



Working Paper Series

Copula-based measurement of interdependence for discrete distributions

Martyna Kobus
Radosław Kurek

ECINEQ WP 2017 - 431

Copula-based measurement of interdependence for discrete distributions*

Martyna Kobus[†]

Polish Academy of Sciences, Poland

Radosław Kurek

Vistula University, Poland

Abstract

We focus on a question that has been long addressed in economics, namely, of one distribution being better than another according to a normative criterion. Our criterion distinguishes between interdependence and behaviour in the margins. Many economics contexts concern interdependence only e.g. complementarities in production function, intergenerational mobility, social gradient in health. We compare bivariate discrete distributions and measure interdependence via a most general measure, namely, a copula (Schweizer and Wolff 1981). For discrete distributions we need to overcome a problem of many copulas associated with a given distribution. Drawing on a copula theory (Carley 2002, Genest and Neslehova 2007) we solve this problem, chose a method to compare copulas which together with first-order stochastic dominance of marginal distributions gives the ordering to compare distributions. We provide a type of Hardy-Littlewood-Pólya result (Hardy et al. 1934), that is, we give implementable characterizations of this ordering (Theorems 1 - 3). As an application, we show how this ordering can be used to measure several phenomena that use either ordinal data (e.g. education-health gradient, bidimensional welfare) or simply discrete distributions (e.g. percentile income distributions of fathers and sons for intergenerational mobility). Welfare measures are easily decomposable into attributes and interdependence.

Keywords: multidimensional welfare, interdependence, ordinal data, copula function.

JEL Classification: D3, D6.

*This work was supported by National Science Centre in Poland i.e. grant 2011/01/N/HS4/01207 for dr Kobus.

[†]**Contact details:** M. Kobus (corresponding author), Institute of Economics, Polish Academy of Sciences, Nowy Swiat 72, 00-330 Warsaw, Poland; Email: mkobus@inepan.waw.pl. R. Kurek, Vistula University, Stokolosy Street 2, 02-787 Warsaw, Poland.

Introduction

Many economics phenomena and concepts are formally equivalent to measuring interdependence in random variables. This is obviously true in finance and insurance, where the value of the portfolio or insurance policy depends on the degree of interdependence of asset returns or insurance claims. The same applies to macroeconomics i.e. the presence of common shocks among firms/sectors, or to matching markets where the efficiency of matching mechanisms depends on the degree of interdependence between matched dimensions (Fernandez and Gali 1999, Lindenlaub 2017). Also, complementarity of the production function (Milgrom and Roberts 1995) is related to measuring interdependence between inputs and so is the measurement of multidimensional welfare and inequality (Atkinson and Bourguignon 1982) i.e. interdependence of well-being dimensions must be taken into account. In program evaluation literature, the estimation of distributional treatment effects concerns the dependence structure between the real and counterfactual distribution of outcomes. This dependence is not observed but can be studied via its bounds (Heckman et al. 1997, Fan and Park 2006).³ As highlighted recently by Chetty et al. (2014) relative intergenerational mobility is in fact interdependence of fathers' and sons' income distribution i.e. rank dependence. Furthermore, the study of the so called income-health gradients, education-health gradients, or more generally the gradient between socioeconomic status and health is the study of interdependence. The measurement of such gradients is an extensive research topic encompassing economics, medical sociology, social epidemiology, medicine and health psychology (see Smith 1999, Evans et al. 2012 for reviews). Other occurrences of interdependence in economics include, for example, preference alignment in coalition formation (Pycia 2012), ex-ante/ex-post reward schemes in groups (Ben-Porath et al. 1997), or the analysis of economics and social networks.

Typically used measures of interdependence, or association, or concordance,⁴ are based on correlation. Correlation is well-suited to capture interdependence in Gaussian settings, but beyond them it has many drawbacks. It picks up only linear association, its value depends on the marginal distributions, and it can be close to zero even in cases of strong dependence. Moreover, it is not scale free so it changes depending on how the data are aggregated. The latter makes it useless for ordinal data for which the only relevant information is the ordering and any monotone transformation of a given scale is equally good. It is easy to construct examples of the reversals of rankings based on correlation for such data (Allison and Foster 2004, Kobus 2015). Most flexible way to study non-Gaussian dependence structures is via a copula (Embrechts 2009).

A copula is a bivariate probability distribution function with uniform marginals. Copulas are well-known in mathematics and statistics due to the celebrated Sklar's theorem (Sklar, 1959) which states that a copula and marginal distributions char-

³Specifying the bounds of the treatment effects distribution is related to a copula theory (Frank et al. 1987). In particular, for discrete distributions of outcomes (e.g. life satisfaction and psychological indicators in assessing Moving to Opportunity (Ludwig et al. 2013)) these bounds can be improved using the results of Carley (2002) which we use a lot in the paper.

⁴We use dependence/association/concordance interchangeably, although they are all different concepts (Nelsen 2006); in our setting, however, we do not need a detailed differentiation.

acterise the joint distribution fully. Thus copula can be thought of as a method to construct joint distributions from marginal distributions. The advantage of using copulas is that one can separate the behaviour in the margins from pure association of random variables, which leads to more flexible modelling and more efficient estimation.⁵ Furthermore, copulas are scale-invariant, as Schweizer and Wolff (1981) note, “it is precisely the copula which captures those properties of the joint distribution which are invariant under (...) strictly increasing transformations.” This property is particularly useful for ordinal variables, that is, association does not depend on the way in which the variable is scaled, or in general, it does not depend on the coarsening of the support i.e. whether it is 1000 or 2000 income brackets. Given these favourable properties of copulas, it is not surprising that copulas are increasingly popular in economics. Fan and Patton (2014) reviews growing literature on the use of copula theory in econometric research, in particular in the area of multivariate modelling (e.g. earnings data (Bonhomme and Robin 2009), auction design (Hubbard et al. 2012) and partial identification (Fan et al. 2014)). Decancq (2012, 2014) is the first to introduce copulas into welfare-economic contexts to measure dependence between dimensions of well-being via measures based solely on copulas.⁶

In this paper we define an ordering to compare bivariate discrete distributions in terms of their interdependence and marginal distributions.⁷ This ordering is essentially bivariate first-order stochastic dominance (Atkinson and Bourguignon 1982), but presented as a collection of marginals and a copula and therefore called increasing concordance. Thus, one distribution dominates the other if it is more interdependent and its marginals first-order stochastically dominate the marginals of the other distribution. We provide the Hardy-Littlewood-Pólya (Hardy et al. 1934) type of result, namely, we prove the equivalence between three notions: (i) increasing concordance ordering (ii) regularity assumptions on the preferences i.e. the unanimity of utilitarian decision makers (social planners), and (iii) elementary transformations that reflect the reduction of interdependence and welfare. Ordinal data, that is data for which we only have the information of the ordering of categories (e.g. self-reported health status, self-declared life satisfaction, educational attainment) typically have discrete distributions.⁸ Such data appear a lot in various branches of economics (i.e. health economics, educational economics, happiness research) and we are mostly interested in comparing their distributions, although the results apply to any discrete

⁵In contrast, more traditional approaches to measuring dependence, use classical families of bivariate parametric distributions (normal, log-normal etc.), but this often has a lot of restrictions e.g. univariate distributions have to belong to the same parametric family, or only positive dependence can be modelled (Joe 2014).

⁶This is, however, for continuous distributions and does not apply to our setting.

⁷Going to more than two dimensions is not straightforward and merits a separate contribution. In two dimensions variables are positively associated (or concordant) if their values are all large together or small together; otherwise, we say, they are discordant or negatively concordant. In three and more dimensions equivalence between discordance and negative concordance breaks down; positive association is no longer minus negative association. The literature on measuring multivariate association is, however, well-established (Joe 1990, Nelsen 2006), and can be used to extend our setting.

⁸There are widely used ordinal indicators with continuous distributions e.g. test scores (Bond and Lang 2013), BMI index.

distributions e.g. to measuring mobility when only percentile income distributions are available. With respect to ordinal data this concerns applications such as measuring the relationship between socioeconomic status (SES) and health when, as is often the case, both SES and health are proxied by an ordinal indicator, or the measurement of welfare for multidimensional ordinal data. The latter is of particular interest because of the recent initiatives of governments (e.g. in UK, France, Germany, Italy, Korea, Spain, Mexico) and international organizations (e.g. OECD) to incorporate non-income dimensions in measuring well-being and progress. Many such dimensions are ordinal.

More specifically, we consider random vectors distributed on $\mathbb{I} := \mathbf{I}_1 \times \mathbf{I}_2 = \{0, \dots, i, \dots, n\} \times \{0, \dots, j, \dots, m\}$, where numbers i, j are arbitrary and we only require that the support preserves the ordering. To measure dependence in such a setting we need to solve the problem of unidentifiability because for discrete distributions copulas are only unique in the values of marginal distributions. Therefore, given two distributions one cannot compare them directly in terms of dependence because their copulas may be defined at different points. More precisely, there is a whole set of copulas associated with a given distribution. The bounds of this set were specified by Carley (2002), that is, each copula associated with a given distribution is between the so called minimum extension copula (i.e. the highest possible negative dependence) and the maximum extension copula (i.e. the highest possible positive dependence). We use this result to construct a so called concordance ordering. Namely, comparing to the standard notion of concordance (Joe 1990), here the whole set of copulas is compared through Carley's bounds. We combine concordance with first order stochastic dominance on marginals to define increasing concordance ordering, which is our main criterion for comparing distributions. It expresses both increasing interdependence and welfare in each dimension. Interdependence can be further summarised by utilising the bounds of well-known dependence measures (e.g. Kendall's tau, Spearman's rho) on the Carley's set (Genest and Neslehova 2007). This allows us to give an interval for the value of intergenerational mobility or education-health gradient, which may be easier to interpret than dominance itself. We show how our dominance criterion can be implemented (Theorem 1) i.e. if the dominance holds then a dominated distribution can be transformed into a dominant distribution via a sequence of probability mass transfers described in the theorem. These elementary transformations are essentially of two types, namely, transformations that affect marginal distributions, that is upward shifts, and transformations that affect only a copula, that is association-increasing switches (Epstein and Tanny 1980, Tchen 1980) but working on a copula distribution. These transformations have their matrix representation (Theorem 2). Increasing concordance is equivalent to the unanimity of welfare functions which are first order dominance - increasing and submodular (Theorem 3). Altogether, Theorems 1 - 3 provide a characterization of the increasing concordance ordering. Main technical difficulty is to show that upward shifts generate a set of copulas that are increasing in the sense of concordance ordering and not every algorithm of transforming marginal distributions possess this property.

We show how the obtained results can be applied to the measurement of education-health gradient, bidimensional welfare and intergenerational mobility. In other words, in all three cases we check what can be said about distributions comparisons when minimal structure is imposed on the data. Education-health gradient is the relation-

ship between education and health. Generally, the higher the education attainment the better the health status (Adler et al. 1994). Such socioeconomic gradients in health are observed in many industrialised countries (Marmot 2006). There is evidence for developing countries as well (Strauss and Thomas 1998). In empirical applications the socioeconomic status is often proxied by income, education, social or occupational class. The relationship between SES and health is typically studied via a concentration curve (van Doorslaer et al. 1997, van Doorslaer and Koolman 2004), however concentration curve is developed for ratio-scale variables and its use for other type of data is inappropriate (Makdisi and Yazbeck 2014). Zheng (2011) measures socioeconomic inequalities in health using bivariate health-income transition matrix, however, the distribution of income is fixed between groups that are compared, which makes this approach in fact unidimensional. Duclos and Echevin (2012) propose a robust method for measuring health-income gradient based on dominance conditions, but it does not separate interdependence and the gradient concerns interdependence only. Based on the National Health Interview Survey 2014 we calculate Kendall's tau lower and upper bound for bivariate distributions of educational attainment and several mental health indicators in four US regions. These bounds are wide, but we can declare West region as unambiguously better in almost all comparisons. There seems to be a steeper education gradient in depression than anxiety.

We then apply our results to the measurement of welfare in bivariate mental health distribution (the feeling of hopelessness and the feeling that everything is an effort in the past 30 days). To the best of our knowledge this is one of the first attempts to do this.⁹ We obtain welfare measures which are inevitably attribute decomposable (Abul Naga and Geoffard 2006), that is, we can decompose a given measure as a function of welfare in each margin and a measure of association and calculate the contributions of all three elements to overall welfare. Attribute decomposability reveals the detail needed to devise and implement policy i.e. think of two regions that have similar overall welfare score but very different contributions structure. They require different policy responses. We find that Midwest is the region with the highest welfare and West is the region with the lowest welfare. The differences between regions, when measured by a welfare index we construct, are small and the contributions of dimensions and association are similar.

As to intergenerational mobility we have a 100×100 centile transition matrix for the U.S. 1980-82 birth cohorts taken from Chetty et al. (2014).¹⁰ Here marginal distributions are the same (i.e. percentile distribution) and we only study interdependence i.e. what is the probability that a child is in a given percentile of his income distribution given that a father was in a given percentile of his income distribution? Interdependence is thus relative mobility and we calculate that it ranges between 0.21 and 0.25, as measured by Kendall's tau coefficient. Overall, the relation is positive although not very strong. Typically, one regresses ranks of children on the ranks of parents, but here we do not impose any restrictions on the shape of dependence. The value 0.21 can be thought of as a lower bound of relative mobility and the value of 0.25 as its upper bound. Furthermore, we analyse the geography of intergenerational

⁹Arndt et al. (2012) study first order stochastic dominance in multiple binary indicators.

¹⁰Available at www.equality-of-opportunity.org.

mobility based on the same dataset. We consider 50 largest community zones and we find that often they can be ranked unambiguously based on Kendall's tau bounds.

The paper is organised as follows. In Section 1 notation and definitions are introduced. In Section 2 we give a brief overview of the relevant parts of a copula theory. We then define an increasing concordance ordering. In Section 3 we prove Theorems 1-3 which provide characterisations of this ordering. Section 4 is devoted to applications. We conclude by pointing to future work.

1. Basic definitions and notation

We define a numerical representation of categories of ordinal variables $\mathbb{I} := \mathbf{I}_1 \times \mathbf{I}_2 = \{0, \dots, n\} \times \{0, \dots, m\}$ which is arbitrary as long as it preserves the ordering. $\mathbf{I}_1, \mathbf{I}_2$ are totally ordered sets and \mathbb{I} is endowed with the usual partial order: $(i, j) \leq (i', j')$ if and only if $i \leq i'$ and $j \leq j'$ for all i and j . Throughout the article \mathbb{I}, n, m are fixed.

Now let f be a probability distribution on the set \mathbb{I} . By defining probability distribution f on \mathbb{I} we make it independent of scale; that is, if there are two different scales with the same number of categories on each dimension, then \mathbb{I} does not change and a given probability distribution can be related to both scales. Obviously we require $\sum_{i=0}^n \sum_{j=0}^m f_{ij} = 1$ and for all $(i, j) \in \mathbb{I}$, $f_{ij} \geq 0$. We define marginal distributions by

$$f_i^1 := \sum_{j=0}^m f_{ij} \quad f_j^2 := \sum_{i=0}^n f_{ij} \quad (1)$$

and cumulative distributions by

$$F_i^1 := \sum_{k=0}^i f_k^1 \quad F_j^2 := \sum_{l=0}^j f_l^2 \quad (2)$$

A multidimensional cumulative distribution function F at (i, j) equals

$$F_{ij} := \sum_{k=0}^i \sum_{l=0}^j f_{kl} \quad (3)$$

Furthermore, we define survival function of distribution f by

$$S_{ij} := \sum_{k=i+1}^n \sum_{l=j+1}^m f_{kl} \quad (4)$$

and marginal survival functions by

$$S_i^1 := \sum_{k=i+1}^n f_k^1 \quad S_j^2 := \sum_{l=j+1}^m f_l^2 \quad (5)$$

Let Λ denote a set of probability distributions with given marginals on all dimensions.

2. Increasing concordance ordering

The increasing concordance ordering we propose combines first order stochastic dominance for marginal distributions and association ordering for copulas. We start with association.

As mentioned in the Introduction, the concept of increasing interdependence is well-captured by an ordering on copulas. Copulas were popularised by Sklar (1959) to study the dependence structure between random variables. From Sklar's theorem the copula of a distribution F is the only information necessary to recover F from its marginal distributions F^1, F^2 . Formally, a bi-dimensional copula $C : [0, 1]^2 \mapsto [0, 1]$ is a function such that

$$F_{ij} = C(F_i^1, F_j^2). \quad (6)$$

Given (6) and utilising probability integral transform, one can think of a copula as a bivariate distribution with uniform marginals. An interesting fact about copulas which is often exploited in econometrics literature (Manski 1988, 1997; Heckman 1997; Hoderlein and Stoye 2014) is Fréchet-Hoeffding inequality which states that for every $(u, v) \in [0, 1]^2$ we have

$$\max(u + v - 1, 0) \leq C \leq \min(u, v)$$

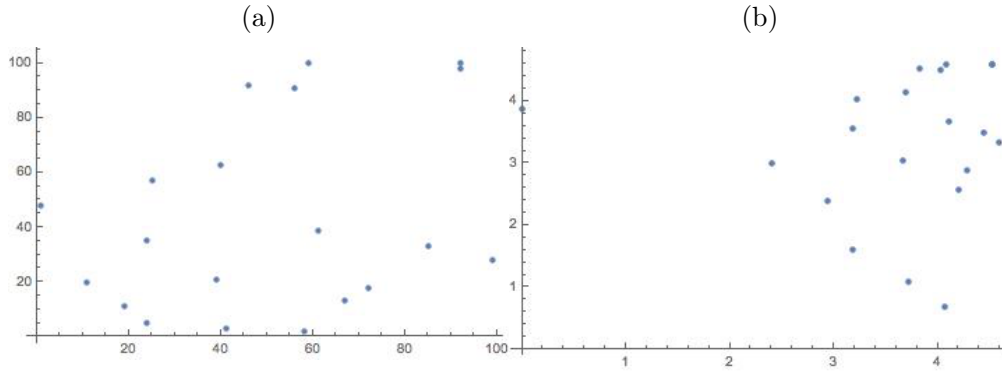
Thus, $M(u, v) = \min(u, v)$ and $W(u, v) = \max(u + v - 1, 0)$ are known as, respectively, Fréchet-Hoeffding upper and lower bound.

Let us consider an example which shows the difference between copulas and cumulative distributions. We assume for a moment that F is a probability distribution on a given scale. There are two dimensions, health status and educational attainment, each with two categories, to which apply scales (1, 2) and (2, 5), respectively (distribution (a) in Figure 1). The value of the copula at, for example, (0.5, 0.5) is the value of F at (1, 2). Because $F^1(1) = 0.5, F^2(2) = 0.5$ we get $C(0.5, 0.5) = 0.5$. Now if we change the scale in which we measure health from 1 and 2 into, respectively, 10 and 20 (distribution (b)) the cdf changes from $F(1, 2) = 0.5$ into $F(10, 20) = 0$, whereas the copula does not change, namely, the copula is still evaluated at (0.5, 0.5) and still equals 0.5. This is what is meant by saying that the copula is invariant to increasing transformations of variables. That the cdf is not is even more clear when looking at Figure 2. The monotone transformation (ln) completely changes the dependence structure. Copula is “designed” exactly for situations when such re-labelling (re-scaling) should not change the underlying distribution. Therefore, it is a particularly useful concept to measure association between variables for which only ordering matters.

For discrete distributions we run into problems of unidentifiability of a copula of a given distribution. This is because Sklar's theorem ensures the uniqueness of a copula only for $\text{Ran}F^1 \times \text{Ran}F^2$, which in case of ordinal data is a proper subset of $[0, 1]^2$. Formally, copulas for discrete variables are subcopulas i.e. $C' : S_1 \times S_2 \mapsto S_3$, where S_1, S_2, S_3 are all subsets of $[0, 1]$ containing 0 and 1. Therefore, there are many functions C that fulfil equation (6). Any subcopula, however, can be extended to a copula, which is a key step in the proof of Sklar's theorem (Nelsen 2006). Given a bivariate distribution F let \mathbb{C}_F be a set of copulas for which equation (6) holds, namely, it is a set of possible extension copulas. Its lower and upper bounds were specified by Carley (2002).

Figure 1: The difference between copula and cdf: an example.

(a) Distribution (a)				(b) Copula of (a)				(c) Distribution (b): transformed health categories			
	Health 1	Health 2			$\frac{1}{2}$	1			Health 10	Health 20	
Edu 2	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$		Edu 2	$\frac{1}{2}$	0	$\frac{1}{2}$
Edu 5	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$		Edu 5	0	$\frac{1}{2}$	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$		1	$\frac{1}{2}$	1			$\frac{1}{2}$	$\frac{1}{2}$	

Figure 2: A scatter plot of (x_i, y_i) sampled randomly from $\{0, 1, \dots, n\} \times \{0, 1, \dots, n\}$ (a) and $(\ln(x_i), \ln(y_i))$ (b).**Proposition 1. Upper and lower bound on \mathbb{C}_F (Carley 2002)**

Let \mathbb{C}_F be the set of copulas which fulfil equation (6) for a given distribution F on \mathbb{I} . The upper Carley bound¹¹ for the set \mathbb{C}_F is given by

$$C_F^+(u, v) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \max(0, \min(u - \alpha_{ij}, v - \beta_{ij}, f_{ij})) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \min_x(u - \alpha_{ij}, v - \beta_{ij}, f_{ij})$$

and the lower Carley bound is equal to

$$C_F^-(u, v) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \max(0, -f_{ij} + \min(u - \gamma_{ij}, f_{ij}) + \min(v - \delta_{ij}, f_{ij}))$$

where

$$\alpha_{ij} = \sum_{k=0}^{i-1} f_k^1 + \sum_{l=0}^{j-1} f_{il}, \quad \beta_{ij} = \sum_{l=0}^{j-1} f_l^2 + \sum_{k=0}^{i-1} f_{kj}$$

¹¹Formally, both upper bounds and lower bounds are shuffles of min denoted by minx (Mikusiński et al. 1992).

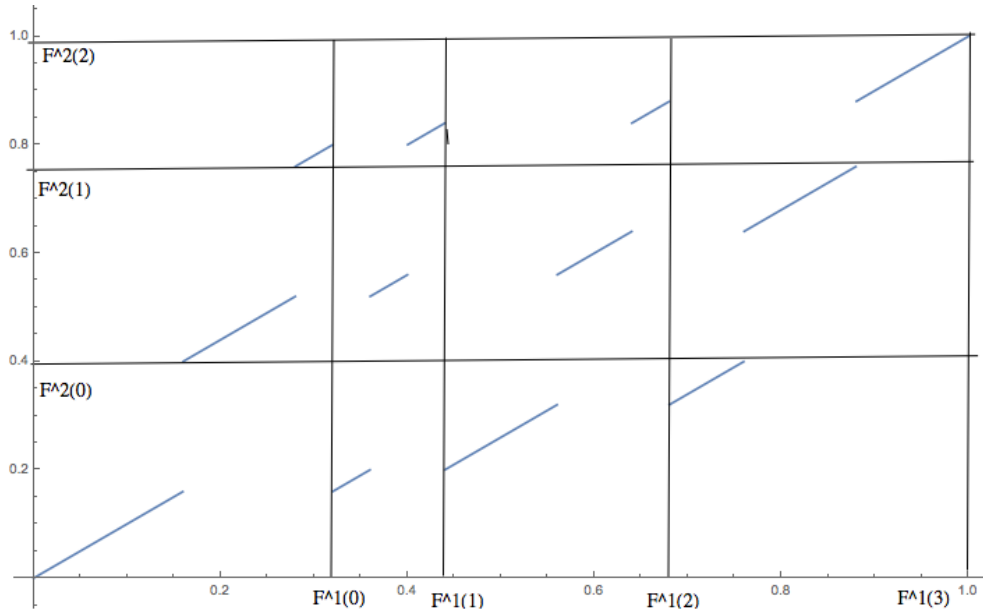
$$\gamma_{ij} = \sum_{k=0}^{i-1} f_k^1 + \sum_{l=j+1}^{\infty} f_{il}, \quad \delta_{ij} = \sum_{l=0}^{j-1} f_l^2 + \sum_{k=i+1}^{\infty} f_{kj}$$

and by convention, an empty sum is equal to zero.

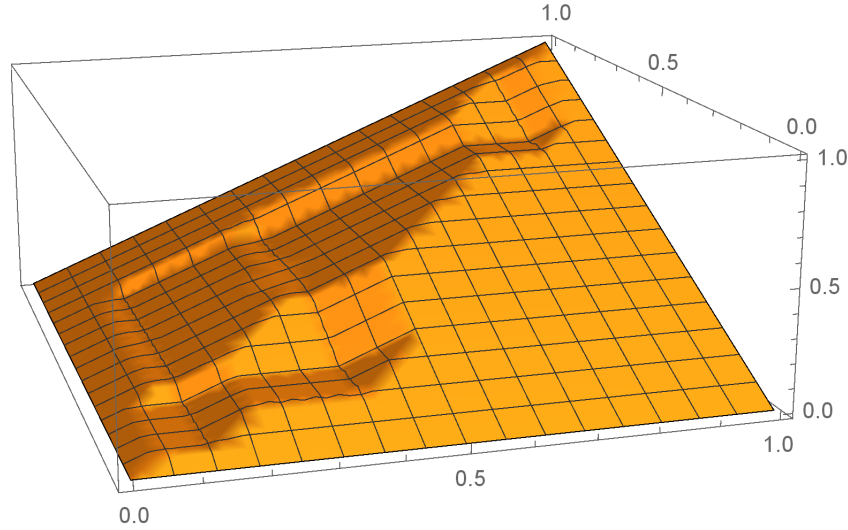
Remark. $\alpha_{(i+1)0} = \gamma_{(i+1)m} = F_i^1$ and $\beta_{0(j+1)} = \delta_{n(j+1)} = F_j^2$

The interpretation of α_{ij} , β_{ij} , γ_{ij} and δ_{ij} is the following. The values of marginal distributions divide $[0, 1]^2$ into blocks $B_{ij} := [i-1, i] \times [j-1, j]$. If we order them from bottom to top, left to right, α_{ij} is the amount of mass in the blocks up to block B_{ij} ; similarly for β_{ij} if we order them left to right, bottom to top. Analogously, if we order them from top to bottom, left to right, γ_{ij} is the amount of mass in the blocks before B_{ij} ; similarly for δ_{ij} if we order them right to left, bottom to top. The upper (lower) Fréchet-Hoeffding bound (copula M) is generated by spreading the mass uniformly along the diagonal (counter-diagonal). Here, for example, for the maximum extension copula, the mass concentrated in each block is spread uniformly along the diagonal as close as to the origin as possible to get the fastest growth of the copula. The diagonals along which the mass is distributed cannot overlap, because projected on the marginals they have to be uniform. Thus for a given distribution the set of its extension copulas might be a proper subset of the set of copulas bounded by Fréchet-Hoeffding bounds. Fréchet-Hoeffding bounds belong to the set of extension copulas for a given distribution only if at the points of uniqueness (values of subcopula) the value of the subcopula agrees with copula M or W e.g. $C''(0.2, 0.3) = 0.2 = \min(0.2, 0.3) = M(0.2, 0.3)$. Figure 3 shows a typical distribution of mass for C_F^+ and Figure 4 shows a maximum extension copula in three dimensions.

Figure 3: A typical distribution of mass (support) for Carley upper bound C_F^+



Let association be captured by \ll which denotes a concordance ordering on the set of copulas. It is a point-wise partial ordering (Nelsen 2006 Definition 2.8.1) such that one copula has higher values than the other for all $(u, v) \in [0, 1]^2$. To make this definition meaningful in a discrete setting we focus on the Carley's set. We first

Figure 4: Maximum extension copula C_F^+ 

define a rotation operation which will be useful in drawing links between maximum and minimum extension copulas.

Definition 1. *Rotation operator, distribution version*

Let $R : \Lambda \rightarrow \Lambda$ be the following rotation operator

$$R(f)(i, j) = f_{j(n-i)}.$$

Operator R rotates distribution matrix of f by 90 degrees clockwise. We will write Rf instead of $R(f)$ and R^2f instead of $R(R(f))$.

Definition 2. *Rotation operator (on the support)*

Let $R : \mathcal{S} \rightarrow \mathcal{S}$ be the following rotation operator

$$R(s)(u, v) = s(v, 1 - u)$$

where \mathcal{S} is set of mass distributions (supports) of C_F^+ and C_F^- .

It follows from Lemma 1 that \mathcal{S} is a well-defined codomain of R . Operator R rotates support of $C_F^+(C_F^-)$ by 90 degrees clockwise.

Lemma 1. Let $s(C_F^+)(s(C_F^-))$ be mass distribution (support) of $C_F^+(C_F^-)$, then

1. $Rs(C_f^+) = s(C_{Rf}^-)$
2. $Rs(C_f^-) = s(C_{Rf}^+)$

Proof. Slightly abusing notation (C_f^+ instead of C_F^+), let us take $\tilde{f} = Rf$. We have

$$\begin{aligned} \alpha_{ij} &= \sum_{k=0}^{i-1} f_k^1 + \sum_{l=0}^{j-1} f_{il} = \sum_{k=0}^{i-1} \tilde{f}_{n-k}^2 + \sum_{l=0}^{j-1} \tilde{f}_{l(n-i)} = \sum_{k=n-i+1}^n \tilde{f}_k^2 + \sum_{l=0}^{j-1} \tilde{f}_{l(n-i)} = \\ &= 1 - \sum_{k=0}^{n-i-1} \tilde{f}_k^2 - \sum_{l=j}^m \tilde{f}_{l(n-i)} = 1 - \delta_{j(n-i)} - \tilde{f}_{j(n-i)} \end{aligned}$$

$$\beta_{ij} = \sum_{l=0}^{j-1} f_l^2 + \sum_{k=0}^{i-1} f_{kj} = \sum_{l=0}^{j-1} \tilde{f}_l^1 + \sum_{k=0}^{i-1} \tilde{f}_{j(n-k)} = \sum_{l=0}^{j-1} \tilde{f}_l^1 + \sum_{k=n-i+1}^n \tilde{f}_{jk} = \tilde{\gamma}_{j(n-i)}$$

It is easy to see that each part of the support of C_f^+ , namely line from $(\alpha_{ij}, \beta_{ij})$ to $(\alpha_{ij} + f_{ij}, \beta_{ij} + f_{ij}) = (\alpha_{i(j+1)}, \beta_{(i+1)j})$ is transformed into the line from $(\tilde{\gamma}_{j(n-i)}, \tilde{\delta}_{j(n-i)} + \tilde{f}_{j(n-i)})$ to $(\tilde{\gamma}_{j(n-i-1)}, \tilde{\delta}_{(j+1)(n-i)} + \tilde{f}_{(j+1)(n-i)}) = (\tilde{\gamma}_{j(n-i)} + \tilde{f}_{j(n-i)}, \tilde{\delta}_{j(n-i)})$, which is part of the support of $C_{\tilde{f}}^-$. Since R is a linear operator and s is the sum of indicators of its parts, statement 1 holds. Statement 2 can be proven similarly. \square

We can use the rotation operator twice to define survival function.

Definition 3. We will call a following function \tilde{C} the survival copula

$$\tilde{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$$

and we have

$$S_{ij} = \tilde{C}(S_i^1, S_j^2)$$

Definition 3 is motivated by the following relation

$$\begin{aligned} S_{ij} &= \sum_{k=i+1}^n \sum_{l=j+1}^m f_{kl} = 1 - \sum_{k=0}^i f_k^1 - \sum_{l=0}^j f_l^2 + \sum_{k=0}^i \sum_{l=0}^j f_{kl} = \\ &= 1 - F_i^1 - F_j^2 + F_{ij} = 1 - (1 - S_i^1) - (1 - S_j^2) + F_{ij} = \\ &= S_i^1 + S_j^2 - 1 + C(F_i^1, F_j^2) = S_i^1 + S_j^2 - 1 + C(1 - S_i^1, 1 - S_j^2). \end{aligned}$$

We have the following remark.

Remark 1. Survival function is a distribution function of $R^2 f$. Hence we have the following results

- $\widetilde{C}_f^+ = C_{R^2 f}^+$
- $\widetilde{C}_f^- = C_{R^2 f}^-$
- $\widetilde{C}_F^+ = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \max(0, \min(\alpha_{i(j+1)} - u, \beta_{(i+1)j} - v, f_{ij}))$
- $\widetilde{C}_F^- = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \max(0, -f_{ij} + \min(\gamma_{i(j-1)} - u, f_{ij}) + \min(\delta_{(i-1)j} - v, f_{ij}))$

using convention that $\alpha_{i(m+1)} = \alpha_{(i+1)0}$, $\beta_{(n+1)j} = \beta_{0(j+1)}$, $\alpha_{n(m+1)} = \beta_{(n+1)m} = 1$, $\gamma_{i(-1)} = \gamma_{(i+1)m}$, $\delta_{(-1)j} = \delta_{n(j+1)}$ and $\gamma_{n(-1)} = \delta_{(-1)m} = 1$.

We are now ready to define an ordering on copulas of discrete distributions.

Definition 4. Concordance

Let C'_F be a unique subcopula of F and let C'_G be a unique subcopula of G . We say that C'_G is more max-concordant/min-concordant/concordant than C'_F i.e. $C'_F \ll C'_G$, if and only if:

1. max-concordant
 $C_F^+(u, v) \leq C_G^+(u, v)$ for all $(u, v) \in [0, 1]^2$
2. min-concordant
 $C_F^-(u, v) \geq C_G^-(u, v)$ for all $(u, v) \in [0, 1]^2$
3. concordant - both 1. and 2. hold

4. *survival-max-concordant*
 $\widetilde{C}_F^+(u, v) \geq C_G^+(u, v)$ for all $(u, v) \in [0, 1]^2$
5. *survival-min-concordant*
 $\widetilde{C}_F^-(u, v) \leq C_G^-(u, v)$ for all $(u, v) \in [0, 1]^2$
6. *survival-concordant* - both 4. and 5. hold

The specific \ll we focus on is induced by a well-known rank correlation coefficient Kendall's τ . Its bounds were specified by Genest and Neslehova (2007) so that for a given distribution its degree of interdependence can be summarised in an interval.

Proposition 2. *Upper and lower bounds for Kendall's τ (Genest and Neslehova 2007)*

$$\tau(C_F^+) = 1 - 4 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{i-1} \sum_{l=j+1}^{\infty} f_{ij} f_{kl} \quad (7)$$

$$\tau(C_F^-) = -1 + 4 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{i-1} \sum_{l=0}^{j-1} f_{ij} f_{kl} \quad (8)$$

$$\tau(\widetilde{C}_F^+) = 1 - 4 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=i+1}^{\infty} \sum_{l=0}^{j-1} f_{ij} f_{kl} \quad (9)$$

$$\tau(\widetilde{C}_F^-) = -1 + 4 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=i+1}^{\infty} \sum_{l=j+1}^{\infty} f_{ij} f_{kl} \quad (10)$$

Lemma 2. $\tau(C_F^+) = \tau(\widetilde{C}_F^+)$ and $\tau(C_F^-) = \tau(\widetilde{C}_F^-)$

Proof.

$$\begin{aligned} \sum_{i=0}^{\infty} \sum_{k=i+1}^{\infty} &= \sum_{k>i} = \sum_{i<k} = \sum_{k=0}^{\infty} \sum_{i=0}^{k-1} \\ \sum_{j=0}^{\infty} \sum_{l=0}^{j-1} &= \sum_{l<j} = \sum_{j>l} = \sum_{l=0}^{\infty} \sum_{j=l+1}^{\infty} \\ \tau(\widetilde{C}_F^+) &= 1 - 4 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=i+1}^{\infty} \sum_{l=0}^{j-1} f_{ij} f_{kl} = 1 - 4 \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{i=0}^{k-1} \sum_{j=l+1}^{\infty} f_{ij} f_{kl} = \tau(C_F^+) \\ \tau(\widetilde{C}_F^-) &= -1 + 4 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=i+1}^{\infty} \sum_{l=j+1}^{\infty} f_{ij} f_{kl} = -1 + 4 \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{i=0}^{k-1} \sum_{j=0}^{l-1} f_{ij} f_{kl} = \tau(C_F^-) \end{aligned}$$

□

So far we considered interdependence only. To complete the definition of the ordering that enables us to compare whole distributions we require first-order stochastic dominance to hold on each marginal distribution.

Definition 5. *First order stochastic dominance*

Fixing $n \geq 1$ and allowing f^j, g^j to be two probability distributions on \mathbf{I}_j .

$$f^j \leq_{FSD} g^j \Leftrightarrow F_i^j \leq G_i^j \text{ for any } i$$

We combine FSD and association measured by the ordering on copulas and define the relation \lesssim_{CFSD} as follows

Definition 6. Increasing concordance (CFSD)

1. $g \lesssim_{C^+FSD} f \iff f^1 \leq_{FSD} g^1, f^2 \leq_{FSD} g^2$ and C'_G is more max-concordant than C'_F
2. $g \lesssim_{C^-FSD} f \iff f^j \leq_{FSD} g^j, f^2 \leq_{FSD} g^2$ and C'_G is more min-concordant than C'_F
3. $g \lesssim_{CFSD} f \iff f^j \leq_{FSD} g^j, f^2 \leq_{FSD} g^2$ and C'_G is more concordant than C'_F
4. $g \lesssim_{\widetilde{C^+}FSD} f \iff f^1 \leq_{FSD} g^1, f^2 \leq_{FSD} g^2$ and C'_G is more survival-max-concordant than C'_F
5. $g \lesssim_{\widetilde{C^-}FSD} f \iff f^j \leq_{FSD} g^j, f^2 \leq_{FSD} g^2$ and C'_G is more survival-min-concordant than C'_F
6. $g \lesssim_{\widetilde{C}FSD} f \iff f^j \leq_{FSD} g^j, f^2 \leq_{FSD} g^2$ and C'_G is more survival-concordant than C'_F

Distribution g is more interdependent than f in the sense of \lesssim_{CFSD} when g is dominated by f according to FSD on marginals and has higher association than f . Relation \lesssim_{CFSD} is reflexive because of reflexivity of FSD relation and it is also transitive, thus it is a quasi-ordering.

3. Characterisation theorems

The following two definitions describe elementary transfers of probability mass that transform a dominated distribution into a dominant distribution. First is an upward shift, well-known to characterise first-order stochastic dominance, and second is an association-increasing transformation introduced independently by Epstein and Tanny (1980) and Tchen (1980). The latter is defined on a copula density, that is,

$$c(i, j) = C(i, j) - C(i - 1, j) - C(i, j - 1) + C(i - 1, j - 1). \quad (11)$$

Definition 7. Upward shift

We say that F^k differs from G^k by an elementary shift if there exist $i_1 < i_2$, such that

$$f_i^k - g_i^k = \begin{cases} \epsilon & \text{if } i = i_1 \\ -\epsilon & \text{if } i = i_2 \\ 0 & \text{otherwise} \end{cases}$$

Definition 8. Association-increasing transformation

We say that C_G differs from C_F by an association-increasing transformation if there exist $i_1 < i_2$ and $j_1 < j_2$ such that

$$c_F(i, j) - c_G(i, j) = \begin{cases} \epsilon, & \text{if } (i, j) = (i_1, j_1) \text{ or } (i_2, j_2) \\ -\epsilon, & \text{if } (i, j) = (i_1, j_2) \text{ or } (i_2, j_1) \\ 0, & \text{otherwise} \end{cases}$$

Please note that association switches do not change the marginal distributions.

We are now ready to prove our main result.

Theorem 1. *Let F, G be a distribution with marginals F^1, F^2 and G^1, G^2 , respectively. Further, let C_F^+, C_G^+ (C_F^-, C_G^-) be their maximal (minimal) extension copulas. $F \preceq_{CFSD} G$ is equivalent to the following condition:*

F can be transformed into G via a finite sequence of elementary shifts and correlation switching transformations i.e. a single step k in the sequence generates a distribution \tilde{F}_k such that it dominates distribution \tilde{F}_{k-1} according to \preceq_{CFSD} .

Instead of \preceq_{CFSD} one can also put \preceq_{C^+FSD} , \preceq_{C^-FSD} , $\preceq_{\tilde{C}FSD}$, $\preceq_{\tilde{C}^+FSD}$, or $\preceq_{\tilde{C}^-FSD}$.

A priori upward shifts that are defined for univariate distributions can be distributed along the other marginal in an arbitrary manner and can produce distributions whose copulas will not be in relation \ll with the copula of the initial distribution e.g. adding $\delta_i = F_i^1 - G_i^1$ to points $(F_i^1, 1)$ for every i will produce distributions that are not more concordant. Therefore, the main difficulty in the proof is to construct an algorithm such that each upward shift generates a distribution whose maximal extension copula is more concordant. We utilise greedy algorithm. Let us consider the first marginal and let i_0 be the lowest category for which $F_{i_0}^1 < G_{i_0}^1$. Let us consider the maximal extension copula of F . It is well-defined in all points of $[0, 1]^2$, in particular in $G_{i_0}^1$. This copula determines a discrete distribution \tilde{F}^1 with the first marginal such that $F_{i_0}^1$ is replaced by $\tilde{F}_{i_0}^1 = G_{i_0}^1$ i.e. following an upward shift.¹² In particular, the maximal extension copula determines how upward shift is distributed along the second dimension. Upward shift involves a move from left to right adding the mass concentrated in-between, which given the distribution of mass in C_F^+ means adding the mass from the bottom (up to category l for the second dimension). This generates a new distribution with a maximal extension copula that is more concordant.

As an example let us consider the following distribution f

f	Category 1	Category 2	Category 3	Category 4
Category D	20	5	0	0
Category C	5	10	5	10
Category B	10	0	10	5
Category A	0	5	10	5

We consider the transfers of 5 from third to the second column, starting from category A. We get extension copulas with supports shown in Figure 5.

Similarly for C^- we consider a distribution \tilde{f} which is a rotated distribution f

\tilde{f}	Category A	Category B	Category C	Category D
Category 1	0	10	5	20
Category 2	5	0	10	5
Category 3	10	10	5	0
Category 4	5	10	5	0

Then, by doing transfers of 5 from third column to the second, starting from category A, we get extension copulas with supports shown in Figure 6.

Elementary transformations have matrix representation.

¹²For simplicity we assume for now that there is enough mass in $i_0 + 1$ to equalise $\tilde{F}_{i_0}^1$ with $G_{i_0}^1$. In general, mass may have to be transferred from more than one category.

Figure 5: An example of a sequence of upward shifts distributed according to C_F^+

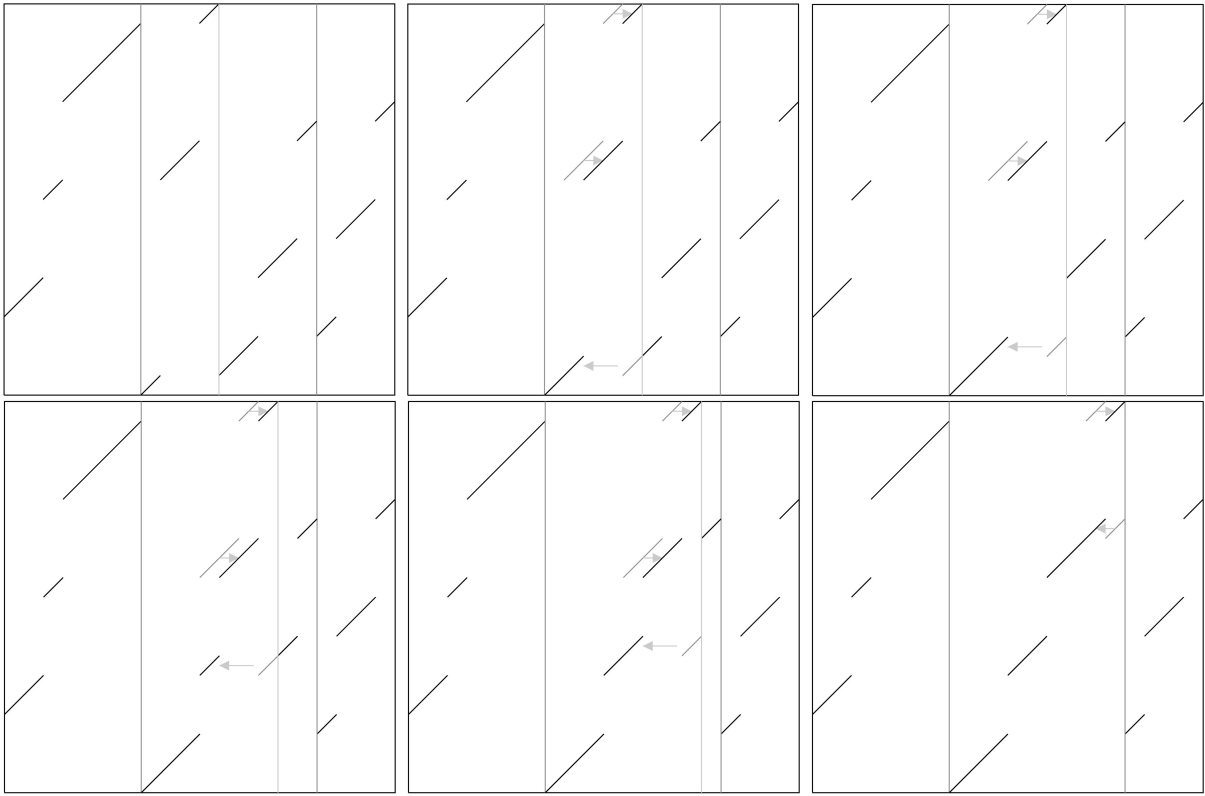
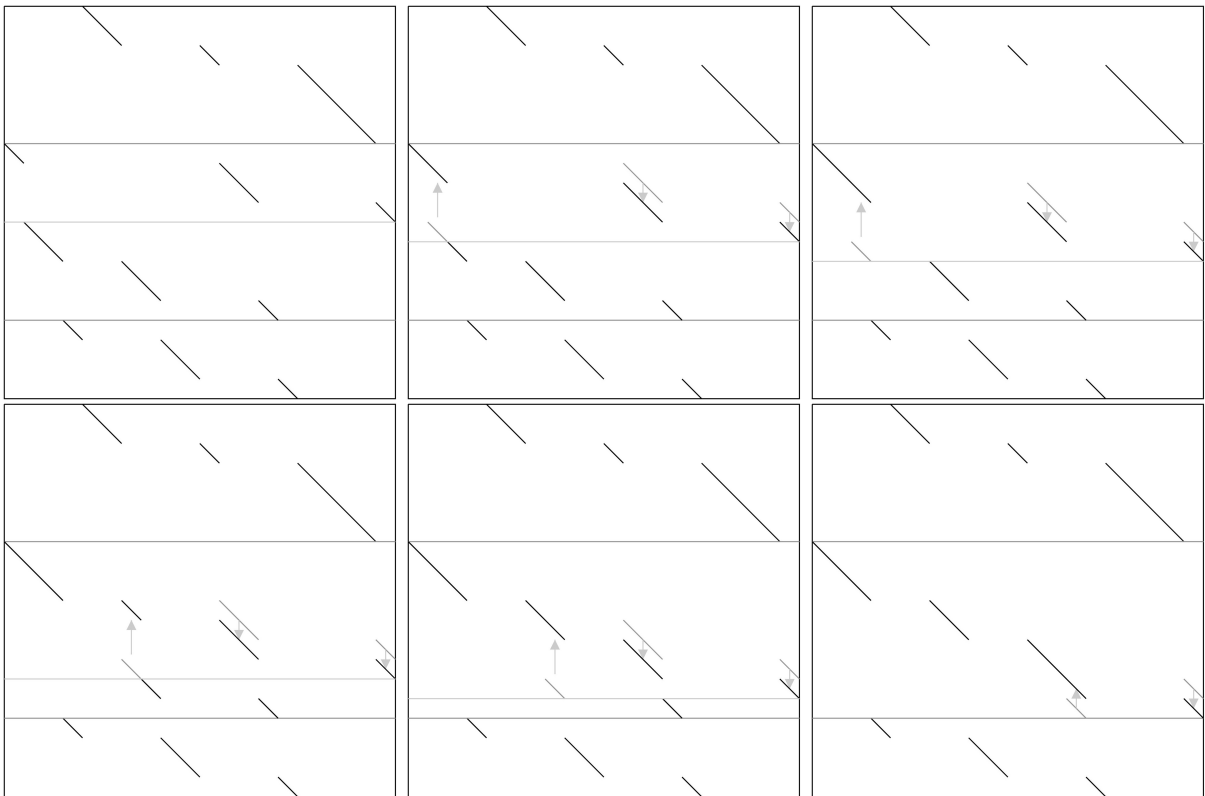


Figure 6: An example of a sequence of upward shifts distributed according to C_F^-



Theorem 2. Let F, G be a distribution with marginals F^1, F^2 and G^1, G^2 , respectively. Further, let C_F^+, C_G^+ (C_F^-, C_G^-) be their maximal(minimal) extension copulas.

$F \preceq_{CFSD} G$ is equivalent to the following condition.

There exist $n \times m$ matrices γ and ϕ such that $f + \gamma + \phi = g$, where γ is a matrix of elementary shifts and ϕ is a matrix of association-increasing transformations.

Similarly to standard results in welfare measurement literature we can relate $CFSD$ relation to a class of welfare functions. Let \mathcal{W} denote the class of functions defined on \mathbb{I} , $w \in \mathcal{W}$ denote its element and

$$\mathbb{E}(w|f) := \sum_{i,j} w(i,j) f_{ij}$$

denote the expected value of w given distribution f .

Theorem 3. Let F, G be a distribution with marginals F^1, F^2 and G^1, G^2 , respectively. Let \mathcal{W}_1^- and \mathcal{W}_1^+ denote functions which are submodular and supermodular, respectively. Let \mathcal{W}_2 be a class of nondecreasing functions with respect to both coordinates.

$F \preceq_{CFSD} G$ is equivalent to

$$\mathbb{E}(w|f) \leq \mathbb{E}(w|g) \text{ for all } w \in \mathcal{W}_1^- \cap \mathcal{W}_2.$$

$F \preceq_{\tilde{CFSD}} G$ is equivalent to

$$\mathbb{E}(w|f) \leq \mathbb{E}(w|g) \text{ for all } w \in \mathcal{W}_1^+ \cap \mathcal{W}_2.$$

Sklar's theorem allows us to characterise functions which are monotone with respect to \preceq_{CFSD} i.e. so called Schur-convex functions.

Theorem 4. Order-preserving functions $P : \Lambda \mapsto \mathbb{R}$ is a function that preserves \preceq_{CFSD} i.e.

$$(f \preceq_{CFSD} g) \Rightarrow (P(f) \leq P(g)). \quad (12)$$

if and only if

$$P(f) = h(C_f^+, \mathbf{p}^1, \mathbf{p}^2), \quad (13)$$

where h is an increasing function. Along the first coordinate it increases with respect to concordance ordering \ll and on the coordinates 2,3 with respect to first-order stochastic dominance.

4. Applications

4.1. Education and health

Below we analyse the relationship between educational attainment and five mental health indicators based on the National Health Interview Survey from year 2014. NHIS is an annual cross-section health survey in the US. The indicators are respondents' 4-category answers to the following questions: During the past 30 days how often did you feel (1) nervous (2) that everything was an effort (3) hopeless (4) restless or fidgety (5) worthless? Typically such relationship is analysed using probit regression, however, Bong and Lang (2014) point out that such cardinalizations are often not justified and lead to conclusions which are easy to reverse. Without going into much detail of their critique, here we check what is the minimum that can be said

about the relationship between education and health when no structure is imposed on the data. We simply calculate Kendall's tau for the Carley's set and compare US regions. For 4-category ordinal variables, these bounds are wide, however, this does not prevent us from declaring West as the region with the lowest education-health gradients in all comparisons and South as the region that dominates in most comparisons (Table 1). It is interesting to notice that association differs substantially for different indicators, that is, for the feelings of nervousness and restlessness it is much lower than for the feelings of hopelessness and worthlessness ($|0.5|$ vs. $|0.8|$). The former feelings are more characteristic of anxiety, whereas the latter are more characteristic of depression. There seems to be a steeper education gradient in depression than anxiety.

Table 1: Kendall's tau bounds for Education-Health distributions

Region	nervous-edu	effort-edu	hopeless-edu	restless-edu	worthless-edu
Northwest	0,5533 -0,5577	0,6761 -0,5985	0,8383 -0,7621	0,5939 -0,5439	0,8845 -0,8186
Midwest	0,5504 -0,5683	0,6684 -0,6086	0,8372 -0,7782	0,5740 -0,5572	0,8657 -0,8164
South	0,5943 -0,5968	0,6875 -0,6030	0,8497 -0,7648	0,6212 -0,5626	0,8911 -0,8199
West	0,5466 -0,5610	0,6329 -0,5932	0,8251 -0,7374	0,5673 -0,5420	0,8506 -0,7924
All	0,5625 -0,5740	0,6655 -0,5998	0,8378 -0,7592	0,5900 -0,5519	0,8728 -0,8110

4.2. Welfare measurement for ordinal data

CFSD relation can be used to rank bivariate distributions of ordinal data according to welfare. As an example of how this works, we consider bivariate health distribution where two dimensions are the feeling of hopelessness and the feeling that everything is an effort in the past 30 days based on the NHIS 2014. We find that Midwest dominates West and that West is dominated by the general population distribution (Table 2). There are in fact a lot of cases of copula dominance but *CFSD* collapses because of lack of dominance on marginal distributions. However, when it comes to the degree of interdependence between two health indicators West exhibits less and South exhibits more association than other regions.

Table 2: Bivariate health distribution and *CFSD* dominance.

Region	Northwest	Midwest	South	West	All
Northwest		\ll $>_{FSD^2}$	\ll	\gg $<_{FSD^1}$	\gg $<_{FSD^1}$
Midwest	\gg $<_{FSD^2}$		\ll_{C^+} \gg_{C^-} $<_{FSD^2}$	$>_{CFSD}$	\gg $<_{FSD^1}$ $<_{FSD^2}$
South	\gg	\gg_{C^+} \ll_{C^-} $>_{FSD^2}$		\gg $<_{FSD^2}$	\gg
West	\ll $>_{FSD^1}$	\ll $>_{FSD^1}$ $>_{FSD^2}$	\ll $>_{FSD^2}$		\ll $>_{FSD^1}$ $>_{FSD^2}$
All	\ll $>_{FSD^1}$	\ll $>_{FSD^1}$ $>_{FSD^2}$	\ll	\gg $<_{FSD^1}$ $<_{FSD^1}$	

Apart from dominance comparisons, it is standard in welfare and inequality measurement literature to use measures. Here, measures have the advantage of being

attribute decomposable, whereas a priori it is difficult to show whether a given measure has this property. Decomposability (Shorrocks 1984) is a desired property of welfare, inequality and poverty measures as it allows for a disaggregated analysis. For an additive function h , Theorem 4 states that functions preserving \preceq_{CAF} are necessarily a special case of attribute decomposable functions (Naga and Geoffard 2006).

Definition 9. Attribute decomposability (Naga and Geoffard 2006)

Function $P : \Lambda \mapsto \mathbb{R}$ is attribute decomposable if and only if there exist univariate inequality indices P_1, P_2 , a measure of association via copula κ and increasing functions g_1, g_2, g_3 such that

$$P(f) = g_1(P_1(f^1)) + g_2(P_2(f^2)) + g_3(\kappa(C_F)) \quad (14)$$

Putting $\tilde{g}_1 = g_1 \circ P_1$, $\tilde{g}_2 = g_2 \circ P_2$, and $\tilde{g}_3 = g_3 \circ \kappa$ we obtain decomposition (13) for an additive function h , and $\tilde{g}_1, \tilde{g}_2, \tilde{g}_3$ behave like h with respect to coordinates. With an additive attribute decomposable index, one can evaluate the contributions of both dimensions and association to overall welfare i.e. taking $\frac{g_1(P_1(f^1))}{P(f)}$. An example of an attribute decomposable welfare measure is the following.

Definition 10. Attribute decomposable welfare measure for ordinal data

$$P(f) = \frac{1}{3} \left(\frac{\mathbb{E}(\tilde{w}|f^1)}{n} + \frac{\mathbb{E}(\bar{w}|f^2)}{m} + \frac{\tau(C_F) + 1}{2} \right), \quad (15)$$

where we put $P_1(f^1) := \mathbb{E}(\tilde{w}|f^1)$, $\tilde{w}(i, j) = i$, $P_2(f^2) := \mathbb{E}(\bar{w}|f^2)$, $\bar{w}(i, j) = j$ and $\kappa(C_F^+) = \tau(C_F^+)$, $\kappa(C_F^-) = -\tau(C_F^-)$ is given by (7).¹³

According to index (15) Midwest has the highest welfare, then its South, Northwest and West, which is the only region for which welfare is lower than welfare in the general population. The differences between regions are small. The values of marginal indices and association are similar, therefore their contributions are roughly equal.

Region	$P_1(f^1)$	$P_2(f^2)$	$\kappa(C_F^+)$	$\kappa(C_F^-)$	$h(C_F^+, f^1, f^2)$	$h(C_F^-, f^1, f^2)$
Northwest	0.8974	0.9518	0.9659	0.6381	0.9440	0.8894
Midwest	0.8996	0.9558	0.9669	0.6493	0.9463	0.8934
South	0.8962	0.9533	0.9678	0.6408	0.9445	0.8900
West	0.8904	0.9485	0.9605	0.6008	0.9397	0.8798
All	0.8955	0.9523	0.9653	0.6308	0.9435	0.8877

4.3. Intergenerational mobility

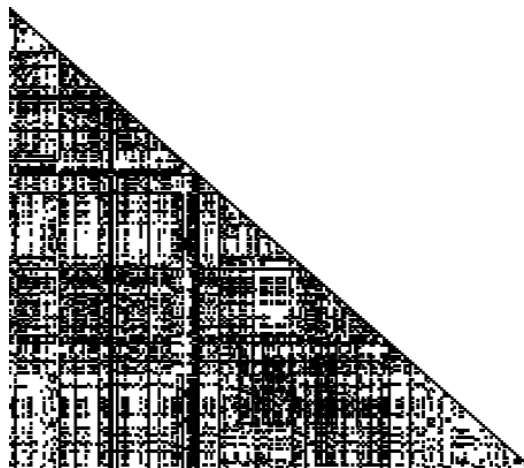
In this section we analyse the 100×100 centile transition matrix for the U.S. 1980-82 birth cohorts taken from Chetty et al. (2014).¹⁴ Here marginal distributions are the same (i.e. percentile distribution) and we only study interdependence i.e. what is the probability that a child is in a given percentile of his income distribution given that a father was in a given percentile of his income distribution? Interdependence is thus relative mobility and we calculate that it ranges between 0.21 and 0.25, as measured

¹³We put $-\tau(C_F^-)$ because the ordering of minimum extension copulas is reversed comparing to FSD.

¹⁴Available at www.equality-of-opportunity.org.

by Kendall's tau coefficient on the Carley's set. Overall, the relation is positive although not very strong. Typically, one regresses ranks on children on the ranks of parents, but here we check what can be said without imposing any restrictions on the shape of dependence. We obtain narrow interval because of a very dense partition of the grid. The value 0.21 can be thought of as a lower bound of relative mobility and the value of 0.25 as its upper bound. Based on the same dataset, we also analyse the geography of intergenerational mobility. We consider 50 largest community zones and we find a lot of comparability between, marked as dark places in Figure 7. These regions can be ranked based on Kendall's tau intervals (which do not cross) in terms of intergenerational mobility. For other types of mobility (e.g. educational mobility, occupational status mobility), where marginal distributions are different, (15) can be used to evaluate the relative importance of welfare in each generation's distribution and of mobility.

Figure 7: Comparability between 50 largest community zones in the US



Concluding remarks

For the case of bivariate discrete distribution we formulate a standard result which characterizes first order stochastic dominance in terms of elementary transformations and a class of welfare functions. The novelty comes from the fact that we explicitly treat interdependence separately from the marginal distributions. Such a setting captures well some problems that appear mostly for ordinal data. We find a method to deal with technical difficulties imposed by a discrete model. There is an inherent uncertainty about the value of dependence for discrete distributions. Other (i.e. than first order stochastic dominance) dominance relations on marginals can be combined with copulas and our proof and description of copulas for discrete distributions makes it easier to study such new relations. Also more work is needed to construct multidimensional measures of inequality based on multivariate concordance literature. Furthermore, there may be alternative definitions of multidimensional welfare ordering, and dependence does not have to be explicit. Nevertheless this may compromise attribute decomposability.

References

- Abul Naga R H, Geoffard P Y. 2006. Decomposition of bivariate inequality indices by attributes, *Economics Letters*, Elsevier, 90(3):362-367.
- Adler N, Boyce T, Chesney M. 1994. Socioeconomic status and health: The challenge of the gradient, *the American Psychologist* 49(1): 15-24.
- Allison R A, Foster J E. 2004. Measuring health inequality using qualitative data, *Journal of Health Economics*; 23(3):505-524.
- Arndt C, Distant R, Hussain M A, Oesterdal L P. 2012. Ordinal Welfare Comparisons with Multiple Discrete Indicators: A First Order Dominance Approach and Application to Child Poverty, *World Development* 40(11): 2290-2301.
- Atkinson A B, Bourguignon F. 1982. The Comparison of MultiDimensioned Distributions of Economic Status, *Review of Economic Studies*, 49(2):183-201.
- Bond T, Lang K. 2013. The Evolution of the Black-White Test Score Gap in Grades K-3: The Fragility of Results, *The Review of Economics and Statistics* 95(5): 1468-1479.
- Bond T, Lang K. 2014. The Sad Truth About Happiness Scales, NBER Working Papers 19950, National Bureau of Economic Research, Inc.
- Bonhomme S, Robin J M. 2009. Assessing the equalizing force of mobility using short panels: France, 1990-2000, *Review of Economic Studies* 76: 63-92.
- Ben-Porath E, Gilboa I, Schmeidler D. 1997. On the Measurement of Inequality under Uncertainty, *Journal of Economic Theory* 75: 194-204.
- Carley H. 2002. Maximum and minimum extensions of finite subcopulas, *Communications in Statistics - Theory and Methods* 31(12): 2151-2166.
- Chetty R, Hendren N, Kline P, Saez E. 2014. Is the United States a Land of Opportunity? Recent trends in intergenerational mobility, *American Economic Review: Papers&Proceedings* 104(5): 141-147.
- Decancq K. 2012. Elementary multivariate rearrangements and stochastic dominance on a Fréchet class, *Journal of Economic Theory* 147(4): 1450-1459.
- Decancq K. 2014. Copula-based measurement of dependence between dimensions of well-being, *Oxford Economic Papers* 66(3): 681-701.
- Duclos J Y, Echevin D. 2012. Health and income: A robust comparison of Canada and the US, *Journal of Health Economics* 30: 293-302.
- Embrechts P. 2009. Copulas: A Personal View, *Journal of Risk and Insurance* 76: 639-650.
- Epstein L G, Tanny S M. 1980. Increasing generalized correlation: a definition and some economic consequences, *The Canadian Journal of Economics* 13(1): 16-34.
- Evans W, Wolfe B, Adler N. 2012. The SES-health gradient: a brief review of the literature, in: *The Biological Consequences of Socioeconomic Inequalities*, Chapter 1, Wolfe B, Evans W, Seeman T (eds.), Russell Sage Foundation.
- Fan Y, Park S S. 2006. Sharp bounds on the distribution of treatment effects and their statistical inference, *Econometric Theory* 26: 931-951.
- Fan Y, Patton A J. 2014. Copulas in econometrics, *Annual Review of Economics* 6: 179-200.
- Fan Y, Sherman R, Shum M. 2014. Identifying treatment effects under data combination, *Econometrica* 82: 811-822.

- Fernandez R, Gali J. 1999. To each according to . . . ? markets, tournaments, and the matching problem with borrowing constraints, *Review of Economic Studies* 66: 799–824.
- Genest Ch, Neslehova J. 2007. A primer on copulas for count data, *ASTIN Bulletin: The Journal of the IAA* 37(2): 475–515.
- Hardy G H, Littlewood J E, Pólya P. 1934. *Inequalities*, Cambridge University Press, Cambridge.
- Heckman J, Smith J, Clements N. 1997. Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts, *Review of Economic Studies* 64: 487–535.
- Hubbard T, Li T, Paarsch HJ. 2012. Semiparametric estimation in models of first-price, sealed-bid auctions with affiliation, *Journal of Econometrics* 168:4–16.
- Joe H. 1990. Multivariate concordance, *Journal of Multivariate Analysis* 35: 12–30.
- Joe H. 2014. *Dependence modelling with copulas*, Chapman and Hall CRC, Monographs on Statistics and Applied Probability 134.
- Kobus M, 2015. Polarization measurement for ordinal data, *Journal of Economic Inequality* 13(2): 275–297.
- Lindenbaum I. 2017. Sorting multidimensional types: theory and application, *Review of Economic Studies* (forthcoming).
- Ludwig J, Duncan G J, Gennetian L A, Katz L F, Kessler R C, Kling J R, Sanbonmatsu L. 2013. Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity, *American Economic Review* 103(3): 226–31.
- Makdisi P, Yazbeck M. 2014. Measuring socioeconomic health inequalities in the presence of multiple categorical information, *Journal of Health Economics* 34: 84–95.
- Marmot M. 2006. Status Syndrome: A Challenge to Medicine, *Journal of the American Medical Association* 295(11): 1304–07.
- Mikusiński P, Sherwood H, Taylor M D. 1992. Shuffles of Min, *Stochastica* 13: 61–74.
- Milgrom P, Roberts J. 1990. The Economics of Modern Manufacturing: Technology, Strategy, and Organization, *American Economic Review* 80: 511–528.
- Nelsen R B. 2006. *An Introduction to Copulas*. 2nd edition, New York: Springer-Verlag.
- Pycia M. 2012. Stability and Preference Alignment in Matching and Coalition Formation, *Econometrica* 80(1): 323–362.
- Schweizer B, Wolff E F. 1981. On nonparametric measures of dependence for random variables, *Annals of Statistics*, 9:870–885.
- Shorrocks, A.F., 1984. Inequality decomposition by population subgroups. *Econometrica* 52 (6), 1369–1385.
- Sklar A. 1959. Fonctions de repartition an dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris*, 8:229–231.
- Smith J P. 1999. Healthy bodies and thick wallets: The dual relation between health and economic status, *Journal of Economic Perspectives* 13(2): 145–166.
- Strauss J, Thomas D. 1998. Health, nutrition, and economic development, *Journal of Economic Literature* 36(2): 766–817.
- Tchen A H. 1980. Inequalities for distributions with given marginals, *The Annals of Probability* 8(4): 814–827.
- van Doorslaer E, Koolman X. 2004. Explaining the differences in income-related health inequalities across European countries, *Health Economics* 13:609–628.

van Doorslaer E, Wagstaff A, Bleichrodt H, Calonge S, Gerdtham U, Gerfin M, Geurts J, Gross L, Hakkinen U, Leu R. 1997. Income-related inequalities in health: some international comparisons, *Journal of Health Economics* 16: 93–112.

Zheng B. 2011. A new approach to measure socioeconomic inequality in health, *Journal of Economic Inequality* 9: 555–577.

Appendix

Theorem 1

Proof. We will start by proving the following lemma.¹⁵

Lemma 3. *Let i_0 be the lowest category for which $F_{i_0}^1 \neq G_{i_0}^1$ i.e. $F_{i_0}^1 < G_{i_0}^1$, and let i_1 be the lowest category such that $F_{i_1}^1 > F_{i_0}^1$. Consider a distribution \tilde{F} such that $\tilde{F}_{ij} = F_{ij}$ for all $i \neq i_0$, $\tilde{F}^2 = F^2$, $\tilde{F}_{i_0}^1 = \min\{F_{i_1}^1, G_{i_0}^1\}$, $\tilde{F}_i^1 = F_i^1$ elsewhere and $\tilde{f}_{i_0j} \geq f_{i_0j}$. Then, $C_{\tilde{F}}^+(F_{i_0}^1, F^2) \geq C_F^+(F_{i_0}^1, F^2)$.*

Proof. Our goal is to show that the following inequality holds

$$C_{\tilde{F}}^+(F_{i_0}^1, F_k^2) \geq C_F^+(F_{i_0}^1, F_k^2)$$

$$\begin{aligned} C_{\tilde{F}}^+(F_{i_0}^1, F_k^2) - C_F^+(F_{i_0}^1, F_k^2) &= \\ &= \sum_{i,j} \min x(F_{i_0}^1 - \tilde{\alpha}_{ij}, F_k^2 - \tilde{\beta}_{ij}, \tilde{f}_{ij}) - \min x(F_{i_0}^1 - \alpha_{ij}, F_k^2 - \beta_{ij}, f_{ij}) =^{16} \\ &= \sum_{i=0}^{i_0} \sum_{j=0}^k \min x(F_{i_0}^1 - \tilde{\alpha}_{ij}, F_k^2 - \tilde{\beta}_{ij}, \tilde{f}_{ij}) - \min x(F_{i_0}^1 - \alpha_{ij}, F_k^2 - \beta_{ij}, f_{ij}) =^{17} \\ &= \sum_{i=0}^{i_0} \sum_{j=0}^k \min x(F_{i_0}^1 - \tilde{\alpha}_{ij}, \tilde{f}_{ij}) - \min x(F_{i_0}^1 - \alpha_{ij}, f_{ij}) =^{18} \\ &= \sum_{i=0}^{i_0} \sum_{j=0}^k \min x(F_{i_0}^1 - \tilde{\alpha}_{ij}, \tilde{f}_{ij}) - f_{ij} = \\ &= \sum_{i=0}^{i_0-1} \sum_{j=0}^k (\tilde{f}_{ij} - f_{ij}) + \sum_{j=0}^k \left[\min x(F_{i_0}^1 - \tilde{\alpha}_{i_0j}, \tilde{f}_{i_0j}) - f_{i_0j} \right] = \\ &= \tilde{F}_{(i_0-1)k} - F_{(i_0-1)k} + \sum_{j=0}^k \left[\min x(F_{i_0}^1 - \tilde{\alpha}_{i_0j}, \tilde{f}_{i_0j}) - f_{i_0j} \right] = \\ &= \sum_{j=0}^k \left[\min x(F_{i_0}^1 - \tilde{\alpha}_{i_0j}, \tilde{f}_{i_0j}) - f_{i_0j} \right] \end{aligned}$$

Let $r = \sup\{r' : F_{i_0}^1 - \tilde{\alpha}_{i_0r'} \geq \tilde{f}_{i_0r'}\}$. Let $\Delta F = F_{i_0}^1 - F_{i_0-1}^1$. Since $F_{i_0}^1 - \tilde{\alpha}_{i_0r} = \Delta F + \tilde{F}_{i_0-1}^1 - \tilde{\alpha}_{i_0r} = \Delta F - \sum_{l=0}^{r-1} \tilde{f}_{i_0l}$, then if $\Delta F - \sum_{l=0}^{r-1} \tilde{f}_{i_0l} \geq \tilde{f}_{i_0r}$ then it is true for any $r' < r$. For $r' > r + 1$ we have $F_{i_0}^1 - \tilde{\alpha}_{i_0r'} < 0$.¹⁸ Finally we obtain:

$$\sum_{j=0}^k \min x(F_{i_0}^1 - \tilde{\alpha}_{i_0j}, \tilde{f}_{i_0j}) = \begin{cases} \sum_{j=0}^k \tilde{f}_{i_0j} & \text{if } k \leq r \\ \sum_{j=0}^r \tilde{f}_{i_0j} + \max(0, F_{i_0}^1 - \tilde{\alpha}_{i_0(r+1)}) & \text{if } k \geq r + 1 \end{cases}$$

¹⁵We study elementary shifts on distribution F^1 , but it can be as well F^2 .

¹⁶Let us keep in mind that $\alpha_{(i_0+1)0} = F_{i_0}^1$, $\beta_{0(k+1)} = F_k^2$ and α_{ij} , β_{ij} are nondecreasing, and that $F_{i_0}^1 < G_{i_0}^1 = \tilde{\alpha}_{(i_0+1)0}$

¹⁷ $F_k^2 - \beta_{ij} = \sum_{l=0}^k f_l^2 - \sum_{l=0}^{j-1} f_l^2 - \sum_{t=0}^{i-1} f_{tj} = \sum_{l=j+1}^k f_l^2 + \sum_{t=i}^n f_{tj} > f_{ij}$, and we get the same if we replace F with \tilde{F} .

¹⁸It is so, because $F_{i_0}^1 - \tilde{\alpha}_{i_0(r+1)} < \tilde{f}_{i_0(r+1)}$ and $\tilde{\alpha}_{i_0r'} = \tilde{\alpha}_{i_0(r'-1)} + \tilde{f}_{i_0(r'-1)}$.

If $k \leq r$ then obviously $\sum_{j=0}^k \min(F_{i_0}^1 - \tilde{\alpha}_{i_0j}, \tilde{f}_{i_0j}) = \sum_{j=0}^k \tilde{f}_{i_0j} \geq \sum_{j=0}^k f_{i_0j}$.
 Let $k \geq r+1$. We want to show that $\sum_{j=0}^r \tilde{f}_{i_0j} + \max(0, F_{i_0}^1 - \tilde{\alpha}_{i_0(r+1)}) \geq \sum_{j=0}^k f_{i_0j}$, but it is sufficient to show it for $k = m$. We have $F_{i_0}^1 - \tilde{\alpha}_{i_0(r+1)} \geq 0$ because $F_{i_0}^1 - \tilde{\alpha}_{i_0r} \geq \tilde{f}_{i_0r}$. We have $\sum_{j=0}^r \tilde{f}_{i_0j} + F_{i_0}^1 - \tilde{\alpha}_{i_0(r+1)} = \sum_{j=0}^r \tilde{f}_{i_0j} + F_{i_0-1}^1 + \sum_{j=0}^m f_{i_0j} - \tilde{F}_{i_0-1}^1 - \sum_{j=0}^r \tilde{f}_{i_0j} = \sum_{j=0}^m f_{i_0j}$ which concludes the proof of the lemma. \square

Now we will prove similar lemma for C_F^- .

Lemma 4. *Let i_0 be the lowest category for which $F_{i_0}^1 \neq G_{i_0}^1$ i.e. $F_{i_0}^1 < G_{i_0}^1$, and let i_1 be the lowest category such that $F_{i_1}^1 > F_{i_0}^1$. Consider a distribution \tilde{F} such that $\tilde{F}_{ij} = F_{ij}$ for all $i \neq i_0$, $\tilde{F}^2 = F^2$, $\tilde{F}_{i_0}^1 = \min\{F_{i_0}^1, G_{i_0}^1\}$, $\tilde{F}_i^1 = F_i^1$ elsewhere and $\tilde{f}_{i_0j} \geq f_{i_0j}$. Then, $C_{\tilde{F}}^-(F_{i_0}^1, F^2) \leq C_F^-(F_{i_0}^1, F^2)$.*

Proof. Our goal is to show that the following inequality holds

$$C_{\tilde{F}}^-(F_{i_0}^1, F_k^2) \leq C_F^-(F_{i_0}^1, F_k^2)$$

$$\begin{aligned} C_{\tilde{F}}^-(F_{i_0}^1, F_k^2) - C_F^-(F_{i_0}^1, F_k^2) &= \\ &= \sum_{i,j} [\max(0, -\tilde{f}_{ij} + \min(F_{i_0}^1 - \tilde{\gamma}_{ij}, \tilde{f}_{ij}) + \min(F_k^2 - \tilde{\delta}_{ij}, \tilde{f}_{ij})) + \\ &\quad - \max(0, -f_{ij} + \min(F_{i_0}^1 - \gamma_{ij}, f_{ij}) + \min(F_k^2 - \delta_{ij}, f_{ij}))] =^{19} \\ &= \sum_{i=0}^{i_0} \sum_{j=0}^k [\max(0, -\tilde{f}_{ij} + \min(F_{i_0}^1 - \tilde{\gamma}_{ij}, \tilde{f}_{ij}) + \min(F_k^2 - \tilde{\delta}_{ij}, \tilde{f}_{ij})) - f_{ij}] + \\ &\quad + \sum_{i=i_0+1}^n \sum_{j=0}^k \max(0, -\tilde{f}_{ij} + \min(F_{i_0}^1 - \tilde{\gamma}_{ij}, \tilde{f}_{ij}) + \min(F_k^2 - \tilde{\delta}_{ij}, \tilde{f}_{ij})) + \\ &\quad + \sum_{i=0}^{i_0} \sum_{j=k+1}^m \max(0, -\tilde{f}_{ij} + \min(F_{i_0}^1 - \tilde{\gamma}_{ij}, \tilde{f}_{ij}) + \min(F_k^2 - \tilde{\delta}_{ij}, \tilde{f}_{ij})) + \\ &\quad + \sum_{i=i_0+1}^n \sum_{j=k+1}^m \max(0, -\tilde{f}_{ij} + \min(F_{i_0}^1 - \tilde{\gamma}_{ij}, \tilde{f}_{ij}) + \min(F_k^2 - \tilde{\delta}_{ij}, \tilde{f}_{ij})) =^{20} \\ &= \sum_{i=0}^{i_0-1} \sum_{j=0}^k (\tilde{f}_{ij} - f_{ij}) + \sum_{i=0}^{i_0-1} \sum_{j=k+1}^m \max(0, F_k^2 - \tilde{\delta}_{ij}) + \\ &\quad + \sum_{j=0}^k [\max(0, \min(F_{i_0}^1 - \tilde{\gamma}_{i_0j}, \tilde{f}_{i_0j})) - f_{i_0j}] + \\ &\quad + \sum_{j=k+1}^m \max(0, -\tilde{f}_{i_0j} + \min(F_{i_0}^1 - \tilde{\gamma}_{i_0j}, \tilde{f}_{i_0j}) + F_k^2 - \tilde{\delta}_{i_0j}) = \\ &= \tilde{F}_{(i_0-1)k} - F_{(i_0-1)k} + \sum_{j=0}^k [\max(0, \min(F_{i_0}^1 - \tilde{\gamma}_{i_0j}, \tilde{f}_{i_0j})) - f_{i_0j}] + \\ &\quad + \sum_{j=k+1}^m \max(0, -\tilde{f}_{i_0j} + \min(F_{i_0}^1 - \tilde{\gamma}_{i_0j}, \tilde{f}_{i_0j}) + F_k^2 - \tilde{\delta}_{i_0j}) = \\ &= \sum_{j=0}^k [\max(0, \min(F_{i_0}^1 - \tilde{\gamma}_{i_0j}, \tilde{f}_{i_0j})) - f_{i_0j}] + \\ &\quad + \sum_{j=k+1}^m \max(0, -\tilde{f}_{i_0j} + \min(F_{i_0}^1 - \tilde{\gamma}_{i_0j}, \tilde{f}_{i_0j}) + F_k^2 - \tilde{\delta}_{i_0j}) \end{aligned}$$

Let $r = \inf\{r' : F_{i_0}^1 - \tilde{\gamma}_{i_0r'} \geq \tilde{f}_{i_0r'}\}$. Let $\Delta F = F_{i_0}^1 - F_{i_0-1}^1$. Since $F_{i_0}^1 - \tilde{\gamma}_{i_0r} = \Delta F + \tilde{F}_{i_0-1}^1 - \tilde{\gamma}_{i_0r} = \Delta F - \sum_{l=r+1}^m \tilde{f}_{i_0l}$, then if $\Delta F - \sum_{l=r+1}^m \tilde{f}_{i_0l} \geq \tilde{f}_{i_0r}$, then it is true for any $r' > r$. For $r' < r-1$ we have $F_{i_0}^1 - \tilde{\gamma}_{i_0r'} < 0$.²¹ Finally we obtain the following. If $k < r-1$ we have

¹⁹Let us keep in mind that $\gamma_{(i_0+1)m} = F_{i_0}^1$, $\beta_{n(k+1)} = F_k^2$, γ_{ij} is nondecreasing and nonincreasing with respect to i and j , respectively, and δ_{ij} is nondecreasing and nonincreasing with respect to j and i , respectively.

²⁰Recall that $\tilde{F}^2 = F^2$ and $\tilde{F}_{i_0}^1 > F_{i_0}^1 = F_{i_0-1}^1 + f_{i_0}^1 = \tilde{F}_{i_0-1}^1 + f_{i_0}^1$.

²¹It is so, because $F_{i_0}^1 - \tilde{\gamma}_{i_0(r-1)} < \tilde{f}_{i_0(r-1)}$ and $\tilde{\gamma}_{i_0(r'-1)} = \tilde{\gamma}_{i_0r'} + \tilde{f}_{i_0r'}$.

$$C_{\tilde{F}}^-(F_{i_0}^1, F_k^2) - C_F^-(F_{i_0}^1, F_k^2) = -(\sum_{j=0}^k f_{i_0j}) + \max(0, -\tilde{f}_{i_0(r-1)} + F_{i_0}^1 - \tilde{\gamma}_{i_0(r-1)} + F_k^2 - \tilde{\delta}_{i_0(r-1)}) + \\ + \sum_{j=r}^m \max(0, F_k^2 - \tilde{\delta}_{i_0j}) = -\sum_{j=0}^k f_{i_0j} \leq 0$$

and if $k \geq r - 1$ we get

$$C_{\tilde{F}}^-(F_{i_0}^1, F_k^2) - C_F^-(F_{i_0}^1, F_k^2) = \max(0, F_{i_0}^1 - \tilde{\gamma}_{i_0(r-1)}) + \\ + \sum_{j=r}^k \tilde{f}_{i_0j} - \sum_{j=1}^k f_{i_0j} + \sum_{j=k+1}^m \max(0, F_k^2 - \tilde{\delta}_{i_0j}) = \\ = \max(0, F_{i_0}^1 - \tilde{\gamma}_{i_0(r-1)}) + \sum_{j=r}^k \tilde{f}_{i_0j} - \sum_{j=1}^k f_{i_0j} +$$

We have $F_{i_0}^1 - \tilde{\gamma}_{i_0r} \geq \tilde{f}_{i_0r}$ so $F_{i_0}^1 - \tilde{\gamma}_{i_0(r-1)} \geq 0$ and we obtain

$$C_{\tilde{F}}^-(F_{i_0}^1, F_k^2) - C_F^-(F_{i_0}^1, F_k^2) = F_{i_0}^1 - \tilde{\gamma}_{i_0(r-1)} + \sum_{j=r}^k \tilde{f}_{i_0j} - \sum_{j=1}^k f_{i_0j} = \\ = f_{i_0}^1 - \sum_{j=r}^m \tilde{f}_{i_0j} + \sum_{j=r}^k \tilde{f}_{i_0j} - \sum_{j=1}^k f_{i_0j} = \sum_{j=k+1}^m f_{i_0j} - \sum_{j=k+1}^m \tilde{f}_{i_0j} \leq 0$$

If $k \leq r$, then obviously $\sum_{j=0}^k \min_x(F_{i_0}^1 - \tilde{\alpha}_{i_0j}, \tilde{f}_{i_0j}) = \sum_{j=0}^k \tilde{f}_{i_0j} \geq \sum_{j=0}^k f_{i_0j}$. Let $k \geq r + 1$. We want to show that $\sum_{j=0}^r \tilde{f}_{i_0j} + \max(0, F_{i_0}^1 - \tilde{\alpha}_{i_0(r+1)}) \geq \sum_{j=0}^k f_{i_0j}$, but it is sufficient to show it for $k = m$. We have $F_{i_0}^1 - \tilde{\alpha}_{i_0(r+1)} \geq 0$ because $F_{i_0}^1 - \tilde{\alpha}_{i_0r} \geq \tilde{f}_{i_0r}$. We have $\sum_{j=0}^r \tilde{f}_{i_0j} + F_{i_0}^1 - \tilde{\alpha}_{i_0(r+1)} = \sum_{j=0}^r \tilde{f}_{i_0j} + F_{i_0-1}^1 + \sum_{j=0}^m f_{i_0j} - \tilde{F}_{i_0-1}^1 - \sum_{j=0}^r \tilde{f}_{i_0j} = \sum_{j=0}^m f_{i_0j}$ which concludes the proof of the lemma. \square

Thus, after an elementary shift on F^1 which is distributed along F^2 according to a maximum (minimum) extension copula we obtain a copula $C_{\tilde{F}_1}^+$ ($C_{\tilde{F}_1}^-$) which is more concordant than the initial copula C_F^+ (C_F^-).

In each step the marginal of F is transformed to equalize the marginal of G . Continuing this procedure we arrive at a marginal G^1 (and analogously at a marginal G^2) obtaining in each step i a more concordant copula $C_{\tilde{F}_i}^+$ i.e. $C_{\tilde{F}_{i-1}}^+ \preceq_{C^+FSD} C_{\tilde{F}_i}^+ \preceq_{C^+FSD} C_G^+$ ($C_{\tilde{F}_{i-1}}^- \preceq_{C^-FSD} C_{\tilde{F}_i}^- \preceq_{C^-FSD} C_G^-$). The final distribution obtained after all elementary shifts has the same marginals as G , therefore it can be transformed into a distribution G via association-increasing transformations according to the procedure described in Epstein and Tanny (1980) and Tchen (1980). \square

Theorem 2

Proof. It is clear from the form of association-increasing transformation that we can express them in the form of $n \times m$ matrices and we have $\Phi = \Phi_0 + \Phi_1 + \Phi_2 + \dots$. Let us focus on upward shifts' matrices.

In step t let i_t be the lowest number for which $F_{i_t}^1 < G_{i_t}^1$ and let $i_{\tilde{t}} > i_t$ be the lowest number such that $f_{i_{\tilde{t}}}^1 > 0$.

$$\gamma_{ij} := \begin{cases} \min \left(G_{i_t}^1 - F_{i_t}^1 - \sum_{k=1}^{j-1} \gamma_{i_tk}, f_{i_{\tilde{t}}j} \right) & i = i_t \\ -\gamma_{i_tj} & i = i_{\tilde{t}} \\ 0 & \text{elsewhere} \end{cases} \quad (16)$$

Similarly for F^2

$$\gamma^{ij} := \begin{cases} \min \left(G_{j_t}^2 - F_{j_t}^2 - \sum_{k=1}^{i-1} \gamma^{kj_t}, f_{ij_t} \right) & j = j_t \\ -\gamma^{ij_t} & j = j_{\tilde{t}} \\ 0 & \text{elsewhere} \end{cases} \quad (17)$$

A matrix of an upward shift along first marginal in step t is $\Gamma_t^1 = (\gamma_{ij})$ while a matrix of an upward shift along second diagonal is $\Gamma_t^2 = (\gamma^{ij})$.

Assuming that $f_{i_{\tilde{t}}}^1 \geq G_{i_{\tilde{t}}} - F_{i_{\tilde{t}}}$ we get

$$\begin{aligned} \tilde{F}_{i_{\tilde{t}}}^1 &= F_{i_{\tilde{t}}}^1 + \sum_{j=0}^n \gamma_{i_{\tilde{t}}j} = F_{i_{\tilde{t}}}^1 + \sum_{j=0}^n \min \left(G_{i_{\tilde{t}}}^1 - F_{i_{\tilde{t}}}^1 - \sum_{k=1}^{j-1} \gamma_{i_{\tilde{t}}k}, f_{i_{\tilde{t}}j} \right) = \\ &= F_{i_{\tilde{t}}}^1 + f_{i_{\tilde{t}}1} + f_{i_{\tilde{t}}2} + \dots + f_{i_{\tilde{t}}(r-1)} + G_{i_{\tilde{t}}}^1 - F_{i_{\tilde{t}}}^1 - \sum_{k=1}^{r-1} \gamma_{i_{\tilde{t}}k} + \\ &+ \sum_{j=r+1}^n \min \left(G_{i_{\tilde{t}}}^1 - F_{i_{\tilde{t}}}^1 - \sum_{k=1}^{j-1} \gamma_{i_{\tilde{t}}k}, f_{i_{\tilde{t}}j} \right) = \\ &= G_{i_{\tilde{t}}}^1 + \sum_{k=1}^{r-1} (f_{i_{\tilde{t}}k} - \gamma_{i_{\tilde{t}}k}) + \sum_{j=r+1}^n \min(0, f_{i_{\tilde{t}}j}) = G_{i_{\tilde{t}}}^1 \end{aligned}$$

We finish our proof by inductive argument. If $f_{i_t}^1 < G_{i_t} - F_{i_t}$ we need to transfer from higher category, so in at most $\frac{(n-1)(n-2)}{2}$ steps we have $\tilde{F}_{i_0+1}^1 \geq G_{i_0}^1$.

After a finite number of steps we obtain a matrix $\Gamma = \Gamma_0^1 + \Gamma_0^2 + \Gamma_1^1 + \Gamma_1^2 + \Gamma_2^1 + \Gamma_2^2 + \dots$. Adding $F + \Gamma$ we obtain a distribution which has the same marginal as G . We can obtain $\tilde{F}^1 = G^1$ in at most cubic time - this algorithm is by no means optimal but it shows the general idea of how to proceed from one distribution to the other. \square

Theorem 3

Proof. Let us rewrite $\mathbb{E}(w|f)$:

$$\begin{aligned} \mathbb{E}(w|f) &= \sum_{i,j} w(i,j) f_{ij} = \\ &= \sum_{i=1}^n \sum_{j=1}^m w(i,j) \left[F_{ij} - F_{(i-1)j} - F_{i(j-1)} + F_{(i-1)(j-1)} \right] + \\ &+ \sum_{i=1}^n w(i,j) \left[F_{i0} - F_{(i-1)0} \right] + \sum_{j=1}^m w(i,j) \left[F_{0j} - F_{0(j-1)} \right] + w(0,0) F_{00} = \\ &= \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} [w(i,j) - w(i,j+1) - w(i+1,j) + w(i+1,j+1)] F_{ij} + \\ &+ \sum_{j=0}^{m-1} [w(n,j) - w(n,j+1)] F_j^2 + \sum_{i=0}^{n-1} [w(i,m) - w(i+1,m)] F_i^1 + \\ &+ w(n,m) F_{nm} \end{aligned}$$

It is straightforward to see that if $F \preceq_{CFSD} G$ then we get:

$$\begin{aligned} \mathbb{E}(w|f) - \mathbb{E}(w|g) &= \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} [w(i,j) - w(i,j+1) - w(i+1,j) + \\ &+ w(i+1,j+1)] (F_{ij} - G_{ij}) + \sum_{j=0}^{m-1} [w(n,j) - w(n,j+1)] (F_j^2 - G_j^2) + \\ &+ \sum_{i=0}^{n-1} [w(i,m) - w(i+1,m)] (F_i^1 - G_i^1) \geq 0 \end{aligned}$$

for all $w \in \mathcal{W}_1 \cap \mathcal{W}_2$. Now we only need to show that $F \preceq_{CFSD} G$ is necessary condition. Let us assume that $F_{kl} \geq G_{kl}$ for some $k < n$, $l < m$. We prove by contradiction. Let $w(i,j) = 0$ for $i \leq k \wedge j \leq l$ and $w(i,j) = 1$ elsewhere. Then $w(i,j) - w(i,j+1) - w(i+1,j) + w(i+1,j+1) = 0$ for all $(i,j) \neq$

(k, l) , $w(k, l) - w(k, l + 1) - w(k + 1, l) + w(k + 1, l + 1) = -1$ and we have also $w(i, m) - w(i + 1, m) = w(n, j) - w(n, j + 1) = 0$ for all (i, j) (in the case where $k = n$ or $l = m$ one of $w(i, m) - w(i + 1, m)$ or $w(n, j) - w(n, j + 1)$ would be equal -1 and the rest of the terms would be equal to 0). Now it is easy to see that

$$\mathbb{E}(w|f) - \mathbb{E}(w|g) = -(F_{kl} - G_{kl}) \leq 0$$

The case of survival concordance and supermodular functions is similar. □

Theorem 4

Proof. Let us first observe that if P is of the form (13) axioms are obviously satisfied. We assume now that the axioms are satisfied. Let

$$D = \{(C_F^+, F^1, F^2) | F \in \Lambda\}.$$

Further, we define $u : D \mapsto \Lambda$ as $u(d) := C_F^+(F^1, F^2)$ where $D \ni d = (C_F^+, F^1, F^2)$. Sklar's theorem ensures that u is an injection. Less formally, u produces a distribution F such that C_F^+ is its maximal extension copula and F^1, F^2 are its marginal distributions. Thus $h = \bar{h} \circ u$ gives us the required representation. We will now show that h increases with respect to its coordinates. Let F, G be such that $C_F^+ \ll C_G^+$ and $F^1 = G^1$ and $F^2 = G^2$. Then from (12) it follows that $P(F) \leq P(G)$ and $h(C_F^+, F^1, F^2) \leq h(C_G^+, G^1, G^2)$, that is, h increases with respect to the ordering on copulas. Similarly one can show that h increases with respect to FSD relation along second and third coordinate. □