf{U}_A)|\mathbf{V}], \mathbb{E}[g(\mathbf{U}_B)|\mathbf{V}]) \\ &\leq 0 + \text{Cov}(\mathbb{E}[f(\mathbf{U}_A)|\mathbf{V}_A], \mathbb{E}[g(\mathbf{U}_B)|\mathbf{V}_B]) = 0, \end{aligned}$$

where the inequality follows from  $\mathbf{U}|\mathbf{V}$  is NA, and the last equality follows from the independence between  $\mathbf{V}_A$  and  $\mathbf{V}_B$ . Hence,  $\mathbf{U}$  is NA.

Assume  $\mathbf{X}$  is NRD. For any fixed  $i$  and  $k$ , by (4), there exist  $x$  and  $v$  such that  $\{U_i = k\} = \{X_i = x, V_i = v\}$ . Then, for any coordinatewise increasing function  $g$ , by the independence between  $V_i$  and  $(X_i, \mathbf{U}_{-i})$ , we have

$$\mathbb{E}[g(\mathbf{U}_{-i})|U_i = k] = \mathbb{E}[g(\mathbf{U}_{-i})|X_i = x, V_i = v] = \mathbb{E}[g(\mathbf{U}_{-i})|X_i = x].$$

Because  $\mathbf{U}_{-i}$  is a function of  $\mathbf{X}_{-i}$  and  $\mathbf{V}_{-i}$ , we can let  $h$  be the function such that  $g(\mathbf{U}_{-i}) = h(\mathbf{X}_{-i}, \mathbf{V}_{-i})$ . Then, due to the independence between  $\mathbf{V}_{-i}$  and  $\mathbf{X}$ ,

$$\mathbb{E}[g(\mathbf{U}_{-i})|X_i = x] = \mathbb{E}[h(\mathbf{X}_{-i}, \mathbf{V}_{-i})|X_i = x] = \int_{[0,1]^{d-1}} \mathbb{E}[h(\mathbf{X}_{-i}, \mathbf{v}_{-i})|X_i = x] d\mathbf{v}_{-i},$$

where  $\mathbf{v}_{-i} = (v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_d)$ . Therefore, for any  $k_1 \leq k_2$ , there exist  $x_1$  and  $x_2$  such that

$$\begin{aligned} \mathbb{E}[g(\mathbf{U}_{-i})|U_i = k_1] &= \int_{[0,1]^{d-1}} \mathbb{E}[h(\mathbf{X}_{-i}, \mathbf{v}_{-i})|X_i = x_1] d\mathbf{v}_{-i}, \\ \mathbb{E}[g(\mathbf{U}_{-i})|U_i = k_2] &= \int_{[0,1]^{d-1}} \mathbb{E}[h(\mathbf{X}_{-i}, \mathbf{v}_{-i})|X_i = x_2] d\mathbf{v}_{-i}. \end{aligned}$$

In addition, by (4), we must have  $x_1 \leq x_2$ . Note that given  $\mathbf{v}_{-i}$ ,  $h(\mathbf{X}_{-i}, \mathbf{v}_{-i})$  is a coordinatewise

increasing function of  $\mathbf{X}_{-i}$ . Hence, we have  $\mathbb{E}[h(\mathbf{X}_{-i}, \mathbf{v}_{-i})|X_i = x_1] \geq \mathbb{E}[h(\mathbf{X}_{-i}, \mathbf{v}_{-i})|X_i = x_2]$  for any  $\mathbf{v}_{-i}$ . Therefore,  $\mathbb{E}[g(\mathbf{U}_{-i})|U_i = k_1] \geq \mathbb{E}[g(\mathbf{U}_{-i})|U_i = k_2]$  and  $\mathbf{U}$  is NRD.

Assume  $\mathbf{X}$  is WNA. For any  $i \in [d]$ , decreasing set  $S \subseteq \mathbb{R}^{d-1}$ , and  $x \in \mathbb{R}$  with  $\mathbb{P}(U_i \leq x) > 0$ ,

$$\begin{aligned} \mathbb{P}(\mathbf{U}_{-i} \in S, U_i \leq x) &= \mathbb{E}[\mathbb{P}(\mathbf{U}_{-i} \in S, U_i \leq x | \mathbf{V})] \\ &\leq \mathbb{E}[\mathbb{P}(\mathbf{U}_{-i} \in S | \mathbf{V}_{-i}) \mathbb{P}(U_i \leq x | V_i)] \\ &= \mathbb{E}[\mathbb{P}(\mathbf{U}_{-i} \in S | \mathbf{V}_{-i})] \mathbb{E}[\mathbb{P}(U_i \leq x | V_i)] \\ &= \mathbb{P}(\mathbf{U}_{-i} \in S) \mathbb{P}(U_i \leq x). \end{aligned}$$

Hence,  $\mathbf{U}$  is WNA.

Assume  $\mathbf{X}$  is NOD. For any  $t_1, \dots, t_d \in \mathbb{R}$ , we have

$$\begin{aligned} \mathbb{P}(U_1 \leq t_1, \dots, U_d \leq t_d) &= \mathbb{E}[\mathbb{P}(U_1 \leq t_1, \dots, U_d \leq t_d | V_1, \dots, V_d)] \\ &\leq \mathbb{E}[\mathbb{P}(U_1 \leq t_1 | V_1) \cdots \mathbb{P}(U_d \leq t_d | V_d)] \\ &= \mathbb{E}[\mathbb{P}(U_1 \leq t_1 | V_1)] \cdots \mathbb{E}[\mathbb{P}(U_d \leq t_d | V_d)] \\ &= \mathbb{P}(U_1 \leq t_1) \cdots \mathbb{P}(U_d \leq t_d). \end{aligned}$$

Similarly, we can show

$$\mathbb{P}(U_1 > t_1, \dots, U_d > t_d) \leq \mathbb{P}(U_1 > t_1) \cdots \mathbb{P}(U_d > t_d).$$

Hence,  $\mathbf{U}$  is NOD.

In conclusion, if  $\mathbf{X}$  satisfies  $\mathfrak{D}$ , then  $\mathbf{U}$  satisfies  $\mathfrak{D}$ , where  $\mathfrak{D}$  is one of the four concepts of negative dependence.

To show the case of positive dependence, we follow a similar route. We take the same  $\mathbf{U}$  as above. Assume  $\mathbf{X}$  is PA. Because  $U_i|V_i$  is an increasing function of  $X_i$ , by Lemma 2,  $\mathbf{U}|\mathbf{V}$  is also PA. Thus, for any given pair of subsets  $A, B$  of  $[d]$  and any given functions  $f$  and  $g$  both coordinatewise increasing or decreasing, we have

$$\begin{aligned} \text{Cov}(f(\mathbf{U}_A), g(\mathbf{U}_B)) &= \mathbb{E}[\text{Cov}(f(\mathbf{U}_A), g(\mathbf{U}_B) | \mathbf{V})] + \text{Cov}(\mathbb{E}[f(\mathbf{U}_A) | \mathbf{V}], \mathbb{E}[g(\mathbf{U}_B) | \mathbf{V}]) \\ &\geq \text{Cov}(\mathbb{E}[f(\mathbf{U}_A) | \mathbf{V}_A], \mathbb{E}[g(\mathbf{U}_B) | \mathbf{V}_B]). \end{aligned}$$

Moreover, given  $\mathbf{X}$ ,  $U_i$  is an increasing function of  $V_i$ . Hence,  $\mathbb{E}[f(\mathbf{U}_A) | \mathbf{V}_A]$  and  $\mathbb{E}[g(\mathbf{U}_B) | \mathbf{V}_B]$  are coordinatewise increasing (or decreasing) with respect to  $\mathbf{V}_A$  and  $\mathbf{V}_B$ , respectively, if  $f$  and  $g$  are

both coordinatewise increasing (or decreasing). Because  $\mathbf{V}$  is PA, we have

$$\text{Cov}(\mathbb{E}[f(\mathbf{U}_A)|\mathbf{V}_A], \mathbb{E}[g(\mathbf{U}_B)|\mathbf{V}_B]) \geq 0,$$

implying that  $\mathbf{U}$  is PA. The proofs for other positive dependence concepts are similar. A partial proof for POD can also be found in [Durante et al. \(2015, Proposition 2.3\)](#).  $\square$

[Genest and Nešlehová \(2007\)](#) also show the dependence preservation results for positive orthant dependence, positive likelihood ratio dependence, and tail dependence in bivariate case.

*Remark 2.* We can clearly see from [Theorem 3](#) that  $C_{\mathbf{X}}^{\perp} = \Pi$  (the independence copula) is the only independent copula in  $\mathcal{C}_{\mathbf{X}}$  when  $\mathbf{X}$  is independent. This fact is used as the basis of the independence test in [Genest et al. \(2019\)](#).

## 6 Two consequences of Theorem 3

We provide two applications in this section to highlight the usefulness of [Theorem 3](#).

### 6.1 Diversification penalty

For random variables  $X$  and  $Y$ , let  $X \geq_{\text{st}} Y$  represent  $\mathbb{P}(X > x) \geq \mathbb{P}(Y > x)$  for all  $x \in \mathbb{R}$ ; this is called the stochastic order. [Chen et al. \(2024a,b\)](#) studied the problem of diversification penalty; that is, whether

$$X \leq_{\text{st}} \sum_{i=1}^d \theta_i X_i \text{ for all } (\theta_1, \dots, \theta_d) \in \Delta_d, \text{ where } X, X_1, \dots, X_d \text{ are identically distributed,} \quad (5)$$

holds under certain marginal distributions and dependence structures. Here,  $\Delta_d$  is the standard simplex defined by  $\Delta_d = \{(\theta_1, \dots, \theta_d) \in [0, 1]^d : \theta_1 + \dots + \theta_d = 1\}$ . When  $X$  is interpreted as a loss, (5) intuitively means that the non-diversified portfolio  $X$  is less dangerous than the diversified portfolio  $\sum_{i=1}^d \theta_i X_i$ . This seems counter-intuitive at first glance, but it indeed happens in the model of [Chen et al. \(2024a\)](#), where  $X$  has infinite mean.

Define the set, for some dependence concept  $\mathfrak{D}$  in [Section 5.1](#),

$$\mathcal{F}_{\mathfrak{D}} = \{\text{distribution of } X : (5) \text{ holds for all } (X_1, \dots, X_d) \text{ that satisfy } \mathfrak{D}\}.$$

[Chen et al. \(2024a\)](#) showed that the Pareto(1) distribution belongs to  $\mathcal{F}_{\text{WNA}}$ , and hence also to  $\mathcal{F}_{\text{NA}}$ ,  $\mathcal{F}_{\text{NRD}}$ , and  $\mathcal{F}_{\text{IN}}$ , where IN stands for independence. Moreover, [Chen et al. \(2024b, Proposition 1\)](#)

showed that  $\mathcal{F}_{\mathfrak{D}}$  for  $\mathfrak{D}$  being WNA, NA, or IN is closed under strictly increasing convex transforms on the random variables. Our next result, which relies on our Theorem 3, addresses non-strictly increasing  $f$  and other notions of dependence, thus generalizing the above result.

**Proposition 2.** *Each of  $\mathcal{F}_{\mathfrak{D}}$  is closed under increasing convex transforms on the random variable.*

*Proof.* Below we first show that each of  $\mathcal{F}_{\mathfrak{D}}$  is closed under strictly increasing convex transforms on the random variable, that is, if the distribution of  $X$  is in  $\mathcal{F}_{\mathfrak{D}}$ , so is the distribution of  $f(X)$  for a strictly increasing convex  $f$ . Assume that  $F \in \mathcal{F}_{\mathfrak{D}}$ ,  $X$  follows  $F$ , and  $Y = f(X)$ , where  $f$  is strictly increasing and convex. Because  $f$  is strictly increasing, if  $(Y_1, \dots, Y_d)$  satisfies  $\mathfrak{D}$ , so does  $(X_1, \dots, X_d)$ , where  $X_i = f^{-1}(Y_i)$  for  $i \in [d]$ , by Lemma 2. Because each of  $X, X_1, \dots, X_d$  has a distribution  $F \in \mathcal{F}_{\mathfrak{D}}$ , we have  $X \leq_{\text{st}} \sum_{i=1}^d \theta_i X_i$ , and this gives, using the convexity of  $f$ ,

$$Y = f(X) \leq_{\text{st}} f\left(\sum_{i=1}^d \theta_i X_i\right) \leq \sum_{i=1}^d \theta_i f(X_i) = \sum_{i=1}^d \theta_i Y_i. \quad (6)$$

To address the case that  $f$  is not strictly increasing, Theorem 3 allows us to find the above  $(X_1, \dots, X_d)$  that satisfies  $\mathfrak{D}$  and such that  $Y_i = f(X_i)$  for  $i \in [d]$ . In particular, using Theorem 3, we can construct  $(U_1, \dots, U_d)$  that follows the checkerboard copula of  $(Y_1, \dots, Y_d)$  and satisfies  $\mathfrak{D}$ , such that

$$(Y_1, \dots, Y_d) = (f \circ g(U_1), \dots, f \circ g(U_d)),$$

where  $g$  is the quantile function of  $X$  and  $f \circ g$  is the quantile function of  $Y$ . Setting  $(X_1, \dots, X_d) = (g(U_1), \dots, g(U_d))$ , we get that  $(X_1, \dots, X_d)$  satisfies  $\mathfrak{D}$ , and this leads to (6).  $\square$

## 6.2 Induced order statistics

Here we demonstrate another application of Theorem 3 in characterizing the distribution of induced order statistics. Consider  $N$  independent and identically distributed bivariate random vectors

$$\begin{pmatrix} \xi_1 \\ \eta_1 \end{pmatrix}, \begin{pmatrix} \xi_2 \\ \eta_2 \end{pmatrix}, \dots, \begin{pmatrix} \xi_N \\ \eta_N \end{pmatrix}.$$

Note that, for  $i \neq j$ ,  $(\xi_i, \eta_i)$  and  $(\xi_j, \eta_j)$  are independent and identically distributed, but  $\xi_i$  and  $\eta_i$  may be correlated and have different marginal distributions. We rank these bivariate vectors according to their first components,  $\xi_i$ :

$$\begin{pmatrix} \xi_{1:N} \\ \eta_{[1:N]} \end{pmatrix}, \begin{pmatrix} \xi_{2:N} \\ \eta_{[2:N]} \end{pmatrix}, \dots, \begin{pmatrix} \xi_{N:N} \\ \eta_{[N:N]} \end{pmatrix}, \quad (7)$$

where  $\xi_{1:N} \leq \xi_{2:N} \leq \dots \leq \xi_{N:N}$  are the order statistics of  $\xi_1, \xi_2, \dots, \xi_N$ . The notation  $\eta_{[i:N]}$  represents the  $i$ -th *induced order statistic* (Bhattacharya, 1974), where the order is induced by another variable  $\xi_i$ . The induced order statistics  $\eta_{[1:N]}, \dots, \eta_{[N:N]}$  are also referred to as *concomitants* of the order statistics  $\xi_{1:N}, \dots, \xi_{N:N}$  (David, 1973).

In the context of constructing impact portfolios, Lo et al. (2024) investigated the joint distribution of  $(\eta_{[1:N]}, \dots, \eta_{[N:N]})$ . In particular, they proved a representation theorem for the joint distribution of  $(\eta_{[1:N]}, \dots, \eta_{[N:N]})$  using the copula of  $(\xi_i, \eta_i)$ . Furthermore, they demonstrated that if  $\xi_i$  is not continuously distributed, the representation theorem holds if and only if the copula of  $(\xi_i, \eta_i)$  is chosen as the (bivariate) checkerboard copula in this paper. This reveals a potential application of the checkerboard copula in portfolio construction.

Lo et al. (2024) also showed that the rank of the odd-order moments of induced order statistics relies on the copula of  $(\xi_i, \eta_i)$ . Assume that  $C$  is a copula of  $(\xi_i, \eta_i)$ . Lo et al. (2024, Theorem EC.5) proved that, for any  $k = 0, 1, \dots$ , if  $C$  is PRD, we have

$$\mathbb{E} \left( \eta_{[1:N]}^{2k+1} \right) \leq \mathbb{E} \left( \eta_{[2:N]}^{2k+1} \right) \leq \dots \leq \mathbb{E} \left( \eta_{[N:N]}^{2k+1} \right), \quad (8)$$

and if  $C$  is NRD, we have

$$\mathbb{E} \left( \eta_{[1:N]}^{2k+1} \right) \geq \mathbb{E} \left( \eta_{[2:N]}^{2k+1} \right) \geq \dots \geq \mathbb{E} \left( \eta_{[N:N]}^{2k+1} \right). \quad (9)$$

In particular, the copula  $C$  can be chosen as the checkerboard copula. Therefore, using Theorem 3, we directly obtain the following result.

**Proposition 3.** *For any  $k = 0, 1, \dots$ , (8) holds if  $(\xi_i, \eta_i)$  is PRD, and (9) holds if  $(\xi_i, \eta_i)$  is NRD.*

The difference between Proposition 3 and Lo et al. (2024, Theorem EC.5) is that the latter imposes the dependence assumption (PRD or NRD) on the copula of  $(\xi_i, \eta_i)$ , while the former imposes a more natural assumption on the random vector  $(\xi_i, \eta_i)$  directly, which is only possible due to our Theorem 3.

## 7 Applications in co-risk measures and portfolio selection

In this section, we use both numerical and empirical experiments to show that the choice of copula impacts the calculation of co-risk measures when the marginal distributions are not continuous. In particular, we consider Marginal ES as defined in (1), which is discussed in our third motivating question in Section 1. Section 7.1 presents a numerical experiment to show that different

copula choices can lead to varying Marginal ES results. Section 7.2 uses real data from the U.S. stock market to illustrate how the choice of copula affects the performance of the portfolio with minimum Marginal ES. Our results not only highlight the importance of copula selection in the computation of co-risk measures in financial practice, but also show that the checkerboard copula is often the most convenient and natural choice that can produce reliable results.

## 7.1 Numerical experiment

Consider a bivariate normal distribution with marginals  $N(0, \sigma^2)$  and correlation  $r$ . Denote this bivariate distribution by  $F_r$ . We choose this distribution because it is well known that, for  $(X_1, X_2) \sim F_r$ , the Marginal ES of  $X_2$  given  $X_1$  at level  $p \in (0, 1)$  can be explicitly computed as

$$\rho(X_2|X_1) = \mathbb{E}[X_2|X_1 > \Phi^{-1}(p)] = \frac{r\sigma}{1-p}\varphi(\Phi^{-1}(p)), \quad (10)$$

where  $\varphi$  and  $\Phi$  are the density and distribution functions of  $N(0, 1)$ , respectively.

Now we conduct a numerical experiment based on this bivariate normal distribution. We draw 1,000 bivariate random vectors,  $(X_1^{(1)}, X_2^{(1)}), \dots, (X_1^{(1,000)}, X_2^{(1,000)})$ , from  $F_r$ . These random values are then rounded to one decimal place to estimate an empirical bivariate distribution, denoted by  $\hat{F}$ . Thus,  $\hat{F}$  is a discrete bivariate distribution on the discrete grid caused by rounding. Next, assuming that  $(\hat{X}_1, \hat{X}_2) \sim \hat{F}$ , we compute the Marginal ES of  $\hat{X}_2$  given  $\hat{X}_1$  at level  $p \in (0, 1)$ , which is  $\rho(\hat{X}_2|\hat{X}_1)$  defined as (1). When calculating the Marginal ES, the following two copulas of  $(\hat{X}_1, \hat{X}_2)$  given by (4) are considered:

- (i) Let  $(V_1, V_2) \sim U([0, 1]^2)$  be independent of  $(\hat{X}_1, \hat{X}_2)$ . That is to use the checkerboard copula  $C^\perp$ .
- (ii) Let  $V_2 \sim U[0, 1]$  be independent of  $\hat{X}_2$ . In addition, given  $\hat{X}_2$ , let  $V_2$  and  $\hat{X}_1$  be comonotonic. Let  $V_1 \sim U[0, 1]$  be independent of  $\hat{X}_1, \hat{X}_2$ , and  $V_2$ . We denote this copula by  $C^+$ .

Therefore, we obtain different values of  $\rho(\hat{X}_2|\hat{X}_1)$  using the two different copulas.

Given  $(r, p)$ , we run the simulation procedure described above 10,000 times and calculate the Marginal ES under both copulas for each run. We choose  $\sigma = 10$ . Table 1 shows the Marginal ES given by (10) (normal formula), the average Marginal ES value, and the mean squared error (MSE)—the mean squared difference between  $\rho(\hat{X}_2|\hat{X}_1)$  and the value in (10)—across the 10,000 runs for each of the two copulas.

Table 1 illustrates that the choice of copula affects the calculation of Marginal ES. Under our setup, the Marginal ES computed using the checkerboard copula is, on average, closer to the result

Table 1: Simulation results for computing Marginal ES under two copulas.

	$p = 0.9$			$p = 0.95$			$p = 0.975$		
$\rho$	0.2	0.3	0.4	0.2	0.3	0.4	0.2	0.3	0.4
normal formula	3.510	5.265	7.020	4.125	6.188	8.251	4.676	7.013	9.351
average with $C^\perp$	3.502	5.257	7.001	4.116	6.171	8.234	4.688	6.992	9.330
average with $C^+$	3.985	5.730	7.455	4.687	6.726	8.767	5.338	7.615	9.934
MSE with $C^\perp$	0.940	0.920	0.891	1.914	1.796	1.673	3.725	3.636	3.329
MSE with $C^+$	1.225	1.192	1.131	2.348	2.173	2.020	4.357	4.187	3.796

obtained using the normal formula (10). This demonstrates that the checkerboard copula is a good candidate for computing co-risk measures for non-continuous random variables.

## 7.2 Empirical study

To further demonstrate the importance of copula selection when computing co-risk measures in financial practice, we use real stock data to calculate the Marginal ES. We obtain daily returns of the S&P 500 Index and five widely traded US stocks—Microsoft, Apple, Google, Nvidia, and Amazon—from 2005 to 2023.<sup>2</sup> To make the distribution discrete, we classify market conditions into five groups based on the daily returns of the S&P 500 Index:  $(-\infty, -3\%]$ ,  $(-3\%, -1\%]$ ,  $(-1\%, +1\%]$ ,  $(+1\%, +3\%]$ , and  $(+3\%, +\infty)$ . These five conditions represent very bad, bad, fair, good, and very good, with corresponding values of  $-2$ ,  $-1$ ,  $0$ ,  $+1$ , and  $+2$ , respectively.

Next, for each individual stock, we use the negative values of its daily returns along with these market condition values over the entire period to estimate an empirical bivariate distribution. Then, based on the empirical bivariate distribution, the two copulas described in Section 7.1 are applied to calculate the Marginal ES of the stock’s loss given the market condition. We also present results computed using the empirical bivariate distributions of daily stock returns and S&P 500 Index returns directly, without classifying the market conditions.

Table 2 shows the Marginal ES of the five stocks under the two copulas, along with the result computed directly from the empirical distributions. The Marginal ES values differ across the three methods, and the results from the checkerboard copula,  $C^\perp$ , are lower than those from the alternative copula,  $C^+$ .

Different choices of copulas can also result in varying financial performance for the portfolio with minimum Marginal ES (minMES). To demonstrate this, we construct minMES portfolios for the five stocks as follows. At the beginning of year  $t$ , we determine the weights of the five stocks

<sup>2</sup>Our data comes from the Center for Research in Security Prices (CRSP). We use data starting in 2005 because Google became publicly traded in August 2004.

Table 2: Marginal ES of the loss of five individual stocks given the loss of the S&P 500 Index.

	Microsoft	Apple	Google	Nvidia	Amazon
$C^\perp$	3.518%	3.545%	3.455%	5.043%	3.712%
$C^+$	4.404%	4.752%	4.523%	6.686%	4.998%
empirical distribution	3.845%	3.905%	3.735%	5.456%	4.054%

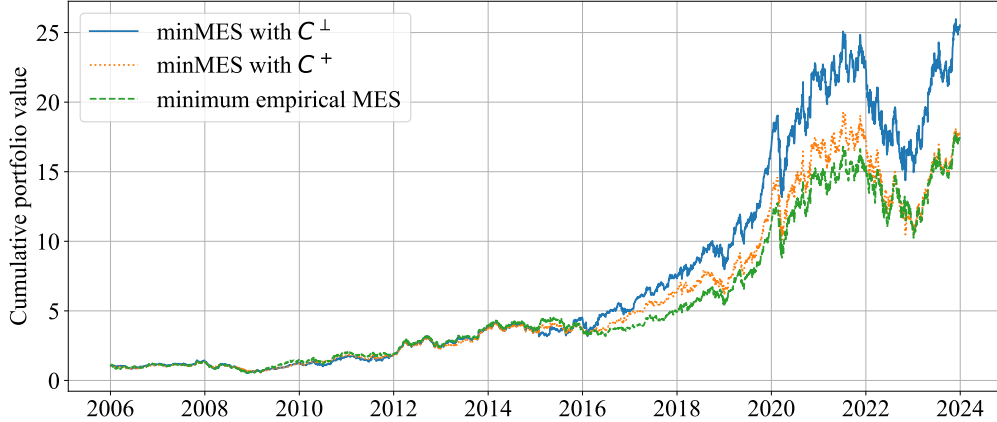


Figure 1: Cumulative values of the minMES portfolios.

that minimize the Marginal ES of the portfolio’s loss given the market condition, using data from year  $t - 1$ . The optimization is subject to the constraints that all weights must be non-negative and sum to 1. These optimal weights are then held throughout year  $t$ .

Figure 1 and Table 3 show the cumulative portfolio values and performance metrics of this minMES strategy for  $p = 0.975$  under the three methods over the entire sample period, respectively. We find that the choice of copula significantly impacts the financial performance of the minMES portfolio. Furthermore, in our empirical study, the checkerboard copula generally achieves better performance, demonstrating that it is a convenient and effective choice for producing reliable results for the considered dataset. Certainly, we do not claim that this advantage is profitable in general, which requires comprehensive empirical analysis.

## 8 Conclusion

We discussed the choice of copula when the marginal distributions are not necessarily continuous. Among all the choices of copulas for a given random vector, the checkerboard copula is the most convenient and natural selection in applications such as simulating from the copula, stressing the distribution, and computing a co-risk measure. It is shown that the checkerboard copula is the most unbiased choice in the sense that it has the largest Shannon entropy among all possible



Table 3: Performance metrics (average annual returns, standard deviation of annual returns, and Sharpe ratio) for the minMES portfolios. The Sharpe ratio is calculated assuming a risk-free rate of 3%.

metrics	average return	standard deviation	Sharpe ratio
$C^\perp$	22.34%	29.89%	0.6470
$C^+$	19.73%	27.65%	0.6050
empirical distribution	20.11%	29.92%	0.5721

copulas for a given random vector. Moreover, the checkerboard copula can preserve the dependence information of the underlying random vector. This preservation property is applied to identify suitable distributions in the context of diversification penalty studied by [Chen et al. \(2024a,b\)](#) and to determine the ranks of the moments of induced order statistics in the context of impact portfolios studied by [Lo et al. \(2024\)](#). Finally, our results indicate that the choice of copula significantly affects the calculation of co-risk measures when the marginal distributions are not continuous. Through numerical experiments and empirical studies, we find that the checkerboard copula can produce reliable results when computing Marginal ES and demonstrate strong performance when constructing minMES portfolios.

## References

- Adrian, T. and Brunnermeier, M. K. (2016). CoVaR. *American Economic Review*, **106**(7), 1705–1741.
- Alam, K. and Saxena, K. M. L. (1981). Positive dependence in multivariate distributions. *Communications in Statistics-Theory and Methods*, **10**(12), 1183–1196.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, **29**(4), 1165–1188.
- Bhattacharya, P. K. (1974). Convergence of sample paths of normalized sums of induced order statistics. *The Annals of Statistics*, **2**(5), 1034–1039.
- Block, H. W., Savits, T. H. and Shaked, M. (1982). Some concepts of negative dependence. *The Annals of Probability*, **10**(3), 765–772.
- Block, H. W., Savits, T. H. and Shaked, M. (1985). A concept of negative dependence using stochastic ordering. *Statistics and Probability Letters*, **3**(2), 81–86.
- Carley, H. (2002). Maximum and minimum extensions of finite subcopulas. *Communications in Statistics-Theory and Methods*, **31**(12), 2151–2166.
- Chen, Y., Embrechts, P. and Wang, R. (2024a). An unexpected stochastic dominance: Pareto distributions, dependence, and diversification. *Operations Research*. <https://doi.org/10.1287/opre.2022.0505>.
- Chen, Y., Embrechts, P. and Wang, R. (2024b). Risk exchange under infinite-mean Pareto models. *arXiv: 2403.20171*.

- Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. John Wiley & Sons, New York.
- Csiszár, I. (1975).  $I$ -divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, **3**(1), 146–158.
- David, H. A. (1973). 18 Concomitants of order statistics. *Bulletin of the International Statistical Institute*, **45**(1), 295–300.
- de Amo, E., Carrillo, M. D., Durante, F. and Sánchez, J. F. (2017). Extensions of subcopulas. *Journal of Mathematical Analysis and Applications*, **452**(1), 1–15.
- Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, **93**(1), 40–57.
- Durante, F., Fernández-Sánchez, J., Quesada-Molina, J. J. and Ubeda-Flores, M. (2015). Convergence results for patchwork copulas. *European Journal of Operational Research*, **247**(2), 525–531.
- Esary, J. D., Proschan, F. and Walkup, D. W. (1967). Association of random variables, with applications. *Annals of Mathematical Statistics*, **38**(5), 1466–1474.
- Faugeras, O. P. (2017). Inference for copula modeling of discrete data: a cautionary tale and some facts. *Dependence Modeling*, **5**(1), 121–132.
- Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York.
- Geenens, G. (2020). Copula modeling for discrete random vectors. *Dependence Modeling*, **8**(1), 417–440.
- Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin*, **37**(2), 475–515.
- Genest, C., Nešlehová, J. G. and Rémillard, B. (2013). On the estimation of Spearman’s rho and related tests of independence for possibly discontinuous multivariate data. *Journal of Multivariate Analysis*, **117**, 214–228.
- Genest, C., Nešlehová, J. G. and Rémillard, B. (2017). Asymptotic behavior of the empirical multilinear copula process under broad conditions. *Journal of Multivariate Analysis*, **159**, 82–110.
- Genest, C., Nešlehová, J. G., Rémillard, B. and Murphy, O. A. (2019). Testing for independence in arbitrary distributions. *Biometrika*, **106**(1), 47–68.
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, **106**(4), 620–628.
- Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications. *Annals of Statistics*, **11**(1), 286–295.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.
- Joe, H. (2015). *Dependence Modeling with Copulas*. Chapman & Hall, London.
- Koliander, G., Pichler, G., Riegler, E. and Hlawatsch, F. (2016). Entropy and source coding for integer-dimensional singular random variables. *IEEE Transactions on Information Theory*, **62**(11), 6124–6154.
- Kuzmenko, V., Salam, R. and Uryasev, S. (2020). Checkerboard copula defined by sums of random variables. *Dependence Modeling*, **8**(1), 70–92.
- Lehmann, E. L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, **37**(5), 1137–1153.
- Li, D. X. (2000). On default correlation: A copula function approach. *Journal of Fixed Income*, **9**(4): 43–54.
- Lo, A. W., Wu, L., Zhang, R. and Zhao, C. (2024). Optimal impact portfolios with general dependence and marginals. *Operations Research*. <https://doi.org/10.1287/opre.2023.0400>.

- Marshall, A. W. (1996). Copulas, marginals, and joint distributions. In *Distributions with Fixed Marginals and Related Topics* (Seattle, WA, 1993), volume 28 of *IMS Lecture Notes-Monograph Series*, 213–222. Institute of Mathematical Statistics, Hayward, CA.
- McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton, NJ: Princeton University Press.
- Mesfioui, M. and Quessy, J. F. (2010). Concordance measures for multivariate non-continuous random vectors. *Journal of Multivariate Analysis*, **101**(10), 2398–2410.
- Millossovich, P., Tsanakas, A. and Wang, R. (2024). A theory of multivariate stress testing. *European Journal of Operational Research*, 318(3), 851–866.
- Moore, D. S. and Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Annals of Statistics*, **3**(3), 599–616.
- Nelsen, R. (2006). *An Introduction to Copulas*. Springer, New York, Second Edition.
- Nešlehová, J. (2004). *Dependence of Non-Continuous Random Variables*. Doctoral dissertation, Universität Oldenburg, Oldenburg, Germany.
- Nešlehová, J. (2007). On rank correlation measures for non-continuous random variables. *Journal of Multivariate Analysis*, **98**(3), 544–567.
- Perrone, E., Solus, L. and Uhler, C. (2019). Geometry of discrete copulas. *Journal of Multivariate Analysis*, **172**, 162–179.
- Piantadosi, J., Howlett, P. and Borwein, J. (2012). Copulas with maximum entropy. *Optimization Letters*, **6**, 99–125.
- Rüschendorf, L. (1981). Stochastically ordered distributions and monotonicity of the OC-function of sequential probability ratio tests. *Statistics: A Journal of Theoretical and Applied Statistics*, **12**(3), 327–338.
- Rüschendorf, L. (2009). On the distributional transform, Sklar’s theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, **139**(11), 3921–3927.
- Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg.
- Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the Expected Shortfall. *Management Science*, **67**, 1413–1429.