The Supermodular Stochastic Ordering

Margaret Meyer Bruno Strulovici*

First draft: February 8, 2010

July 8, 2010

Abstract

This paper uses the stochastic dominance approach to study orderings of interdependence for n-dimensional random vectors. Supermodularity (Topkis, 1968) of an objective function is a natural property to capture a preference for greater interdependence. We characterize the partial ordering on n-dimensional distributions which is equivalent to one distribution's yielding a higher expectation than another for all supermodular objective functions. Though the "supermodular stochastic ordering" has previously been characterized for the special case of bivariate distributions, our results apply to random vectors with an arbitrary number of dimensions, and exploit duality in polyhedral description of the ordering. In particular, supermodular dominance is equivalent to one distribution being derivable from another by a sequence of nonnegative "elementary transformations." We develop several methods for determining whether such a sequence exists. For the special case of random vectors with conditionally i.i.d. components ("mixture distributions"), we provide sufficient conditions for supermodular dominance; these conditions have a natural interpretation as a non-parametric ordering of the relative size of aggregate vs. idiosyncratic shocks. We also characterize the symmetric supermodular ordering and provide a set of sufficient conditions for symmetric supermodular dominance. Finally, we describe applications of our approach and results to a range of questions in welfare economics, matching markets, social learning, insurance, and finance.

*Email addresses: margaret.meyer@economics.ox.ac.uk b-strulovici@northwestern.edu

1 Introduction

In many economic contexts, it is of interest to know whether one set of random variables displays a greater degree of interdependence than another. In this paper, we use the stochastic dominance approach to study a range of notions of greater interdependence, focusing particularly on the supermodular stochastic ordering.

The stochastic dominance approach to assessing interdependence relates orderings of interdependence expressed directly in terms of joint probability distributions to orderings expressed indirectly through properties of objective functions whose expectations are used to evaluate distributions. Since the expected values of additively separable objective functions depend only on marginal distributions, attitudes towards interdependence must be represented through non-separability properties. We argue that the property of supermodularity (Topkis, 1968) of an objective function is a natural property with which to capture a preference for greater interdependence. Supermodularity of a function captures the idea that its arguments are complements, not substitutes: When an increasing function of two or more variables is supermodular and the values of any two variables are increased together, the resulting increase in the function is larger than the sum of the increases that would result from increasing one or the other of the values separately. Our main objective in this paper is to characterize the partial ordering on distributions of $n-$dimensional random vectors which is equivalent to one distribution's yielding a higher expectation than another for all supermodular objective functions. Following the statistics literature, we refer to this partial ordering as the "supermodular stochastic ordering" (Shaked and Shanthikumar, 1997).

There are many branches of economics where the supermodular stochastic ordering is a valuable tool for comparing distributions with respect to their degree of interdependence. Section 2 describes applications of our methods and results to the assessment of i) ex post inequality under uncertainty; ii) multidimensional inequality; iii) the efficiency of matching in the presence of informational or search frictions; iv) the effect of network structure on conformity of behavior or beliefs in social learning situations; v) the dependence among claims in a portfolio of insurance policies or among assets in a financial institution's portfolio.

For the special case of two-dimensional random vectors, the economics and statistics literatures have provided a complete characterization of the supermodular ordering. Specifically, Epstein and Tanny (1980) and Tchen (1980), among others, have shown that one bivariate distribution dominates another according to the supermodular ordering if and only if the first distribution dominates the second in the sense of both upper-orthant and lower-orthant dominance. Hu, Xie, and Ruan (2005) have shown that this equivalence continues to hold in three dimensions in the special case of Bernoulli random vectors, but the equivalence breaks down for more than three dimensions (Joe, 1990) and even in three dimensions for larger supports (Müller and Scarsini, 2000). In general, the supermodular ordering is strictly stronger than the combination of upper-orthant and lower-orthant dominance.

Focusing on the case where the random vectors to be compared have discrete supports on a lattice, we characterize the supermodular ordering for more than two dimensions. In Section 4, we use duality to prove (Theorem 1) that one distribution is preferred to the other by every supermodular objective function if and only if the first distribution can be derived from the other by a sequence of nonnegative "elementary transformations". Intuitively, our elementary transformations play a similar role to the mean-preserving spreads defined by Rothschild and Stiglitz (1970) for univariate distributions to capture the notion of greater riskiness.

In the current context, where our concern is with interdependence between dimensions rather than with riskiness in a single dimension, our elementary transformations leave all marginal distributions unaffected. Holding fixed the realizations of $n-2$ of the random variables comprising the random vector, our elementary transformations increase the probability that the remaining two variables will take on (relatively) high values together or (relatively) low values together and reduce the probability that one will be high and the other low. For multivariate distributions, our elementary transformations provide a local characterization of the notion of "greater interdependence". They are a natural generalization to multivariate distributions of the bivariate "correlation-increasing transformations" defined by Epstein and Tanny (1980). In another sense, though, our definition of elementary transformations is more restrictive than Epstein and Tanny's, in that our transformations affect only adjacent points in the support; because of this restriction, as we prove (Theorem 3), our transformations are all extreme, in the sense that none can be expressed as a positive linear combination of the others.

Section 5 shows how our restrictive definition of elementary transformations allows a simple constructive proof of the known characterization of the supermodular ordering for bivariate distributions. For any pair of bivariate distributions with identical marginals, if we allow elementary transformations to be given weights that are either positive or negative, then there is a unique weighted sequence of elementary transformations of our form that converts one distribution into the other. Therefore, two bivariate distributions can be ranked according to the supermodular ordering if and only if the weights in the unique sequence are all non-negative.

For pairs of distributions f, g in three or more dimensions, even with our restrictive definition of elementary transformations (and even confining attention to distributions with identical marginals), there are many weighted sequences of elementary transformations that convert one distribution into the other. How, then, can we determine whether g dominates f according to the supermodular ordering? In Section 6, we develop three different methods for assessing whether in fact g can be derived from f by a sequence of elementary transformations with nonnegative weights. The first approach is constructive and builds on the result that none of our elementary transformations is redundant. This constructive approach allows us, for distributions on supports with small numbers of nodes, to directly derive inequalities which are necessary and sufficient for supermodular dominance of g over f to hold.

A second approach is to formulate a linear program, based on the set of elementary transformations on the discrete support, such that the optimum value of the program is zero if and only if there exist non-negative weights on elementary transformations that will convert f to g. This method, like the first approach, has the advantage of constructing an explicit sequence of elementary transformations. However, it also has the drawback that one has to solve a different linear program for each pair of distributions to be compared.

Our third method is based on Minkowski's and Weyl's representation theorems for polyhedral cones, and it allows us to compute once and for all, for any given support, a minimal set of inequalities that characterize the stochastic supermodular ordering. This method can be used for optimization problems such as mechanism design or analysis of optimal policy, where each mechanism or policy generates a multivariate distribution, and the set of mechanisms or policies to be compared is large. Specifically, we develop an algorithm, based on the "double description method" conceptualized by Motzkin et al. (1953) and developed by Avis and Fukuda (1992) to generate, for any given multidimensional support, the set of extreme rays of the cone of supermodular functions. Each extreme ray corresponds to one of the minimal set of inequalities defining the supermodular ordering.

In some applications, it is natural to focus on objective functions that are symmetric.

Section 7 studies the ordering on distributions that corresponds to one distribution's generating a higher expected value than another for all symmetric supermodular objective functions. We term this ordering the symmetric supermodular ordering and show in Theorem 5 that two distributions can be ranked according to the symmetric supermodular ordering if and only if the "symmetrized" versions of the distributions satisfy supermodular dominance. We then use this result to characterize (in Proposition 4) the symmetric supermodular ordering for any number of dimensions and l points in the support of each dimension in terms of a closely related ordering on an $l-1$-dimensional support. For n-dimensional random vectors representing n independent lotteries, we identify in Theorem 6 sufficient conditions for symmetric supermodular dominance and show that these conditions have a natural interpretation in terms of lower dispersion among one set of lotteries than another.

Section 8 studies the special case of multivariate distributions generated as follows: first, a univariate probability distribution is drawn randomly, according to some distribution. Then, all random variables are drawn independently from that common distribution. The resulting multivariate distribution is a *mixture* of conditionally i.i.d. random variables. Since the common distribution is ex ante uncertain, this creates some positive dependence between the random variables. We compare the interdependence of two random vectors each of which is a mixture of conditionally i.i.d. random variables. Specifically, we provide in Theorem 7 sufficient conditions for two (symmetric) mixture distributions to be ranked according to the supermodular ordering. The sufficient conditions we identify have a natural interpretation as a non-parametric ordering of the relative size of aggregate vs. idiosyncratic shocks. At a formal level, moreover, they are very closely related to the sufficient conditions for symmetric supermodular dominance identified in Theorem 6.

Section 9 extends our approach of using duality results for polyhedral cones to characterize a range of other stochastic orders. We identify the set of elementary transformations that correspond to dominance with respect to all objective functions satisfying both supermodularity and componentwise convexity, or supermodularity and full convexity. Convexity on lattices is a nontrivial concept, and our characterization of it in terms of elementary transformations is an interesting result in itself.

Section 10 provides a brief discussion of how our approach to assessing interdependence relates to analyses of copulas, and Section 11 contains concluding remarks.

2 Applications

Our methods and results are applicable to a wide range of questions in economics and related fields. Consider first some applications in welfare economics. In many group settings where individual outcomes (e.g. rewards) are uncertain, members of the group may be concerned, ex ante, about how unequal their ex post rewards will be (Meyer and Mookherjee, 1987; Ben-Porath et al, 1997; Gajdos and Maurin, 2004; Kroll and Davidovitz, 2003; Adler and Sanchirico, 2006; Chew and Sagi, 2010). (This concern is distinct from concerns about the mean level of rewards and about their riskiness.) As argued by Meyer and Mookherjee (1987), an aversion to ex post inequality can be formalized by adopting an ex post welfare function that is supermodular in the realized utilities of the different individuals. We then want to know: Given two mechanisms for allocating rewards (formally, two joint distributions of random utilities), when can we be sure that one mechanism generates higher expected welfare than the other, for all supermodular ex post welfare functions? Our characterization results for the supermodular ordering allow us to answer this question.

Consider a specific illustration. Intuitively, when groups dislike ex post inequality, tournament reward schemes, which distribute a fixed set of rewards among individuals, one to each person, should be particularly unappealing, since they generate a form of negative correlation among rewards: if one person receives a higher reward, this must be accompanied by another person's receiving a lower reward. This intuitive reasoning suggests the conjecture that tournaments should be dominated, in the sense of the supermodular ordering, by reward schemes that provide each individual with the same marginal distribution over rewards but determine rewards independently. Meyer and Mookherjee (1987) proved this conjecture for an arbitrary number of individuals (dimensions), but only for the special case of a symmetric tournament (one in which each individual has an equal chance of winning each of the rewards), and their method of proof was laborious. Here, we allow tournaments to be arbitrarily asymmetric across individuals, and we compare expected ex post welfare under a tournament with that under the reward scheme which for each individual yields the same marginal distribution of rewards as he faced under the tournament but which allocates rewards independently. We show that for all symmetric supermodular ex post welfare functions, expected welfare is lower under the tournament.

A second application in welfare economics concerns comparisons of inequality or poverty when separate data are available on different dimensions of economic status, for example, income, health, and education (Atkinson and Bourguignon, 1982, Bourguignon and Chakravarty, 2002, and Decancq, 2009). Depending on whether the different attributes are regarded as complements or substitutes at the individual level, the function aggregating the attributes into an individual welfare measure will be supermodular or submodular. Our characterization results for the supermodular ordering provide the conditions under which one multidimensional distribution can be ranked above another for all welfare measures in the given class. Furthermore, we develop constructive methods for checking supermodular dominance that can be easily applied to the comparison of empirical distributions.

Another set of microeconomic applications concerns comparisons of the efficiency of two-sided or many-sided matching mechanisms when the outcomes of the matching process are subject to frictions. Consider, for example, settings where different categories of workers (e.g. newly-qualified and experienced, or technical and managerial) are matched with firms. Suppose that workers within each category, as well as firms, are heterogeneous and that the production function giving the output of a matched set of workers at a given firm, as a function of the workers' types and the firm's type, is supermodular. In the absence of any frictions, the efficient matching would be perfectly assortative, matching the highest-quality worker in each category with the highest-quality firm, the next-highest-quality workers with the next-highest-quality firm, etc. Such a matching would correspond to a "perfectly correlated" joint distribution of the random variables representing quality in each category (dimension). When, however, matches are formed based only on noisy or coarse information (McAfee, 2002), or when search is costly (Shimer and Smith, 2000), or when signaling is constrained by market imperfections such as borrowing constraints (Fernandez and Gali, 1999), perfectly assortative matching will generally not arise. In these settings, our characterization of the supermodular ordering can be used to assess when one matching mechanism will generate higher expected output than another, for all supermodular production functions. Fernandez and Gali (1999) and Meyer and Rothschild (2004) apply existing two-dimensional results to compare matching institutions, but multi-dimensional applications remain largely unexplored. One exception is Prat (2002), but he compares only a perfectly correlated joint distribution with an independent one, and Lorentz (1953) has shown that the former is preferred to the latter for all supermodular objective functions.

The stochastic supermodular ordering could also prove a valuable tool for studying how "social structure" influences the degree of interdependence ("conformity") of individual beliefs or choices in social learning situations. Recent studies of communication and learning in social networks (e.g. Golub and Jackson, 2009, Acemoglu et al, 2008, and Acemoglu et al, 2009) examine settings in which individuals learn by communicating with and/or observing the behavior of others, and the social structure that influences communication and/or observation is described by a network. These studies examine the limiting beliefs/choices of the community as the number of individuals interacting and/or the number of periods of interaction grows large, focusing on whether or not the limiting beliefs/choices match the truth. Particular interest is attached to how the structure of the network in which individuals are embedded affects the results. The focus on the limiting cases of infinitely large communities or infinitely repeated interaction is, at least in part, for tractability. The stochastic supermodular ordering could be used to study theoretically the degree of interdependence in behavior in finite communities interacting over a finite number of periods, examining questions such as how changes in the network structure or in the nature of communication opportunities affect the degree of conformity of individual beliefs and choices. The supermodular ordering could also prove a useful tool for analyzing experimental data on interdependence of behavior in social networks (see, for example, Choi, Gale, and Kariv, 2005 and 2009).

Macroeconomists need to be able to gauge and compare levels of "systematic risk". At the level of a single country, this involves assessing the degree of covariation among levels of output in different sectors, while at the level of the world economy, it involves assessing the degree of interdependence among output levels in different countries. In both of these cases, the assessments are naturally multidimensional rather than simply two-dimensional. Hennessy and Lapan (2003) have proposed using the supermodular stochastic ordering to make such comparisons.

In the actuarial literature, the supermodular ordering has recently received considerable attention as a means of comparing the degrees of dependence among claims in a portfolio of insurance policies (see Müller and Stoyan, 2002, and Denuit, Dhaene, Goovaerts, and Kaas, 2005). In finance, the supermodular ordering has been proposed as a method for assessing the dependence among asset returns in a portfolio (Epstein and Tanny, 1980) and as a method for assessing the interdependence between a single institution's portfolio and the market as a whole (Patton, 2009). Moreover, financial economists have recently shown increased interest in developing measures of interdependence for the components of the financial system as a whole and not just for individual assets. Brunnermeier and Adrian (2009), for example, study interdependence among financial institutions, with the objective of developing measures of "systemic risk" that capture the degree of comovement among individual institutions' entry into states of financial distress.

3 General Setting

This section introduces the general setting analyzed in the paper.

Distribution Support We consider multivariate distributions with the same number, n, of variables and identical, finite support. Formally, let L_i denote the finite, totally ordered set of values taken by the i^{th} random variable, and let L denote the cartesian product of L_i's. For all applications, and in what follows, L_i is a finite subset of $\mathbb{R}$ and L is a finite lattice of $\mathbb{R}^n$ with the following partial order: $x \leq y$ if and only if $x_i \leq y_i$ for all $i \in N = \{1, \ldots, n\}$. If l_i denotes the cardinality of L_i, then L has $d = \prod_{i=1}^n l_i$ elements.

As a specific example, let $L_{l_1,\ldots,l_n}$ denote the lattice of $\mathbb{R}^n$ with $L_i = \{0, \ldots, l_i - 1\}$. Thus, for example, $L_{2,2}$ consists of the vertices of the unit square in $\mathbb{R}^2$ based at the origin: $L_{2,2} = \{0,1\}^2$. Similarly, $L_{2,2,2}$ consists of the vertices of the unit cube of $\mathbb{R}^3$ based at the origin: $L_{2,2,2} = \{0,1\}^3$.

For any $x \in L$, let $x + e_i$ denote the element y of L, whenever it exists, such that $y_j = x_j$ for all $j \in N \setminus \{i\}$ and y_i is the smallest element of L_i greater than but not equal to x_i. For example, in $L_{2,2}$, $(0,0) + e_1 = (1,0)$ and $(1,0) + e_2 = (0,0) + e_1 + e_2 = (1,1)$.

Lattice vs. Vector Structures. The lattice structure of the support L and its corresponding order is used to compare distributions. In particular, supermodularity of objective functions is defined with respect to that partial order. One may label the d elements (or "nodes") of L and view real functions on L as vectors of $\mathbb{R}^d$, where each coordinate of the vector corresponds to the value of the function at a specific node of L. This representation will prove particularly important for dual characterizations of interdependence relations. A multivariate distribution whose support is L (or a subset of L) can be represented as an element of the unit simplex Δ_d of $\mathbb{R}^d$.

Orderings of Multivariate Distributions. For any function $w : L \to \mathbb{R}$ and distribution $f \in \Delta_d$, the expected value of w given f is the scalar product of w with f, seen as vectors of $\mathbb{R}^d$:

$$E[w|f] = \sum_{x \in L} w(x) f(x) = w \cdot f,$$

where $\cdot$ denotes the scalar product of w and f in $\mathbb{R}^d$. To any class $\mathcal{W}$ of functions on L corresponds an ordering of multivariate distributions:

$$f \prec_{\mathcal{W}} g \quad \Leftrightarrow \quad \forall w \in \mathcal{W}, \quad E[w|f] \leq E[w|g] \tag{1}$$

The main purpose of this paper is to better understand the orders defined according to such classes of functions, starting with the stochastic supermodular ordering, which is based on supermodular functions.

4 The Stochastic Supermodular Ordering

Supermodular Functions and Elementary Transformations For any $x, y \in L$, denote by $x \wedge y$ the component-wise minimum (or "meet") of x and y, i.e., the element of L such that $(x \wedge y)_i = \min\{x_i, y_i\} \in L_i$ for all $i \in N$. Let $x \vee y$ similarly denote the component-wise maximum (or "join") of x, y. A function w is said to be *supermodular* (on L) if $w(x \wedge y) + w(x \vee y) \geq w(x) + w(y)$ for all $x, y \in L$. Supermodular functions are characterized by the following property (see Topkis, 1968):

$$w \in \mathcal{S} \quad \Leftrightarrow \quad w(x + e_i + e_j) + w(x) \geq w(x + e_i) + w(x + e_j) \tag{2}$$

for all $i \neq j$ and x such that $x + e_i + e_j$ is well-defined (i.e., such that x_i is not the upper bound of L_i and x_j is not the upper bound of L_j). For any $x \in L$ such that $x + e_i + e_j$ is well-defined, let $t^x_{i,j}$ denote the function on L such that

$$t^x_{i,j}(x) = t^x_{i,j}(x + e_i + e_j) = -t^x_{i,j}(x + e_i) = -t^x_{i,j}(x + e_j) = 1 \tag{3}$$

and $t^x_{i,j}(y) = 0$ for all other nodes y of L. We call these functions the *elementary transformations* on L. Let $\mathcal{T}$ denote the class of all elementary transformations.

For example, for $L_{2,2}$, there is a single elementary transformation, which is defined by $t(1,1) = t(0,0) = 1$ and $t(1,0) = t(0,1) = -1$. For $L_{2,2,2}$, there are six elementary transformations, one corresponding to each face of the unit cube. For $L_{3,3}$, there are four elementary transformations, corresponding to the four values of x , namely $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$, such that $x + e_i + e_j$ is well defined. Observe that our definition of elementary transformations confines attention to transformations that i) affect only *two* of the n dimensions (as illustrated by the example of $L_{2,2,2}$) and ii) affect values only at four *adjacent* points in the lattice, x, $x+e_i$, $x+e_j$, and $x+e_i+e_j$ (as illustrated by the example of $L_{3,3}$).

With this notation, (2) can be re-expressed as

$$w \in \mathcal{S} \Leftrightarrow w \cdot t \geq 0 \quad \forall t \in \mathcal{T}. \tag{4}$$

Now that we have a formal characterization of the class of supermodular functions, we can formally define the (stochastic) supermodular ordering:

$$f \prec_{\mathcal{S}PM} g \Leftrightarrow \forall w \in \mathcal{S}, \quad E[w|f] \leq E[w|g] \tag{5}$$

If $f \prec_{\mathcal{S}PM} g$, we will say that distribution g is *more interdependent* than distribution f.

Dual Characterization When does a random vector Y, distributed according to g, exhibit more interdependence among its components than another random vector X, distributed according to f? What modifications to the distribution of a random vector increase interdependence among the random variables composing it? The answer is given in the following theorem.

THEOREM 1 (SUPERMODULAR ORDERING) *$f \prec_{\mathcal{S}PM} g$ if and only if there exist* nonnegative *coefficients $\{\alpha_t\}_{t\in\mathcal{T}}$ such that, with f, g, and t seen as vectors of* $\mathbb{R}^d$,

$$g = f + \sum_{t\in\mathcal{T}} \alpha_t t. \tag{6}$$

Proof. Equation (6) holds if and only if $g - f$ belongs to the convex cone $\mathcal{T}^C$ generated by $\mathcal{T}$, i.e., defined by $\mathcal{T}^C = \{\sum_{t\in\mathcal{T}} \alpha_t t : \alpha_t \geq 0 \quad \forall t \in \mathcal{T}\}$. From (4), $\mathcal{S}$ is the dual cone of $\mathcal{T}^C$. Since $\mathcal{T}^C$ is closed and convex, this implies (see Luenberger, 1969, p. 215) that $\mathcal{T}^C$ is the dual cone of $\mathcal{S}$. That is,

$$\delta \in \mathcal{T}^C \Leftrightarrow w \cdot \delta \geq 0 \quad \forall w \in \mathcal{S}.$$

By definition of the stochastic supermodular ordering (see (5)), the above equation exactly means that $f \prec_{\mathcal{S}PM} g$ if and only if $g - f \in \mathcal{T}^C$, which shows the result. ■

Coarsening For many applications, the choice of a particular support seems somewhat arbitrary. For example, when comparing several empirical distributions of inequality across various components (such as income, health, and education), the distribution depends on the way data has been aggregated into discrete categories. It is natural, then, to ask whether our notion of greater interdependence is robust with respect to further aggregation. Theorem 1 provides a way to answer this question.

Define a *coarsening* M of some support L by a partitioning of each L_i into M_i, consisting of $m_i \leq l_i$ components of consecutive elements of L_i. For example, if $L = \{0,1,2,3\} \times \{0,1,2\}$, one possible coarsening of L is $M = \{\{0,1\},\{2,3\}\} \times \{\{0\},\{1,2\}\}$. To any coarsening M of L corresponds a surjective map $\phi : L \to M$ such that $\phi(x) = \phi(x')$ if and only if x_i and x'_i belong to the same element y_i of M_i for all i. Each element of M represents a hyper-rectangle resulting from slicing L along (possibly) each dimension. For any distribution f on L and any coarsening M of L, let f^M denote the "coarsened version" of f, which is defined by

$$f^M(y) = \sum_{x \in L : \phi(x) = y} f(x).$$

To indicate dependence with respect to the chosen support, let $\mathcal{S}(L)$ denote the set of all supermodular functions with domain L.

THEOREM 2 (COARSENING INVARIANCE) *If $f \prec_{\mathcal{S}(L)} g$, then for any coarsening M of L, $f^M \prec_{\mathcal{S}(M)} g^M$.*

Proof. Suppose that $f \prec_{\mathcal{S}(L)} g$. By Theorem 1, this implies the existence of nonnegative coefficients α_t such that

$$g = f + \sum_{t \in \mathcal{T}(L)} \alpha_t t, \tag{7}$$

where $\mathcal{T}(L)$ is the set of elementary transformations on L. Let Φ denote the operator which to any function w on L associates the function on M defined by $\Phi(w)(y) = \sum_{x \in L : \phi(x) = y} w(x)$. Φ is a linear operator, and by construction, $f^M = \Phi(f)$. Applying Φ to (7) yields

$$g^M = f^M + \sum_{t \in \mathcal{T}(L)} \alpha_t \Phi(t).$$

Now observe that for $t = t^x_{i,j} \in \mathcal{T}(L)$, $\Phi(t)$ belongs to $\mathcal{T}(M)$ if $\phi(x)$, $\phi(x+e_i)$, $\phi(x+e_j)$, and $\phi(x+e_i+e_j)$ are all distinct, and $\Phi(t)(y) = 0$ for all $y \in M$ otherwise. Therefore,

$$g^M = f^M + \sum_{t \in \mathcal{T}(M)} \alpha_t t,$$

for some nonnegative coefficients α'. Another application of Theorem 1 then implies that $f^M \prec_{\mathcal{S}(M)} g^M$, which concludes the proof. ■

Thus, if distribution g is more interdependent than distribution f on a given support L, then on any coarsening M of L, the coarsened version of g, g^M, is more interdependent than the coarsened version of f, f^M.

In the next several sections, we develop a range of methods for determining, given a pair of distributions f and g, whether g is more interdependent than f. These methods apply the characterization result of Theorem 1 and are greatly facilitated by two aspects of our approach. The first is our restriction to a *finite* support L. The second is the manner in which we have defined the elementary transformations on L, requiring that the transformations affect only two of the n dimensions and affect values at only adjacent points in the lattice. These two features of our approach imply that it is very straightforward, either manually or algorithmically, to list the entire set $\mathcal{T}$ of elementary transformations on any given L. Furthermore, given a pair of distributions f, g, when we search for a representation of $g - f$ as a nonnegative weighted sum $\sum_{t\in\mathcal{T}} \alpha_t t$, we can be certain that none of the elementary transformations in $\mathcal{T}$ is redundant, as demonstrated by the following:

THEOREM 3 *All elements of* $\mathcal{T}$ *are extreme rays of* $\mathcal{T}^C$*, the convex cone generated by* $\mathcal{T}$*.*

Proof. Without loss of generality, we prove the claim for $L = L_{l_1,\ldots,l_n}$ (other cases are treated with an obvious modification of the function w below). Consider a point $x \in L$ and a pair of dimensions i, j such that the elementary transformation $t^* \equiv t_{i,j}^{x-e_i-e_j}$ is well-defined. Suppose that, contrary to the claim, there exist nonnegative coefficients α_s such that

$$t^* = \sum_{s\in\mathcal{T}\setminus\{t^*\}} \alpha_s s. \tag{8}$$

Let us define the function w on L by $w(x) = (\frac{3}{4})2^{\sum_k x_k}$ and, for $y \neq x$, $w(y) = 2^{\sum_k y_k}$. It is easy to check that w is supermodular. Moreover, w makes a nonnegative scalar product with all elementary transformations and a positive scalar product with all elementary transformations except for those whose highest corner is x. Since t^* is one of the elementary transformations whose highest corner is x, taking the scalar product of w with both sides of (8) implies that

$$0 = \sum_{s\in\mathcal{T}\setminus\{t\}} \alpha_s (w \cdot s).$$

This equation in turn implies that $\alpha_s = 0$ for all transformations s except possibly those whose highest corner is x. However, t^* cannot be a positive linear combination of only elementary transformations whose highest corner is x. To see this, observe that any elementary transformation s (other than t^*) whose highest corner is x must take value 0 at $x - e_i - e_j$, whereas t^* evaluated at $x - e_i - e_j$ equals 1. ■

For the special case of two dimensions, a stronger result is easily shown: It is impossible to write any elementary transformation $t \in \mathcal{T}$ as a sum, with weights of *arbitrary* sign, of other elementary transformations in $\mathcal{T}$. However, for three or more dimensions, this stronger condition does not hold, as the following example demonstrates: For $L = \{0,1\}^3$, $t_{13}^{(0,0,0)} = t_{13}^{(0,1,0)} - t_{23}^{(1,0,0)} + t_{23}^{(0,0,0)}$.

The constructive methods we develop for determining whether a distribution g is more interdependent than a distribution f also exploit an important implication of the relation $f \prec_{\mathcal{SPM}} g$, namely that f and g have identical univariate marginal distributions. To see why this holds, note that for any dimension $i \in \{1, \ldots, n\}$ and any $k \in L_i$, the functions $\overline{w}(x) = I_{\{x_i \geq k\}}$ and $\underline{w}(x) = I_{\{x_i < k\}}$ are both supermodular. Therefore $f \prec_{\mathcal{SPM}} g$ implies that, for all $i \in \{1, \ldots, n\}$ and any $k \in L_i$,

$$\begin{aligned} 0 \leq E[\overline{w}|g] - E[\overline{w}|f] &= \sum_{x:\, x_i \geq k} g(x) - \sum_{x:\, x_i \geq k} f(x) \\ \text{and} \quad 0 \leq E[\underline{w}|g] - E[\underline{w}|f] &= \sum_{x:\, x_i < k} g(x) - \sum_{x:\, x_i < k} f(x), \end{aligned} \tag{9}$$

and these inequalities together imply that f and g have identical univariate marginal distributions. This conclusion also follows from the characterization of Theorem 1, given that for any elementary transformation $t \in \mathcal{T}$ and for any α, $f + \alpha t$ and f have the same marginal distributions.

5 Two Dimensions

Theorem 1 tells us that, given two distributions f, g, determining whether $f \prec_{\mathcal{SPM}} g$ is equivalent to determining whether the difference vector $\delta = g - f$ can be decomposed into a nonnegative weighted sum of elementary transformations. For the special case of bivariate distributions ($n = 2$), we now show that, given how we have defined elementary transformations, this determiniation is extremely simple. Given f, g with identical marginal distributions and defined on $L = L_{l_1,l_2} \equiv \{0, \ldots, l_1 - 1\} \times \{0, \ldots, l_2 - 1\}$, the difference vector δ is fully described by its values at $(l_1 - 1) \times (l_2 - 1)$ points (the remaining values being pinned down by the condition of identical marginals), and there are exactly $(l_1 - 1) \times (l_2 - 1)$ (linearly independent) elementary transformations defined as in (3).

Therefore, there is a *unique* decomposition of δ into a weighted sum of elementary transformations $t \in \mathcal{T}$, where the weights α_t can have *arbitrary* signs. Since the decomposition is unique, $f \prec_{SPM} g$ if and only if the weight on every elementary transformation in the decomposition is nonnegative.

It is also straightforward to identify the weight on each elementary transformation in the unique decomposition, as a function of the difference vector δ. To simplify notation, note that with only two dimensions, given an arbitrary $z \in L$, we can write t^z instead of $t^z_{i,j}$ for the elementary transformation defined in (3). Also, let $\alpha(z)$ denote α_{t^z}. The elementary transformation t^z is well-defined for $z \in \{0, \ldots, l_1 - 2\} \times \{0, \ldots, l_2 - 2\} \equiv L_{(l_1-1),(l_2-1)}$. With only two dimensions, for any given $z \in L_{(l_1-1),(l_2-1)}$, there are at most four elementary transformations $t \in \mathcal{T}$ that take on non-zero values at z: t^z, $t^{(z-e_1)}$, $t^{(z-e_2)}$, and $t^{(z-e_1-e_2)}$. If $z = (z_1, 0)$, then $z - e_2$ is not well-defined; it is convenient in this case to say that $t^{(z-e_2)}$ is identically 0. Similarly, if $z = (0, z_2)$, then $z - e_1$ is not well-defined, and in this case we say that $t^{(z-e_1)}$ is identically 0. With these conventions, it follows that for any $z \in L_{(l_1-1),(l_2-1)}$,

$$\begin{aligned}\delta(z) &= \alpha(z)t^z(z) + \alpha(z-e_1)t^{(z-e_1)}(z) + \alpha(z-e_2)t^{(z-e_2)}(z) + \alpha(z-e_1-e_2)t^{(z-e_1-e_2)}(z) \\ &= \alpha(z) - \alpha(z-e_1) - \alpha(z-e_2) + \alpha(z-e_1-e_2), \qquad (10)\end{aligned}$$

where the second line follows from the definition of elementary transformations in (3).

A simple inductive process allows us to solve the equations (10) for the weights $\alpha(z)$. Start with $z = (0, 0)$. Since the only elementary transformation that takes on a non-zero value on $(0, 0)$ is $t^{(0,0)}$, (10) reduces to $\delta(0, 0) = \alpha(0, 0)$. Thus the weight $\alpha(0, 0)$ on $t^{(0,0)}$ in the unique decomposition of δ is $\delta(0, 0)$. Proceed now to $z = (1, 0)$. Since the only two elementary transformations that take on non-zero values on $(1, 0)$ are $t^{(1,0)}$ and $t^{(0,0)}$, (10) reduces to $\delta(1, 0) = \alpha(1, 0) - \alpha(0, 0)$, and hence $\alpha(1, 0) = \delta(0, 0) + \delta(1, 0)$. Straightforward induction arguments then show that for $z = (z_1, 0)$, $\alpha(z_1, 0) = \sum_{i=0}^{z_1} \delta(i, 0)$; for $z = (0, z_2)$, $\alpha(0, z_2) = \sum_{j=0}^{z_2} \delta(0, j)$; and finally for $z = (z_1, z_2)$, $\alpha(z_1, z_2) = \sum_{i=0}^{z_1} \sum_{j=0}^{z_2} \delta(i, j)$. If we define G and F as the cumulative distribution functions corresponding to g and f, respectively, then we have $G(z_1, z_2) - F(z_1, z_2) = \sum_{i=0}^{z_1} \sum_{j=0}^{z_2} \delta(i, j)$. Thus, in the unique decomposition of $\delta = g - f$ into a weighted sequence of elementary transformations, the weight $\alpha(z)$ on the transformation t^z is the difference $G(z) - F(z)$. Since $f \prec_{SPM} g$ if and only if every elementary transformation has a nonnegative weight in the decomposition, it follows that for two dimensions,

$$f \prec_{SPM} g \quad \Leftrightarrow \quad G(z) - F(z) \geq 0 \quad \forall z \in L. \qquad (11)$$

Note that (11) is written for all $z \in L$ and not just for all $z \in L_{(l_1-1),(l_2-1)}$, because identical marginals is a necessary condition for $f \prec_{\mathcal{SPM}} g$ and ensures that for $z = (l_1 - 1, 0)$ or $z = (0, l_2 - 1)$, $G(z) - F(z) = 0$.

For random variables $(Y_1, \ldots, Y_n)$ and $(X_1, \ldots, X_n)$ with distribution g and f, respectively, define the survival functions $\overline{G}$ and $\overline{F}$ by $\overline{G}(z) = P(Y \geq z)$ and $\overline{F}(z) = P(X \geq z)$. In the special case of two dimensions, if g and f have identical marginal distributions, then $\overline{G}(z) - \overline{F}(z) = G(z - e_1 - e_2) - F(z - e_1 - e_2)$, so

$$G(z) - F(z) \geq 0 \quad \forall z \in L \quad \Leftrightarrow \quad \overline{G}(z) - \overline{F}(z) \geq 0 \quad \forall z \in L. \tag{12}$$

Joe (1990) has defined a notion of greater interdependence for multivariate distributions which he terms the "concordance order": g dominates f according to the concordance order, written $f \prec_{\mathcal{CONC}} g$, if for all $z \in L$, both $G(z) - F(z) \geq 0$ and $\overline{G}(z) - \overline{F}(z) \geq 0$ hold. For bivariate distributions, by combining (11) and (12) we can conclude that

$$f \prec_{\mathcal{SPM}} g \quad \Leftrightarrow \quad f \prec_{\mathcal{CONC}} g. \tag{13}$$

The equivalence between the supermodular order and the concordance order for bivariate distributions is well known and has been proved by Levy and Parousch (1974), Epstein and Tanny (1980), and Tchen (1980). The latter two papers both developed constructive proofs that $f \prec_{\mathcal{CONC}} g$ implies $f \prec_{\mathcal{SPM}} g$ by defining a notion of a simple "correlation increasing" transformation.[1] Their proofs were considerably more complex than our argument above, for two reasons. First, they did not restrict their simple transformations to affect values at only *adjacent* points in the support. Second, they sought a weighted sequence of transformations that, when added to distribution f, yielded g *and* that produced, after each individual step, a probability distribution. Our Theorem 1 makes clear that, in searching for a decomposition of $g - f$ into a weighted sum $\sum_{t \in \mathcal{T}} \alpha_t t$, it is irrelevant whether or not partial sums of the form $f + \sum_{t \in \mathcal{U} \subset \mathcal{T}} \alpha_t t$ are actual probability distributions. And with elementary transformations defined as in (3), the decomposition of $g - f$ into $\sum_{t \in \mathcal{T}} \alpha_t t$ is, for two dimensions, unique, with $\alpha_{t^z} \equiv \alpha(z) = G(z) - F(z)$.[2]

Now

$$G(z) - F(z) = P(Y \leq z) - P(x \leq z) = EI_{\{Y \leq z\}} - EI_{\{X \leq z\}} = I^z \cdot (g - f),$$

[1] Levy and Paroush's proof assumed continuous distributions and used integration by parts.

[2] In Section 9, we provide an analogous characterization, in a unidimensional setting, of the convex ordering, also known in economics as the ordering of "greater riskiness", as characterized by Rothschild and Stiglitz (1970).

where $I^z(x) \equiv I_{\{x \leq z\}}$, the indicator function of the lower-orthant set $\{x | x \leq z\}$. Therefore, the nonnegativity requirement on the weights α_t in the unique decomposition of $g - f$ into $\sum_{t \in \mathcal{T}} \alpha_t t$ is equivalent to the requirement that, for all $z \in L$, the function $I^z(x)$ have a higher expectation under g than under f. These indicator functions of lower orthant sets are in fact the extreme rays of the cone of supermodular functions in two dimensions. An implication of the uniqueness, in two dimensions, of the decomposition $g - f = \sum_{t \in \mathcal{T}} \alpha_t t$ is that, in this special case, there is a one-to-one mapping associating with each elementary transformation $t^z \in \mathcal{T}$ the only extreme ray I^z of the cone of supermodular functions with which the transformation makes a strictly positive scalar product.

For more than two dimensions, however, many decompositions of $g - f$ into weighted sums of elementary transformations exist, and as a consequence such a one-to-one mapping between elementary transformations and extreme supermodular functions does not exist. In addition, for more than two dimensions, the supermodular ordering and the concordance ordering are no longer equivalent in general. These features make it considerably more difficult to determine, given a pair of distributions f and g, whether or not $f \prec_{\mathcal{SPM}} g$ when the underlying random vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ have three or more dimensions.

6 Constructive Methods for Comparing Distribution Interdependence

For three or more dimensions, how can one determine whether $f \prec_{\mathcal{SPM}} g$? We provide several answers to this question, all of which apply the characterization result in Theorem 1 and Theorem 3's result that all elementary transformations as defined in (3) are extreme.

The simplest approach involves specifying the sequence consisting of all elementary transformations, with attached weights, and then identifying, by construction, necessary and sufficient conditions on $g - f$ for the existence of a set of nonnegative weights such that the weighted sequence sums to $g - f$. This approach extends that adopted for two dimensions. However, because for more than two dimensions there is not a unique decomposition of $g - f$ into a weighted sum of ET's, it is impossible to apply the simple inductive process described in Section 5. Nevertheless, direct constructive methods can be used for other special cases, and they provide a number of insights into the structure of the supermodular ordering in higher dimensions. We have characterized the supermodular ordering in several such cases, and present three of them. The first, simplest example is the cube, that is, the case where $L = \{0,1\}^3$. The second example is the case where $L = \{0,1\}^4$ and where we confine attention to distributions satisfying a symmetry property that we term "top-to-bottom symmetry" (defined precisely below). The third example is the case where $L = \{0,1,2\}^3$ and where we impose a different form of symmetry, symmetry across dimensions. We defer discussion of this example until Section 7, where we analyze the symmetric supermodular ordering in detail.

A second approach to determining whether g is more interdependent than f is to formulate a linear program, based on the set of elementary transformations on L, such that the optimum value of the program is zero if and only if there exist non-negative coefficients $\{\alpha_t\}_{t\in\mathcal{T}}$ such that $g - f = \sum_{t\in\mathcal{T}} \alpha_t t$. This method, like the first approach, has the advantage of constructing an explicit sequence of elementary transformations that, added to f, result in g. However, it also has the drawback that one has to solve a different linear program for each pair of distributions to be compared.

A third method, based on Minkowski's and Weyl's representation theorems for polyhedral cones, allows one to compute once and for all, for any given support L, a minimal set of inequalities that characterize the stochastic supermodular ordering, such that $f \prec_{\mathcal{SPM}} g$ if and only if the vector $g - f$ satisfies these inequalities. This method can be used for optimization problems such as mechanism design or analysis of optimal policy, where each mechanism or policy generates a multivariate distribution, and the set of mechanisms or policies is large. In such settings, one must compare many distributions, and this so-called "double description method" may significantly reduce computations.

6.1 Supermodular Ordering on the Three-Dimensional and Four-Dimensional Cubes

The material in subsection 6.1 now appears in Meyer and Strulovici (2010).

Consider the case of three dimensions, where each dimension has two points in its support, i.e., $L = \{0,1\}^3$. The difference vector $\delta = g - f$ is represented in Figure 1.

Since for f and g to be ranked according to the supermodular ordering it is necessary that they have identical marginal distributions, once the values of $\delta(1,1,1) \equiv a$, $\delta(0,1,1) \equiv b_1$, $\delta(1,0,1) \equiv b_2$, and $\delta(1,1,0) \equiv b_3$ are specified, the remaining values are determined. For

$L = \{0,1\}^3$, there are six elementary transformations, corresponding to the six faces of the cube. Denote the transformations on the three upper faces (those faces with $(1,1,1)$ as a vertex) $\bar{t}_{12}$, $\bar{t}_{13}$, $\bar{t}_{23}$, where $\bar{t}_{12}(1,1,1) = \bar{t}_{12}(0,0,1) = -\bar{t}_{12}(0,1,1) = -\bar{t}_{12}(1,0,1) = 1$, and $\bar{t}_{13}$ and $\bar{t}_{23}$ are defined analogously. Denote the transformations on the three lower faces (those with $(0,0,0)$ as a vertex) $\underline{t}_{12}$, $\underline{t}_{13}$, $\underline{t}_{23}$, where $\underline{t}_{12}(0,0,0) = \underline{t}_{12}(1,1,0) = -\underline{t}_{12}(0,1,0) = -\underline{t}_{12}(1,0,0) = 1$, and $\underline{t}_{13}$ and $\underline{t}_{23}$ are defined analogously. Also denote the weight on $\bar{t}_{ij}$ by $\overline{\alpha}_{ij}$ and that on $\underline{t}_{ij}$ by $\underline{\alpha}_{ij}$. Then a set of six weights $\{\overline{\alpha}_{ij}, \underline{\alpha}_{ij}\}_{i \neq j}$ constitutes a weighted decomposition of $\delta = g - f$ into a sum of ET's if and only if

$$a = \overline{\alpha}_{12} + \overline{\alpha}_{13} + \overline{\alpha}_{23} \quad \text{and} \quad \forall i,j,k \in \{1,2,3\}, i \neq j \neq k, \quad b_i = -\overline{\alpha}_{ij} - \overline{\alpha}_{ik} + \underline{\alpha}_{jk}. \tag{14}$$

By adding the first equation in (14) to each of the other three in turn, the four equations above can be transformed into

$$a = \overline{\alpha}_{12} + \overline{\alpha}_{13} + \overline{\alpha}_{23} \quad \text{and} \quad \forall i,j,k \in \{1,2,3\}, i \neq j \neq k, \quad a + b_i = \overline{\alpha}_{jk} + \underline{\alpha}_{jk}. \tag{15}$$

By Theorem 1, $f \prec_{\mathcal{SPM}} g$ if and only if there exist nonnegative weights $\{\overline{\alpha}_{ij}, \underline{\alpha}_{ij}\}_{i \neq j}$ satisfying (15).

Proposition 1 (Supermodular Ordering on the Three-Dimensional Cube) *If the support $L = \{0,1\}^3$, $f \prec_{\mathcal{SPM}} g$ if and only if $f \prec_{\mathcal{C}} g$.*

Since the indicator functions $I_{\{x \geq z\}}$ and $I_{\{x \leq z\}}$ are both supermodular for all $z \in L$, it follows, as is well known, that for any support L, $f \prec_{\mathcal{SPM}} g$ implies $f \prec_{\mathcal{C}} g$. While Hu, Xie, and Ruan (2005, pp. 188-9) have proved the reverse implication for $L = \{0,1\}^3$ using the tool of "majorization with respect to weighted trees", we provide here a simple constructive proof.

Proof. First observe that $f \prec_{\mathcal{C}} g$ implies that f and g must have identical marginal distributions and that, with $L = \{0,1\}^3$ and identical marginals, $f \prec_{\mathcal{C}} g$ if and only if the following five inequalities are satisfied by the components of the difference vector δ (as defined in Figure 1:

$$a \geq 0, \quad a + b_i \geq 0 \quad \forall i \in \{1,2,3\}, \quad \text{and} \quad 2a + \sum_{i=1}^{3} b_i \geq 0. \tag{16}$$

The first four inequalities above correspond to $\overline{G}(z) - \overline{F}(z) \geq 0$ for z equal to $(1,1,1)$, $(0,1,1)$, $(1,0,1)$, and $(1,1,0)$, respectively. The fifth corresponds to $G(z) - F(z) \geq 0$ for $z = (0,0,0)$, given that f and g must have identical marginals.

We now show constructively that if δ satisfies the inequalities (16), then there exist non-negative weights $\{\overline{\alpha}_{ij}, \underline{\alpha}_{ij}\}_{i\neq j}$ satisfying (15). Set

$$\overline{\alpha}_{ij} = a\left(\frac{a+b_k}{3a+\sum_{i=1}^{3} b_i}\right) \quad \text{and} \quad \underline{\alpha}_{ij} = (2a+\sum_{i=1}^{3} b_i)\left(\frac{a+b_k}{3a+\sum_{i=1}^{3} b_i}\right). \tag{17}$$

It is apparent by inspection that the equations (15) are satisfied and that, if the inequalities in (16) hold, then $\overline{\alpha}_{ij} \geq 0$ and $\underline{\alpha}_{ij} \geq 0$. Therefore, it follows from Theorem 1 that $f \prec_{\mathcal{C}} g$ implies $f \prec_{\mathcal{SPM}} g$. ■

For three dimensions, if there is at least one dimension i for which L_i has cardinality greater than 2, then the supermodular order is strictly stronger than the concordance order. The following example proves this claim:

EXAMPLE 1: Let $L = \{0,1,2\} \times \{0,1\} \times \{0,1\}$ and let f, g have difference vector $g - f = \epsilon(t_{23}^{(0,0,0)} - t_{23}^{(1,0,0)} + t_{23}^{(2,0,0)})$, where $\epsilon > 0$. It is easy to check that $f \prec_{\mathcal{C}} g$. However, for the supermodular function $w(x) = \max\{(\sum_{i=1}^{3} x_i) - 2, 0\}$, $w \cdot (g - f) = \epsilon(2 \cdot 1 - 1 \cdot 1 - 1 \cdot 1 - 1 \cdot 1) < 0$, so it is not the case that $f \prec_{\mathcal{SPM}} g$. This example can be embedded in any support L strictly larger than $L = \{0,1,2\} \times \{0,1\} \times \{0,1\}$ to show that the same conclusion holds.

For four or more dimensions, Joe (1990) has provided an example showing that the supermodular order is strictly stronger than the concordance order, even when for each dimension i, $L_i = \{0,1\}$. In our notation, Joe's example has $g - f = \delta = \epsilon(t_{34}^{(0,0,0,0)} - t_{34}^{(1,0,0,0)} - t_{34}^{(0,1,0,0)} + t_{34}^{(1,1,0,0)})$ and $w(x) = \frac{1}{2}|(\sum_{i=1}^{4} x_i) - 1|$.

Nevertheless, the insight behind our constructive proof of Proposition 1 for the three-dimensional cube can be extended to characterize the supermodular ordering for larger supports, as we now illustrate for the case where $L = \{0,1\}^4$.

Consider four-dimensional random vectors with support $L = \{0,1\}^4$, and for simplicity confine attention to random vectors whose distributions satisfy a symmetry condition we term "top-to-bottom symmetry": For any $z \in \{0,1\}^4$, $P(X = z) = P(X = 1-z)$. Top-to-bottom symmetry arises naturally in various matching settings. For example, let the four dimensions represent managers, supervisors, workers, and firms, and suppose that for each dimension, there is one representative (individual or firm) with high quality ($z_i = 1$) and one with low quality ($z_i = 0$). Production requires forming a "team" consisting of exactly one manager, one supervisor, one worker, and one firm, and the output of such a team is a supermodular function of the qualities of each of its four components. Supermodularity of the production function implies that it would be output-maximizing for the four high-quality individuals/firm to be matched and for the four low-quality individuals/firm to be matched. However, informational frictions may prevent such an outcome being reached and cause the matching process to be stochastic. Nevertheless, as long as the stochastic process is certain to generate two teams, each consisting of one representative from each dimension, the distribution over teams satisfies "top-to-bottom symmetry". For such a setting, we now construct a set of inequalities for two distributions over teams (i.e., two matching processes) which are necessary and sufficient for one distribution to generate higher expected output than the other, for all supermodular functions.

Let the random vectors X and Y have distributions f, g on $L = \{0,1\}^4$ satisfying top-to-bottom symmetry. For any such f, g the difference vector $\delta = g - f$ can be represented as in Figure 2. Note that the assumption of top-to-bottom symmetry implies that f and g have identical marginal distributions and that for all i, $P(X_i = 1) = P(Y_i = 1) = 1/2$.

A construction analogous to that used for the three-dimensional cube allows us to prove:

Proposition 2 (Supermodular Ordering on the Four-Dimensional Cube) *If the support $L = \{0,1\}^4$ and f and g satisfy top-to-bottom symmetry, then $f \prec_{SPM} g$ if and only if*

$$P(\sum_{i=1}^{4} Y_i = 4) \geq P(\sum_{i=1}^{4} X_i = 4),$$
$$2P(\sum_{i=1}^{4} Y_i = 4) + P(\sum_{i=1}^{4} Y_i = 3) \geq 2P(\sum_{i=1}^{4} X_i = 4) + P(\sum_{i=1}^{4} Y_i = 3),$$
$$\text{and}\, \forall i \neq j, \quad P(Y_i = 1, Y_j = 1) \geq P(X_i = 1, X_j = 1).$$

In terms of the components of the difference vector δ, as defined in Figure 2, the inequalities in Proposition 2 correspond to

$$a \geq 0, \quad 2a + \sum_{i=1}^{4} b_i \geq 0, \quad \text{and} \quad a + b_i + b_j + c_{ij} \geq 0 \quad \forall i,j \in \{1,2,3,4\}, i < j.$$

Since in the example from Joe (1990) described above, f and g satisfy top-to-bottom symmetry, that example shows that even when we restrict attention to distributions satisfying top-to-bottom symmetry, the supermodular ordering on the four-dimensional cube is strictly stronger than the concordance ordering. The same conclusion follows from observing that there is no $z \in \{0,1\}^4$ for which the second inequality in Proposition 2 can be rewritten as $G(z) - F(z) \geq 0$ or $\overline{G}(z) - \overline{F}(z) \geq 0$.

6.2 The Linear Programming Approach: Comparing Two Specific Distributions

From Theorem 1, $f \prec_{\mathcal{SPM}} g$ if and only there exist nonnegative coefficients $\{\alpha_t\}_{t\in\mathcal{T}}$ such that $g - f = \sum_{t\in\mathcal{T}} \alpha_t t$. Given a specific pair of distributions f and g, we can formulate the problem of determining whether such a set of coefficients exists as a linear programming problem. Let $\tau = |\mathcal{T}|$ denote the number of elementary transformations on L, and let E denote the $d \times \tau$-matrix whose columns are the d-dimensional vectors consisting of all elementary transformations of L. Theorem 1 can be re-expressed as $f \prec_{\mathcal{SPM}} g$ if and only if there exists $\alpha \in \mathbb{R}^\tau$ such that i) $\alpha \geq 0$ and ii) $E\alpha = g - f$. Now define the d-dimensional vector δ^+ such that $\delta_i^+ = |(g-f)_i|$, and let E^+ denote the matrix whose i^{th} row, denoted E_i^+, satisfies $E_i^+ = (-1)^{\varepsilon_i} E_i$, where $\varepsilon_i = 1$ if $(g-f)_i < 0$ and 0 otherwise. The condition $E\alpha = g - f$ can be re-expressed as $E^+\alpha = \delta^+$. Now consider the following[3] linear program (A):

$$\min_{(\alpha,\beta)\in\mathbb{R}^\tau\times\mathbb{R}^d} \sum_{i=1}^{d} \beta_i$$

subject to

$$E^+\alpha + \beta = \delta^+, \quad \alpha \geq 0, \quad \beta \geq 0.$$

Theorem 4 (Pairwise Comparison) *The linear program (A) always has an optimal solution. $f \prec_{\mathcal{SPM}} g$ if and only if the optimum value is zero, and in that case $g = f + \sum_{t\in\mathcal{T}} \alpha_t^* t$, where (α^*, β^*) is any minimizer of (A) and $\beta^* = 0$.*

Proof. There always exists a feasible vector (α, β), namely $(\alpha, \beta) = (0, \delta^+)$. Moreover, the value function is nonnegative since the feasibility constraints require that β have nonnegative components, and therefore the optimum is nonnegative. If $f \prec_{\mathcal{SPM}} g$, there exists $\alpha^* \geq 0$ such that $E^+\alpha^* = \delta^+$, so the optimum value of program (A) must indeed be zero, since that value is achieved by $(\alpha, \beta) = (\alpha^*, 0)$. Reciprocally, if there exists (α^*, β^*) such that the value of the program is zero, then necessarily $\beta^* = 0$ and $E^+\alpha^* = \delta^+$. ■

[3] This corresponds to the auxiliary program for the determination of a basic feasible solution described in Bertsimas and Tsitsiklis (1997, Section 3).

6.3 The Double Description Method

The linear programming approach just described has the drawback of requiring a new program to be solved each time a new pair of distributions is to be compared. When many distributions are to be compared, for example as part of a larger optimization problem, it is more convenient to have an explicit representation of the stochastic supermodular ordering for the common support of these distributions. We now provide such a representation in the form of a list of inequalities that are satisfied by the vector $g - f$ if and only if $f \prec_{\mathcal{S}PM} g$. For any given finite support L, these inequalities are computed once and for all, a computation which is made possible by the support's finiteness.

Recall that $f \prec_{\mathcal{S}PM} g$ if $g - f$ makes a nonnegative scalar product with all supermodular functions on L, seen as vectors of $\mathbb{R}^d$. This condition can be reduced to a finite set of inequalities by exploiting the geometric properties of $\mathcal{S}$. $\mathcal{S}$ is a convex cone characterized by the fact that w is supermodular (i.e., belongs to $\mathcal{S}$) if and only if it makes a nonnegative scalar product with all elementary transformations on L. In matrix form, $\mathcal{S} = \{w \in \mathbb{R}^d : Aw \geq 0\}$, where $A = E'$ is the matrix whose rows consist of all elementary transformations (i.e., the transpose of the matrix E introduced earlier). A is called the *representation matrix* of the polyhedral cone $\mathcal{S}$. Minkowski's theorem states that to any representation matrix corresponds a *generating matrix* R such that

$$Ax \geq 0 \quad \Leftrightarrow \quad x = R\lambda \quad \text{for some } \lambda \geq 0.$$

The columns of the matrix R are the extreme rays of the cone $\mathcal{S}$. There exists a finite number of such extreme rays. The stochastic supermodular ordering is entirely determined by the extreme rays:

$$E[w|f] \leq E[w|g] \quad \forall w \in \mathcal{S} \quad \Leftrightarrow \quad R'(g - f) \geq 0.$$

Minkowski's theorem thus proves the existence, for any finite support L, of a finite list of inequalities that entirely characterize the stochastic supermodular ordering on L. How can we determine the extreme rays of the cone of supermodular functions? The *double description method*, conceived by Motzkin et al. (1953) and implemented by Fukuda and Prodon (1996) and Fukuda (2004), builds on Minkowski's and Weyl's representation theorems for polyhedral cones. A polyhedral cone can be represented either by a set of inequalities (i.e., by the intersection of a number half-spaces) or by extreme rays. The double description method provides an algorithm to determine one description from the other. Luckily, the set of elementary transformations is trivially computable, and can be automatically generated for any given support L. From this input, the double description method can compute the set of extreme supermodular functions. Using Fukuda's algorithm for the double description method, we have computed the stochastic supermodular order for a range of problems that are intractable by hand. In the Appendix, we illustrate the method for the case where $L = \{0,1\}^4$ and no symmetry assumptions of any sort are imposed.

Complexity of the Double Description Method Although the double description method is very useful in theory, its computational complexity is unsurprisingly exponential in the size of L. Keeping in mind the potential applications of the stochastic supermodular ordering, we now provide an exact computation of the algorithm's complexity.

Avis and Bremner (1995) show that the double description algorithm described by Motzkin et al. (1953) has complexity $O(p^{\lfloor d/2 \rfloor})$ where d is the dimension of the space and p is the number of inequalities defined by the representation matrix. Given a finite lattice $L = \times_{i=1}^n L_i$ of $\mathbb{R}^n$ with $|L_i| = l_i$, the dimension of the vector space generated by associating a dimension to each node of L is $d = \prod_{i=1}^n l_i$. To compute the number p of inequalities, first recall Theorem 3, which states that all of the elementary transformations $t \in \mathcal{T}$ are extreme, so it is impossible to reduce the number of inequalities required to check supermodularity by removing redundant elementary transformations. Therefore, p equals the number of elementary transformations on L, which it is straightforward to calculate:

$$p = \sum_{1 \le i < j \le n} (l_i - 1)(l_j - 1)\Pi_{k \notin \{i,j\}} l_k.$$

Suppose, for example, that l_i is exactly l for each of the n dimensions. Then $p = \frac{n(n-1)}{2}(l-1)^2 l^{n-2} \sim \frac{n(n-1)}{2} l^n$ and $d = l^n$. Therefore, the complexity of the double description method is $O(\exp(l^n(n \log l + 2 \log n)))$. In practice, therefore, the stochastic supermodular ordering can only be computed via this method for "small-size" problems. However, the "size" of a problem can be reduced by aggregating data into coarser categories. As Theorem 2 showed, aggregation of data preserves the supermodular ordering. Therefore, despite its potential complexity, the double description method can in practice easily be used in conjunction with data coarsening to achieve a tractable comparison of distributions.

7 Symmetric Supermodular Ordering

In many contexts, it is natural to assume that the supermodular objective functions being used to compare distributions are symmetric with respect to the components of the random vectors. We now define the symmetric supermodular ordering, show formally (Theorem 5) how it relates to the supermodular ordering, and provide some characterization results (Propositions 3 and 4). Theorem 6 develops useful sufficient conditions for the symmetric supermodular ordering to hold and applies these results to some welfare-economic and contract-theoretic examples.

Call a lattice $L = \times_{i=1}^{n} L_i$ symmetric if $L_i = L_j$ for all $i \neq j$. For a symmetric lattice, let the cardinality of L_i equal l, so the lattice has $d = l^n$ nodes. Let θ denote a real function on a symmetric lattice L, or equivalently a vector of $\mathbb{R}^d$. Depending on the context, θ can represent an objective function w or a probability distribution f. We will say that the function θ is *symmetric on* L if $\theta(z) = \theta(\sigma(z))$ for all $z \in L$ and for all permutations $\sigma(z)$ of z.

For distributions g and f on a symmetric lattice L, we will say that g dominates the distribution f according to the **symmetric supermodular ordering**, written $f \prec_{SSPM} g$, if and only if $E[w|f] \leq E[w|g]$ for all symmetric supermodular functions w on L.

For an arbitrary (not necessarily symmetric) function θ, the *symmetrized version of* θ, θ^{symm}, is defined as follows: for any z,

$$\theta^{symm}(z) = \frac{1}{n!} \sum_{\sigma \in \Sigma(n)} \theta(\sigma(z)), \tag{18}$$

where $\Sigma(n)$ is the set of all permutations of $\{1, \ldots, n\}$. Importantly, if w is a supermodular function, then w^{symm} is supermodular. For a symmetric supermodular function w, let $\mathcal{W}^{symm}(w)$ denote the set of supermodular functions $\hat{w}$ on L such that the symmetrized version of $\hat{w}$ is w, i.e., $\hat{w}^{symm} = w$. Note that $\{\mathcal{W}^{symm}(w)\}$ is a partition of the set of all supermodular functions on the symmetric lattice L.

We can now state the following useful result:

THEOREM 5 *Given a pair of distributions f, g defined on L, the following three statements are equivalent:*

i) $f \prec_{SSPM} g$*;*

ii) $f^s \prec_{SSPM} g^s$;

iii) $f^s \prec_{SPM} g^s$.

Proof. To show that i) $\Rightarrow$ ii) $\Rightarrow$ iii): If for all symmetric supermodular w, $w \cdot f \leq w \cdot g$, then for all symmetric supermodular w, $w \cdot f^s \leq w \cdot g^s$. This is ii). In turn, if for some symmetric supermodular w, $w \cdot f^s \leq w \cdot g^s$, then $\hat{w} \cdot f^s \leq \hat{w} \cdot g^s$ for all $\hat{w} \in \mathcal{W}^s(w)$. Therefore, $w \cdot f^s \leq w \cdot g^s$ for all symmetric supermodular w implies that $\hat{w} \cdot f^s \leq \hat{w} \cdot g^s$ for all supermodular $\hat{w}$, which is iii).

To show that iii) $\Rightarrow$ i): If for all supermodular w, $w \cdot f^s \leq w \cdot g^s$, then for all supermodular w, $w^s \cdot f^s \leq w^s \cdot g^s$. This is equivalent to $w^s \cdot f^s \leq w^s \cdot g^s$ for all symmetric supermodular w^s. This in turn implies that for all symmetric supermodular w^s, $w^s \cdot f \leq w^s \cdot g$. ■

In words, Theorem 5 states that one can characterize the symmetric supermodular order in terms of the supermodular order applied to symmetric distributions. Furthermore, when attention is restricted to symmetric distributions, the supermodular order is equivalent to the symmetric supermodular one. Theorem 5 can be used to simplify the analysis of the symmetric supermodular ordering by focusing on symmetric distributions.

This theorem is also important with respect to some economic applications of the theory, particularly welfare analysis. Indeed, focusing on symmetric ex post objective functions amounts to assuming a form of ex post anonymity across individuals: one does not care whether 1 got the high prize and 2 the low prize, or vice versa. However, it does not impose anything on the ex ante fairness of mechanisms. For example: a mechanism that randomizes with equal probability between giving the high prize to 1 and low prize to 2 or vice versa yields the same expected ex post welfare as a mechanism that always gives 1 a high prize and 2 a low one. A justification for the focus on symmetric distributions is precisely that we can make any mechanism fair ex ante by randomizing equally across all possible player permutations before applying the initial mechanism. In that sense, symmetry provides both an ex post anonymous and ex ante anonymous mechanism.

In the analysis to follow, we will mostly focus on the symmetric supermodular ordering, keeping in mind the interpretation in terms of symmetrized distributions provided by Theorem 5.

7.1 Binary Variables, n Dimensions

Consider the hypercube $L = \{0,1\}^n$. With symmetric objective functions, only the number of 1's, $c(x) = \sum_{i=1}^n I_{\{x_i=1\}}$, contained in any x matters for the objective. Thus, an equivalent representation of L is $\tilde{L}^1 = \{0,1,\ldots,n\}$. To any distribution f on L we can associate a distribution $\tilde{f}$ on $\tilde{L}^1$ defined by $\tilde{f}(k) = \sum_{x:c(x)=k} f(x)$ for each $k \in \tilde{L}^1$. Similarly, to any symmetric function $w : L \to \mathbb{R}$, corresponds another function $\tilde{w} : \tilde{L}^1 \to \mathbb{R}$ such that $w(x) = \tilde{w}(c(x))$.

Moreover, w is symmetric and supermodular on L if and only if $\tilde{w}$ is convex on $\tilde{L}^1$. To prove this, note that supermodularity of w is equivalent to $w(x) + w(x + e_i + e_j) \geq w(x + e_i) + w(x + e_j)$, for all nodes x with zero i^{th} and j^{th} components. Symmetry of w then allows us to write the inequality as $\tilde{w}(k) + \tilde{w}(k+2) \geq 2\tilde{w}(k+1)$, where $k = c(x)$. Since this holds for all $k \in \{0,1,\ldots,n-2\}$, this shows convexity of $\tilde{w}$. The reverse implication is proved similarly. An alternative method of proof, which will be generalized in the following subsection, is to use duality: recall from Theorem 1 that supermodular functions are characterized by the dual cone of elementary transformations, $t^x_{i,j}$, as defined in (3). When the transformation $t^x_{i,j}$, defined on L, is projected onto $\tilde{L}^1$, the result is an elementary transformation of the form $\tilde{t}^k$ such that $\tilde{t}^k(k) = \tilde{t}^k(k+2) = 1$, $\tilde{t}^k(k+1) = -2$, and $\tilde{t}^k(y) = 0$ for all other $y \in \tilde{L}^1$, where $k = c(x)$. Such a function $\tilde{t}^k$ is an elementary transformation characterizing convexity on a one-dimensional, equally-spaced grid (see Section 9). Since their dual cones are equivalent, it follows therefore that symmetric supermodular functions on L are equivalent to convex functions on $\tilde{L}^1$.

This shows a key relation between supermodularity and convexity:

Proposition 3 *On $L = \{0,1\}^n$, $f \prec_{SSPM} g$ if and only if $\tilde{g}$ dominates $\tilde{f}$ according to the convex ordering on $\tilde{L}^1$.*

7.2 l-Point Supports and n Dimensions

Now consider the case $L = \{0,1,\ldots,l-1\}^n$, and for $k \in \{1,\ldots,l-1\}$ and $x \in L$, define $\bar{c}^k(x) = \sum_{i=1}^n I_{\{x_i \geq k\}}$ and $\bar{c}(x) = (\bar{c}^1(x),\ldots,\bar{c}^{l-1}(x))$. $\bar{c}^k(x)$ counts the number of components of x that are at least as large as k, and $\bar{c}(x)$ is the "cumulative count vector" corresponding to x. The vector $\bar{c}(x)$ lies in $\tilde{L}^{l-1}$, an $(l-1)$-dimensional subset of $\{0,1,\ldots,n\}^{l-1}$. Any function $w : L \to \mathbb{R}$ that is symmetric can be expressed as a function $\tilde{w} : \tilde{L}^{l-1} \to \mathbb{R}$ such that $w(x) = \tilde{w}(\bar{c}(x))$. To any distribution f on L we can associate a distribution $\tilde{f}$ on $\tilde{L}^{l-1}$ defined by $\tilde{f}(y) = \sum_{x:\bar{c}(x)=y} f(x)$ for each $y \in \tilde{L}^{l-1}$.

To generalize Proposition 3, we need the following definition:

A function $\tilde{w}$ on $\tilde{L}^{l-1}$ is *componentwise-convex* if for any $y \in \tilde{L}^{l-1}$ and $k = \{1, 2, \ldots, l-1\}$ such that $y + 2e_k \in \tilde{L}^{l-1}$, $\tilde{w}(y) + \tilde{w}(y + 2e_k) \geq 2\tilde{w}(y + e_k)$.[4] Equivalently, $\tilde{w}$ on $\tilde{L}^{l-1}$ is componentwise-convex if and only if it makes a positive scalar product with any elementary transformation defined by a function t_k^y on $\tilde{L}^{l-1}$ such that

$$t_k^y(y) = t_k^y(y + 2e_k) = 1 \qquad\qquad t_k^y(y + e_k) = -2, \tag{19}$$

and $t_k^y(z) = 0$ for all other $z \in \tilde{L}^{l-1}$.

Proposition 3 can now be generalized to

PROPOSITION 4 *On $L = \{0, 1, \ldots, l-1\}^n$, $f \prec_{SSPM} g$ if and only if $\tilde{g}$ dominates $\tilde{f}$ according to the supermodular and componentwise-convex ordering on $\tilde{L}^{l-1}$.*

Proposition 4 is proved by showing that w is symmetric and supermodular on L if and only if $\tilde{w}$ is supermodular and componentwise-convex on $\tilde{L}^{l-1}$.[5] To show this, we use the dual approach and show that any "supermodular" elementary transformation $t_{i,j}^x$ on L, as defined in (3), maps either into a "supermodular" elementary transformation on $\tilde{L}^{l-1}$ or into an elementary transformation on $\tilde{L}^{l-1}$ of the form in (19) characterizing componentwise-convexity. For transformations $t_{i,j}^x$ such that $\bar{c}(x + e_i) = \bar{c}(x + e_j)$, there exists a $k \in \{1, \ldots, l-1\}$ such that $\bar{c}(x + e_i) = \bar{c}(x + e_j) = \bar{c}(x) + e_k$ and $\bar{c}(x + e_i + e_j) = \bar{c}(x) + 2e_k$; therefore, such transformations on L map into transformations on $\tilde{L}^{l-1}$ of the form in (19). For transformations $t_{i,j}^x$ such that $\bar{c}(x + e_i) \neq \bar{c}(x + e_j)$, there exist $k, m \in \{1, \ldots, l-1\}$ such that $\bar{c}(x + e_i) = \bar{c}(x) + e_k$, $\bar{c}(x + e_j) = \bar{c}(x) + e_m$, and $\bar{c}(x + e_i + e_j) = \bar{c}(x) + e_k + e_m$; therefore, these transformations on L map into transformations on $\tilde{L}^{l-1}$ of the form in (3).

Meyer and Strulovici (2010) provide an explicit characterization of the symmetric supermodular ordering for the case of three dimensions with three points in the support for each dimension.

---

[4]See Section 9 for more detail.

[5]Since $\tilde{L}^{l-1}$ is not a lattice, we say that $\tilde{w}$ is supermodular if, whenever y and z belong to $\tilde{L}^{l-1}$ and are such that $y \wedge z$ and $y \vee z$ also belong to $\tilde{L}^{l-1}$, where the meet and join operate on $\mathbb{R}^{l-1}$, $\tilde{w}(y \wedge z) + \tilde{w}(y \vee z) \geq \tilde{w}(y) + \tilde{w}(z)$.

Propositions 3 and 4 are useful because, even as the dimension of the underlying support L increases, the dimensions of the derived supports $\tilde{L}^1$ and $\tilde{L}^{l-1}$ remain unchanged.

7.3 Sufficient Conditions for the Symmetric Supermodular Ordering

Let A and B denote two $n \times m$ *row-stochastic* matrices, i.e., matrices such that each row has nonnegative components which sum to 1. Also suppose that for each $j \leq m$, the j^{th} column of A and B have equal sum.

For concreteness, think of each row of A as describing the lottery among m prizes to some individual i, for $i \leq n$. The first column corresponds to the lowest prize, the second column to the second-lowest, etc. Let these lotteries be *independently* distributed across individuals. Thus $A_{i,j}$ is the probability that i receives prize j independently of what others receive. We will call X and Y the random vectors of prizes that individuals receive under distributions defined by A and B, respectively.

For an arbitrary row-stochastic matrix Q, let $\bar{Q}$ denote the *cumulative sum matrix* of Q, defined by $\bar{Q}_{i,j} = \sum_{k=j}^{m} Q_{i,k}$. There is a one-to-one mapping between row-stochastic matrices and their cumulative-sum equivalents, so slightly abusing notation we will use $\bar{A} \prec_{SSPM} \bar{B}$, $A \prec_{SSPM} B$, and $X \prec_{SSPM} Y$ equivalently.

Say that Q is *stochastically ordered* if for each k, $\bar{Q}_{i,k}$ is weakly increasing in i. This is equivalent to the requirement that for all $i \in \{2, \ldots, n\}$, the ith row of Q dominates the $(i-1)$th row in the sense of first-order stochastic dominance. Intuitively, this means that under the distribution described by Q, high-index individuals are more likely to receive high prizes.

Given an arbitrary row-stochastic matrix Q and its associated cumulative sum matrix $\bar{Q}$, define $\bar{Q}^{so}$ as the matrix obtained from $\bar{Q}$ by reordering each of its columns from the smallest to the largest element. If Q is stochastically ordered, then $\bar{Q}^{so} = \bar{Q}$. We will say that A dominates B according to the *cumulative column majorization criterion*, denoted $A \succ_{CCM} B$, if for all k, the k^{th} column vector of $\bar{A}$ majorizes[6] the k^{th} column vector of

[6] A vector a majorizes a vector b if i) the vectors have identical sums, and ii) for all k, the sum of the k largest entries of a is weakly greater than the sum of the k largest entries of b (see Hardy, Littlewood, and Polya (1952)).

$\bar{B}$. That is, $A \succ_{CCM} B$ if for each $k \in \{1, \ldots, m\}$ and for each $l \in \{1, \ldots, n\}$

$$\sum_{i=l}^{n} \bar{A}^{so}_{i,k} \geq \sum_{i=l}^{n} \bar{B}^{so}_{i,k},$$

with equality holding for $l = 1$.

THEOREM 6 *Let A and B be two $n \times m$ row-stochastic matrices such that, for each $j \leq m$, the jth column of A and B have equal sums. If A is stochastically ordered and $A \succ_{CCM} B$, then $X \prec_{SSPM} Y$.*

There are several ways to interpret and apply Theorem 6. Recall that Theorem 5 showed that the statements $f \prec_{SSPM} g$ and $f^s \prec_S g^s$ are equivalent. In this context, this means that using the symmetric supermodular order to compare the distributions generated by the independent lotteries over prizes described by matrices A and B is equivalent to using the supermodular order to compare the symmetrized versions of these distributions. Importantly, the symmetrized versions of these distributions are not independent, so supermodular dominance of one symmetrized distribution over another reflects greater interdependence of the former over the latter. Thus, the original comparison of independent distributions according to the symmetric supermodular ordering can be interpreted as a comparison of interdependence of symmetrized distributions. Theorem 6 provides a sufficient condition for the symmetrized version of the distribution generated by the set of lotteries in matrix B to display greater interdependence than the symmetrized version of the distribution generated by A.[7]

To illustrate this interpretation of the theorem, suppose that $m = n$ (the number of prizes equals the number of individuals) and that we focus on matrices A and B that are bistochastic, i.e., both their rows and their columns all sum to 1. A tournament is a mechanism that allocates, according to some random process, the n prizes to the n individuals in such a way that each individual receives exactly one prize. Any tournament is fully described by the probability it assigns to each of the $n!$ possible prize allocations, and a tournament can be summarized by a bistochastic matrix Q, where the ith row of

[7] Hu and Yang (2004, Thm. 3.4) showed that for any stochastically ordered row-stochastic matrix A, the symmetrized version of the distribution of X (which is not in general independent) is supermoduarly dominated by the independent symmetric distribution with identical marginals to the symmetrized version of X. (In fact, Hu and Yang proved this result by showing something stronger, that the symmetrized version of the distribution of X displays negative association.) Hu and Yang's result for supermodular dominance corresponds to the special case of Theorem 6 where the rows of the matrix B are all identical.

Q describes individual i's marginal distribution over the n prizes. A symmetric tournament is one in which each of the $n!$ possible prize allocations is equally likely, and such a tournament is summarized by the bistochastic matrix all of whose entries are $1/n$. Given an arbitrarily asymmetric tournament and the bistochastic matrix Q which summarizes the marginal distributions it generates, consider the reward scheme which gives each individual the same marginal distribution over rewards as he receives in the tournament but which determines rewards independently. We term this reward scheme the "randomized independent scheme" (RIS) associated with the given tournament. Theorem 6 implies that given any asymmetric tournament, the associated RIS generates a distribution over rewards that dominates the distribution generated by the tournament according to the symmetric supermodular ordering.

To see why this conclusion follows from the theorem, let A be the $n \times n$ identity matrix and B the bistochastic matrix summarizing the marginal distributions over prizes generated by an arbitrary asymmetric tournament T. What is the symmetrized version of the distribution generated by the independent (degenerate) lotteries in A? It is the distribution which assigns probability $1/(n!)$ to each of the $n!$ possible allocations of prizes to individuals in any tournament. This symmetric distribution is in fact the symmetrized version of the distribution of prizes resulting from any, arbitrarily asymmetric tournament. The symmetrized version of the distribution generated by B is the symmetrized version of the distribution of prizes under the RIS associated with the original tournament T. When the matrix A is the identity matrix, it is clearly stochastically ordered, and it also clearly dominates any other bistochastic matrix according to the cumulative column majorization criterion. Therefore, the symmetrized version of the distribution generated by A is supermoduarly dominated by the symmetrized version of the distribution generated by B. Equivalently, for any symmetric supermodular objective function, expected welfare is lower under any arbitrary tournament than under the RIS associated with it.

Theorem 6 has applications outside the welfare-economic context discussed above. Suppose that row i of the row-stochastic matrix Q now represents the distribution of output on the ith of n tasks over the m possible output levels, indexed by j, and suppose that output levels on the different tasks are independently distributed. Suppose that the production function is supermodular in the output levels on the different tasks, reflecting the fact that tasks are complementary inputs, and suppose also that tasks are identical ex ante, so the production function is symmetric with respect to the vector of task outputs. Two row-stochastic matrices with matching column sums then describe two different production settings in which, for each possible output level, the average probability (over all tasks) of its being realized is the same. Theorem 6 then identifies conditions under which expected production is higher in one setting than another for all symmetric supermodular production functions.

Bond and Gomes (2009) have recently analyzed a special case of the setting just described. An agent chooses levels of effort $\{e_i\}$ on n tasks, where $e_i \in [\underline{e}, \overline{e}]$. For each task, output is either success or failure, and by exerting effort e_i on task i, the agent incurs total effort cost $\sum_{i=1}^{n} e_i$ and produces a probability of success on task i of e_i. Given the effort choices, the outputs are independently distributed. The principal's benefit is a convex function of the total number of successes. Bond and Gomes ask, for a given total amount of effort $\sum_{i=1}^{n} e_i < n$ (and, hence, given total cost of effort), what is the socially efficient allocation of effort across tasks? They show that it is socially efficient for the agent to exert equal effort on all tasks. However, under any incentive scheme rewarding him as a function of the total number of successes achieved, the agent will choose either the minimum ($\underline{e}$) or the maximum ($\overline{e}$) level of effort on each task. Bond and Gomes show that, given the total amount of effort exerted, the allocation chosen by the agent actually minimizes expected social surplus.

The two conclusions summarized above follow from Proposition 3 and Theorem 6. With binary output levels on the tasks and a benefit function for the principal that is symmetric across tasks, the benefit function can be described either as a convex function of the total number of successes or as a symmetric supermodular function of the vector of task outputs. The effort allocation determines an $n \times 2$ row-stochastic matrix, the first column of which is the vector of success probabilities on the n tasks, and holding the total level of effort fixed corresponds to ensuring that any matrices being compared have matching column sums. In the special case where $m = 2$, any row-stochastic matrix can be converted into a stochastically ordered one by reordering rows (an operation which will have no effect on the expected value of a symmetric objective function). Therefore, with $m = 2$, we can deduce from Theorem 6 that, holding total effort fixed, if one effort allocation corresponds to a vector of success probabilities that majorizes the vector corresponding to another allocation, then the former allocation generates lower expected social surplus, for all symmetric supermodular benefit functions. The final step is to observe that a vector of success probabilities in which all entries are equal is majorized by all vectors with the same total over entries; and a vector in which all probabilities are either 0 or 1 majorizes all vectors with the same total (which are not permutations of it).[8]

We have examples showing that Theorem 6 does not hold if we relax either the assumption that A is stochastically ordered or that $A \succ_{CCM} B$.

Theorem 6 has the following useful corollary, which is proved in the Appendix.

Corollary 1 *For any n and any m-dimensional probability vector p, there exists a unique $n \times m$ row-stochastic matrix A' whose j^{th} column, for each j, sums to np_j and such that $A' \prec_{SSPM} B$ for all $n \times m$ row-stochastic matrices B with the same column sums as A'.*

In settings where the objective function is symmetric and supermodular, the corollary identifies, within the class of distributions generated from row-stochastic matrices as described above, the *worst* distribution. Equivalently, where the objective function is symmetric and submodular, the corollary identifies the *optimal* distribution within the specified class. For arbitrary n, m, and probability vector p, the matrix A' identified by the corollary is the one in which the lotteries described by the rows are as disparate as possible, subject to their average equaling the vector p. In the welfare-economic context described above, the matrix A' is the one that treats individuals as differently as is consistent with the constraint on the average distribution of rewards. In the production context, the matrix A' is the one in which the resources allocated to the various tasks are as different as is feasible, given the overall resource constraint.

8 Using the Supermodular Ordering to Compare Interdependence of Mixture Distributions

Consider multivariate distributions generated as follows: first, a univariate probability distribution is drawn randomly, according to some distribution. Then, all random variables are drawn independently from that common distribution. The resulting multivariate distribution is a *mixture* of conditionally i.i.d. random variables. Since the common distribution is ex ante uncertain, this creates some positive dependence between the random

[8] Bond and Gomes's results follow from a result due to Karlin and Novikoff (1963), which is the special case of Theorem 6 when $m = 2$.

variables.[9] Intuitively, learning that $X_i = k$ raises the posterior probability that the common distribution is one which assigns high probability to k, and hence makes it more likely that $X_j = k$ for $j \neq i$. This section compares the interdependence of two random vectors each of which is a mixture of conditionally i.i.d. random variables. Specifically, we provide sufficient conditions for two (symmetric) mixture distributions to be ranked according to the supermodular ordering. The sufficient conditions we identify have a natural interpretation as a non-parametric ordering of the relative size of aggregate vs. idiosyncratic shocks.

The results of this section are useful in finance and insurance contexts, where mixtures of conditionally i.i.d. random variables are frequently used to model positively dependent risks in a portfolio: the realization of the common distribution represents an aggregate shock or common factor which affects all the elements of the portfolio (see, for example, Cousin and Laurent, 2008). Our results also have applications in macroeconomics, where assessments of the relative importance of aggregate vs. sectoral shocks are of importance for understanding variation and covariation of output levels (see, for example, Foerster, Sarte, and Watson, 2008).

Consider a univariate distribution on the support $\{0, \ldots, l-1\}$, described by its upper cumulative vector $\bar{p}$, i.e., $\bar{p}_k = Pr[X \geq k]$. Let $(X_1, \ldots, X_n)$ be i.i.d. variables with distribution $\bar{p}$. Given a supermodular objective function w on $\mathbb{R}^n$, define $u^w(\bar{p})$ by

$$u^w(\bar{p}) = E[w(X_1, X_2, \ldots, X_n)|\bar{p}].$$

The function u^w is defined on a convex subset of the vector space $\mathbb{R}^l$, and inherits some properties from the supermodularity of w, as shown in the following.

Proposition 5 *If w is supermodular, u^w is supermodular and componentwise convex.*

Proof. Changing any component $\bar{p}_k$ affects all random variables and hence has a complicated effect on u^w. To prove the result, it is therefore useful to consider, as an intermediate step, a more general domain where each of the independent variables X_i has its own distribution vector $\bar{p}^i$ on the support $\{0, \ldots, l-1\}$. Accordingly, define

$$v^w(\bar{p}^1, \ldots, \bar{p}^n) = E[w(X_1, \ldots, X_n)|\bar{p}^1, \ldots, \bar{p}^n].$$

We use the following lemma.

[9] Shaked (1977) defines such random variables as "positively dependent by mixture".

Lemma 1 *For any supermodular w, $v^w(\bar{p}^1, \ldots, \bar{p}^n)$ has the following properties:*

- $\frac{\partial^2 v}{\partial \bar{p}^i_r \partial \bar{p}^i_s} = 0$ *for all* $i \in \{1, \ldots, n\}$ *and* $r, s \in \{0, \ldots, l-1\}$.
- $\frac{\partial^2 v}{\partial \bar{p}^i_r \partial \bar{p}^j_s} \geq 0$ *for all* $i \neq j \in \{1, \ldots, n\}$ *and* $r, s \in \{0, \ldots, l-1\}$.

The first part of the lemma is standard, and comes from linearity of the objective with respect the probability distribution, which holds also in terms of the cumulative distribution vector. The second part comes from supermodularity of w. Indeed, by the discrete equivalent of an integration by parts,[10] we have

$$\frac{\partial v}{\partial \bar{p}^i_r} = E[w(X_{-i}, r) - w(X_{-i}, r-1)],$$

and, applying the same transformation to the (difference) function $w(x_{-i}, r) - w(x_{-i}, r-1)$,

$$\frac{\partial^2 v}{\partial \bar{p}^i_r \partial \bar{p}^j_s} = E[w(X_{-(i,j)}, r, s) + w(X_{-(i,j)}, r-1, s-1) - w(X_{-(i,j)}, r-1, s) - w(X_{-(i,j)}, r, s-1)],$$

which is nonnegative, by supermodularity of w.

To conclude the proof of the proposition, observe that $u(\bar{p}) = v(\bar{p}, \ldots, \bar{p})$. Second-order derivatives of u only involve second-order derivatives of v. The above lemma then shows the result. ■

We now compare different ways of generating the common distribution $\bar{p}$.

We will compare two matrices, $\bar{A}$ and $\bar{B}$, which both have l columns and q rows. For each matrix, each row is a cumulative probability distribution of the form analyzed above. A matrix generates a mixture distribution of the type defined earlier, where the univariate distribution $\bar{p}$ can take q possible values, and we assume here that each distribution is equally likely to be selected.[11] Matrix $\bar{A}$ generates the mixture distribution for X and $\bar{B}$ the mixture distribution for Y. We assume that $\bar{A}$ and $\bar{B}$ have *identical column sums.* This ensures that, for each realization k, the expected probability that $X_i \geq k$ equals the expected probability that $Y_i \geq k$, in other words, that the common marginal distribution of the X_i is the same as the common marginal distribution of the Y_i.[12]

[10] The continuous integration by parts would be $\int u(x)dG(x) = \int u'(x)F(x)$, where G is the usual cumulative distribution and F is the upper cumulative distribution.

[11] Generalizations are easy. For example, one could replicate rows to give particular distributions arbitrarily more weight.

[12] We do *not* assume that rows put strictly positive weight on all outcomes. In terms of interdependence, we therefore allow, for example, that observing the outcome from one random variable rules out some distributions that did not put any weight on that particular outcome.

In the theorem below, the hypotheses are the same as those of Theorem 6. But the manner in which multivariate distributions are generated from matrices is completely different here and in Theorem 6. And the conclusions of the two theorems are also different.

THEOREM 7 *Suppose that X and Y are random vectors generated by matrices $\bar{A}$ and $\bar{B}$, respectively, such that*

- *$\bar{A}$ is stochastically ordered, i.e., $\bar{A}_{i,k}$ is weakly increasing in i for all k.*
- *$\bar{A}$ dominates $\bar{B}$ according to the* cumulative column majorization criterion*, i.e., for each k, the k^{th} column vector of $\bar{A}$ majorizes the k^{th} column vector of $\bar{B}$.*

Then $X \succ_{SPM} Y$.

The random vectors X and Y have symmetric distributions so that, as far as their comparison is concerned, the supermodular ordering is equivalent to the symmetric supermodular ordering (see Theorem 5). Just as for Theorem 6, we have examples showing that Theorem 7 does not hold if we relax either the assumption that A is stochastically ordered or that $A \succ_{CCM} B$.[13]

The proof of Theorem 7, which is in Section D of the Appendix, is based on the following lemma.

LEMMA 2 *Suppose that $q = 2$ and that there exists a nonnegative vector ε such that for all $k \in \{1, l-1\}$,*

- $\bar{B}(1,k) = \bar{A}(1,k) + \varepsilon_k$
- $\bar{B}(2,k) = \bar{A}(2,k) - \varepsilon_k$
- $\bar{A}(2,k) \geq \bar{A}(1,k) + \varepsilon_k$

Then, $X \succ_{SPM} Y$.

[13] Jogdeo (1978) showed that for any stochastically ordered row-stochastic matrix A, the distribution of X generated from it displays association, a dependence concept defined in Esary, Proschan, and Walkup (1967). It follows from this and Theorem 2 of Meyer and Strulovici (2010) that the distribution of X dominates its independent counterpart (the independent distribution with identical marginals to X) according to the supermodular ordering. Jogdeo's result, weakened to supermodular dominance, corresponds to the special case of Theorem 7 where the rows of the matrix B are all identical.

The function $u = u^w$ is polynomial in $\bar{p}$ and hence twice differentiable. Moreover, it is componentwise convex and supermodular, which implies that its second-order derivatives are everywhere nonnegative on its domain. We need to show that for any vectors x, y and $\varepsilon \geq 0$ such that $x + \varepsilon \leq y$, the following inequality holds

$$u(x) + u(y) \geq u(x + \varepsilon) + u(y - \varepsilon)$$

Equivalently, we need to show that

$$u(x+\varepsilon) - u(x) = \int_0^1 \sum_i u_i(x + \alpha\varepsilon)\varepsilon_i d\alpha \leq \int_0^1 \sum_i u_i(y - \varepsilon + \alpha\varepsilon)\varepsilon_i d\alpha = u(y) - u(y - \varepsilon),$$

where u_i denotes the i^{th} derivative of u. Let $\delta = y - \varepsilon - x \geq 0$.

$$u_i(y - \varepsilon + \alpha\varepsilon) - u_i(x + \alpha\varepsilon) = \int_0^1 \sum_j u_{ij}(x + \alpha\varepsilon + \beta\delta)\delta_j d\beta,$$

which is nonnegative since all second-order derivatives are nonnegative. Integrating these inequalities with respect to α shows the result.

The hypotheses of Theorem 7 ensure that the rows of the matrix $\bar{A}$ are "more different" from one another than are the rows of the matrix $\bar{B}$. Since the rows represent the possible cumulative probability distributions from which the n variables are independently drawn, the hypotheses ensure that for the random vector X, these distributions are more different than for the random vector Y. Given that the X_i have the same marginal distribution as the Y_i (ensured by the requirement that $\bar{A}$ and $\bar{B}$ have identical column sums), the conditions in Theorem 7 can be interpreted as ensuring that aggregate shocks are relatively more important in the distribution of X while idiosyncratic shocks are relatively more important in the distribution of Y. At one extreme, where the matrix $\bar{B}$ has all rows identical, the mixture distribution reflects no common shock; at the other extreme, where the matrix A takes the form of the matrix $\bar{A}'$ identified by Corollary 1, the mixture distribution displays as much common uncertainty as possible, given the specified marginal distribution.

9 Characterizations of Difference-Based Orderings

This section generalizes the approach of Section 4 to provide characterizations of a class of orderings which we call "difference-based orderings," which have a particular linear structure which allows the use of duality theorems. We use the general duality approach to characterize orders combining supermodularity and componentwise convexity, or full convexity. Since convexity on lattices is a nontrivial concept, we also show how to characterize it in terms of elementary transformations, which is an interesting result in itself.

Recall from (1) that any class $\mathcal{W}$ of functions on L defines an order by $f \prec_{\mathcal{W}} g \Leftrightarrow E[w|f] \leq E[w|g] \quad \forall w \in \mathcal{W}$. We begin by stating formally the intuitive fact that larger classes of functions make it harder to compare distributions, hence result in a coarser order.

Theorem 8 (Order Monotonicity) *If $\mathcal{C} \subset \mathcal{D}$ and $f \prec_{\mathcal{D}} g$, then $f \prec_{\mathcal{C}} g$.*

Proof. Trivial and omitted.

Theorem 8 implies that any property of the order generated from a class of objective functions must be inherited by the order generated from any larger class of objective functions. This implication is illustrated in the next result, which implies that if g dominates f according to the stochastic supermodular ordering, then $Cov(Y_i, Y_j) \geq Cov(X_i, X_j)$ for any $i \neq j$ and random vectors X and Y respectively distributed according to f and g.

The Quadratic Ordering We now consider the subset $\mathcal{Q}$ of supermodular functions that are quadratic, i.e., of the form[14] $w(x) = w_0 + \sum_i w_i x_i + \sum_{i \neq j} w_{ij} x_i x_j$ for some real coefficients w_0 and $\{w_i\}$ and some nonnegative coefficients $\{w_{ij}\}_{i \neq j}$. Such functions are supermodular, as is easily checked. Let X and Y denote random vectors distributed according to f and g, respectively.

Theorem 9 (Quadratic Ordering) *$f \prec_{\mathcal{Q}} g$ if and only if $E[X_i] = E[Y_i]$ for all i and $Cov(X_i, X_j) \leq Cov(Y_i, Y_j)$ for all $i \neq j$.*

Proof. Since for all i, the functions $w(x) = x_i$ and $w(x) = -x_i$ are in $\mathcal{Q}$, $f \prec_{\mathcal{Q}} g$ implies that $\sum_{x_i \in L_i} x_i f_i \leq \sum_{x_i \in L_i} x_i g_i$ and $\sum_{x_i \in L_i} x_i f_i \geq \sum_{x_i \in L_i} x_i g_i$, where f_i (resp. g_i) is f's (resp. g's) marginal distribution along the i^{th} component. Therefore, $E[X_i] = E[Y_i]$ for all i. Since for all $i \neq j$, $w(x) = x_i x_j$ is in $\mathcal{Q}$, $f \prec_{\mathcal{Q}} g$ implies that $E[X_i X_j] \leq E[Y_i Y_j]$ for all $i \neq j$. To prove the reverse implication, observe that for any $w(x) = w_0 + \sum_i w_i x_i + \sum_{i \neq j} w_{ij} x_i x_j$ for some real coefficients w_0 and $\{w_i\}$ and some nonnegative coefficients

[14] We rule out functions x_i^2 in order to get an equivalence in the next theorem. For the entire class of supermodular quadratic functions, necessity of covariance relations is implied by combining Theorems 9 and 8.

$\{w_{ij}\}_{i\neq j}$,

$$E[w|g] - E[w|f] = \sum_i w_i \left(E[Y_i] - E[X_i]\right) + \sum_{i\neq j} w_{ij} \left[Cov(Y_i, Y_j) - Cov(X_i, X_j)\right] \geq 0,$$

so $E[X_i] = E[Y_i]$ for all i and $Cov(X_i, X_j) \leq Cov(Y_i, Y_j)$ for all $i \neq j$ imply $f \prec_{\mathcal{Q}} g$. ■

The Componentwise Convex/Concave Ordering In several applications, objective functions may have other properties than supermodularity. For example, if the objective is a welfare function and each variable entering the multivariate distribution represents the random income of an individual, componentwise concavity may express the social planner's preference for reducing risk faced by each individual. We now show how the duality approach in the case of the stochastic supermodular ordering can be extended to such situations. In what follows, we consider the case of objective functions that are supermodular and componentwise convex, but the case of supermodular, componentwise concave objective functions can be analyzed similarly.

In Section 4, we used the fact that supermodular functions are characterized by a list of inequalities which correspond to nonnegativity of their scalar product with all elementary transformations of the type defined in 3. To accommodate the introduction of other types of elementary transformations, let $\mathcal{T}(\mathcal{S})$ denote the set of elementary transformations characterizing $\mathcal{S}$.

A function w is componentwise convex if for any i in N and x, y in L such that $x_j = y_j$ for all $j \neq i$ and any $\lambda \in [0, 1]$ such that $\lambda x + (1-\lambda)y$ belongs to L, $w(\lambda x + (1-\lambda)y) \leq \lambda w(x) + (1-\lambda)w(y)$. Let $\mathcal{X}$ denote the set of componentwise convex functions on L.

To simplify the exposition, we assume that for each $i \in N$, $L_i = \{0, 1, \ldots, l_i - 1\}$, that is, in each dimension, the points in the support are equally spaced. We briefly discuss below how to extend our characterizations to more general lattices.

For any x and i, let t_i^x denote the function on L that vanishes everywhere except at nodes x, $x + e_i$, and $x + 2e_i$, such that

$$t_i^x(x) = t_i^x(x + 2e_i) = 1 \quad \text{and} \quad t_i^x(x + e_i) = -2, \tag{20}$$

and let $\mathcal{T}(\mathcal{X})$ denote the set all such functions. When added to the distribution of a random vector Y, the transformation t_i^x leaves the marginal distributions of Y_j, $j \neq i$, unaffected and increases the spread of the marginal distribution of Y_i, while leaving the mean of Y_i unchanged. Relative to Rothschild and Stiglitz's (1970) definition of a "mean-preserving spread", the elementary transformations defined here are both a generalization, in that they are defined for multidimensional distributions, and a specialization, in that, for the single dimension they affect, they affect values at only three *adjacent* points in the lattice.[15] As is easily checked, these elementary transformations entirely characterize componentwise convex functions, that is:

$$w \in \mathcal{X} \Leftrightarrow w \cdot t \geq 0 \quad \forall t \in \mathcal{T}(\mathcal{X}).$$

Proceeding as in Section 4, we can characterize the set of distributions ordered according to $\mathcal{X}$ as follows.

THEOREM 10 (COMPONENTWISE CONVEX ORDERING) *$f \prec_{\mathcal{X}} g$ if and only if there exist nonnegative coefficients α_t, $t \in \mathcal{T}(\mathcal{X})$, such that*

$$g = f + \sum_{t \in \mathcal{T}(\mathcal{X})} \alpha_t t.$$

The proof is analogous to the proof of Theorem 1 and therefore omitted.

For the supermodular ordering, we showed in Section 5 that the case of two dimensions is special in that, for any two distributions f, g with identical marginals, there is a unique decomposition of $g - f$ into a weighted sum of elementary transformations $t \in \mathcal{T}(\mathcal{S})$, where the weights α_t can have arbitrary signs. For the componentwise-convex ordering, the case of one dimension is special in an analogous sense. Specifically, if $n = 1$, for any two distributions f, g with identical means, there is a unique decomposition of $g - f$ into a weighted sum of elementary transformations $t \in \mathcal{T}(\mathcal{X})$, where the weights α_t can have arbitrary signs.[16]. Given this uniqueness, it follows from Theorem 10 that $f \prec_{\mathcal{X}} g$ if and only if the weight on every elementary transformation in the decomposition is nonnegative. To identify the weight on each elementary transformation in the unique decomposition, we adopt the notational conventions used in Section 5 and also note that for $L = \{0, 1, \ldots, l-1\}$, we can write $z+1$ instead of $z+e_i$. For any $z \in \{0, 1, \ldots, l-3\}$,

[15] If for some i the points in L_i are not equally spaced, the definition (20) can be generalized to $t_i^x(x) = 1$, $t_i^x(x+e_i) = -(\frac{|(x+2e_i)-(x)|}{|(x+2e_i)-(x+e_i|)}$, and $t_i^x(x+2e_i) = \frac{|(x+e_i)-(x)|}{|(x+2e_i)-(x+e_i)|}$. Fishburn and Lavalle (1995) have noted the convenience of working with supports that are evenly-spaced grids, but used summation by parts rather than defining elementary transformations. Müller and Scarsini's (2001) definition of a "mean-preserving local spread" is similar in motivation to our definition but in practice more complex to work with.

[16] For one-dimensional distributions f, g on $L = \{0, 1, \ldots, l-1\}$, with identical means, the difference vector δ is fully described by its values at $l-2$ points, and there are exactly $l-2$ (linearly independent) elementary transformations defined as in (20)

there are at most three elementary transformations $t \in \mathcal{T}(\mathcal{X})$ that take on non-zero values at z: t^z, $t^{(z-1)}$, and $t^{(z-2)}$. We can then write:

$$\begin{aligned} \delta(z) &= \alpha(z)t^z + \alpha(z-1)t^{(z-1)}(z) + \alpha(z-2)t^{(z-2)}(z) \\ &= \alpha(z) - 2\alpha(z-1) + \alpha(z-2), \end{aligned} \tag{21}$$

where the second line uses the definition of elementary transformations $t \in \mathcal{T}(\mathcal{X})$ in (20). Solving for the weights $\{\alpha(z)\}$ in terms of $\{\delta(z)\}$ yields $\alpha(z) = \sum_{i=0}^{z}(i+1)\delta(z-i)$. Thus, for one-dimensional distributions f, g with equal means,

$$f \prec_{\mathcal{X}} g \quad \Leftrightarrow \quad \sum_{i=0}^{z}(i+1)\,[g(z-i) - f(z-i)] \geq 0 \quad \forall z \in \{0, 1, \ldots, l-3\}. \tag{22}$$

The inequalities in (22) are the discrete analogs of Rothschild and Stiglitz's (1970) "integral conditions". They show that for one dimension, where the sets of convex and componentwise convex functions are identical, the extreme rays of the cone of componentwise convex functions are the functions $w(x) = \max\{z+1-x, 0\}$ for $z \in \{0, 1, \ldots, l-3\}$. Furthermore, in this special case of one dimension, there is a one-to-one mapping associating with each elementary transformation $t^z \in \mathcal{T}(\mathcal{X})$ the only extreme ray $w(x) = \max\{z+1-x, 0\}$ with which it makes a strictly positive scalar product.

For multidimensional distributions, determining whether g dominates f according to the componentwise convex ordering requires combining Theorem 10 with the analog of one of the constructive methods developed in Section 6 for the supermodular ordering.

Combined Properties of Objective Functions As mentioned earlier, one may be interested in classes of objective functions that satisfy both supermodularity and other properties. Such additional restrictions are important as they may refine the resulting order on distributions (from Theorem 8), i.e., allow one to compare distributions that were not comparable under the stochastic supermodular order. The following result, based on duality, provides a general method to characterize the order based on objective functions that combine several properties. Let $\mathcal{C}$ and $\mathcal{D}$ denote two classes of functions that are each stable under positive combinations (i.e., $\mathcal{C}$ and $\mathcal{D}$ are convex cones seen as subsets of $\mathbb{R}^d$). Also let $\mathcal{T}$ and $\mathcal{U}$ denote their respective sets of elementary transformations: In this generalized setting, elementary transformations are the extreme rays of the dual cones of $\mathcal{C}$ and $\mathcal{D}$.

THEOREM 11 (COMBINED CLASSES) *$f \prec_{\mathcal{C}\cap\mathcal{D}} g$ if and only if there exist nonnegative* *coefficients α_t and β_u such that*

$$g = f + \sum_{t \in \mathcal{T}} \alpha_t t + \sum_{u \in \mathcal{U}} \beta_u u.$$

Proof. The dual cone of the intersection of two polyhedral cones is equal to the (Minkowski) sum of the dual cones (see Goldman and Tucker, 1956). Therefore, $f \prec_{\mathcal{C} \cap \mathcal{D}} g$ if and only if $g - f$ belongs to $\mathcal{C}^* + \mathcal{D}^*$, where $\mathcal{C}^*$ and $\mathcal{D}^*$ are respectively the dual cones of $\mathcal{C}$ and $\mathcal{D}$. Since these dual cones are the convex hulls of $\mathcal{T}$ and $\mathcal{U}$, the result obtains. ■

Theorem 11 applies to any set of properties that can be described by polyhedral cones.

Corollary 2 *Let $\mathcal{SX}$ denote the set of objective functions that are both supermodular and componentwise convex. Then $f \prec_{\mathcal{SX}} g$ if and only if there exists a sequence of elementary transformations of either type t_i^x (defined in (20)) or type $t_{i,j}^x$ (defined in (3)) that, added to f, yield g.*

Müller and Scarsini (2010) have derived a similar characterization for the set of objective functions that are both submodular and componentwise concave, a set they term "inframodular". Like us, they employ duality methods, but rather than combining classes of functions, along with their associated elementary transformations, they define a single type of elementary transformation that corresponds to the class of inframodular functions and prove their equivalence result directly.

Convexity In multidimensional settings, discrete convexity is harder to characterize than discrete componentwise convexity. The very concept of convexity in discrete multidimensional settings has received a number of definitions, several of which are compared in Murota and Shioura (2001). We focus here on a notion, natural to economists, of convex-extensibility. A function $w : L \to \mathbb{R}$ is *convex extensible* if there exists a convex function $\bar{w} : \mathbb{R}^n \to \mathbb{R}$ such that $w(x) = \bar{w}(x)$ for all $x \in L$. Concavity is defined similarly. This definition is natural in economic settings: it characterizes usual convexity or concavity properties of an objective function defined on all possible outcomes in a situation where only discrete outcomes are available.[17] To apply the duality technique used so far in this section, we need to characterize convexity by a set of inequalities, each of which corresponds to an elementary transformation. For example, suppose that $L = \{0, 1, 2\}^2$. In this case, convexity is clearly a stronger requirement than componentwise convexity: the two diagonals of the square each imply a convexity relation that involves both dimensions. As a first guess, then, could it be that discrete convexity on L is characterized by the

[17] Although natural in economics, this definition of discrete convexity is criticized by Murota (1998).

componentwise convexity inequalities plus the inequalities $w(0,0) + w(2,2) \geq 2w(1,1)$ and $w(0,2) + w(2,0) \geq 2w(1,1)$? It turns out that this set of inequalities is not enough to guarantee convexity. For example, consider the function w on L with the following values:

w	$x_1 = 0$	$x_1 = 1$	$x_1 = 2$
$x_2 = 0$	0	1	2
$x_2 = 1$	1	1	1
$x_2 = 2$	2	1	2

The two inequalities above are satisfied, as are all those defining componentwise convexity. However, even though $(1,1)$ is the barycenter of $(0,0)$, $(1,2)$ and $(2,1)$ with equal weights, we have $w(1,1) > \frac{1}{3}(w(0,0)+w(1,2)+w(2,1))$, which precludes the existence of a convex function $\bar{w}$ extending w.

For real variables, the following relations are equivalent for any convex set $\mathcal{X}$ of $\mathbb{R}^n$ and $w : \mathcal{X} \to \mathbb{R}$:

$$w(\alpha x + (1-\alpha)y) \leq \alpha w(x) + (1-\alpha)w(y) \qquad \forall (x,y,\alpha) \in \mathcal{X}^2 \times [0,1]$$

$$w\left(\sum_{i=1}^{p} \alpha_i x_i\right) \leq \sum_{i=1}^{p} \alpha_i w(x_i) \quad \forall (x,\alpha) \in \mathcal{X}^p \times \Delta_{p-1}$$

However, this equivalence fails for discrete variables, as the above example illustrates. In that example, all convexity conditions involving convex combinations of two variables are satisfied, but convexity is violated by a convex combination of three variables. The reason is that the usual induction argument to reduce a p-variable convex relation to a 2-variable one fails, as the intermediate convex combinations it involves typically do not belong to the lattice.

How, then, can we characterize convex-extensibility? What convexity inequalities must a function w defined on an n-dimensional lattice satisfy in order to guarantee that it can be extended to a convex function of continuous variables? The answer is that one needs to consider only convex combinations of at most $(n+1)$ variables. The following characterization is new to our knowledge, although a similar statement based on epigraph comparisons for a slightly different class of functions appears in Kiselman (2005), and a method of proof using LP duality for local convex extensions is given in Murota (2003).

Theorem 12 (Discrete Convexity) *Let L denote any finite Cartesian lattice of $\mathbb{R}^n$. The following two statements are equivalent:*

- *(i) w is convex extensible.*

- *(ii) For all* $(x_0, \ldots, x_n) \in L$ *and* $\alpha \in \Delta_n$*,*

$$w\left(\sum_{i=0}^{n} \alpha_i x_i\right) \leq \sum_{i=0}^{n} \alpha_i w(x_i).$$

Proof. Clearly (i) implies (ii). We now show the reverse. For all $x \in \mathbb{R}^n$, Let

$$\bar{w}(x) = \sup_{(p,\gamma)\in\mathbb{R}^n\times\mathbb{R}} \{p \cdot x + \gamma | p \cdot y + \gamma \leq w(y) \quad \forall y \in L\}. \tag{23}$$

By construction, $\bar{w}$ is convex and such that $\bar{w}(x) \leq w(x)$ for all $x \in L$. We will show that $\bar{w}(x) \geq w(x)$ for all $x \in L$, which will conclude the proof. Since L is finite, the number d of constraints defining (23) is finite, and the objective is well defined and finite. By strong LP duality (see e.g. Bertsimas and Tsitsiklis, 1997, Theorem 4.4), this implies that for all $x \in \mathbb{R}^n$,

$$\bar{w}(x) = \inf_{\lambda\in\mathbb{R}^d} \left\{ \sum_{y\in L} \lambda_y w(y) | \sum_{y\in L} \lambda_y y = x, \sum_{y\in L} \lambda_y = 1, \lambda_y \geq 0 \right\}.$$

Moreover, there exists a basic feasible solution $\lambda^* \in \mathbb{R}^d$ to this dual program, i.e., such that λ^* vanishes except for a set $Y(x)$ of at most $n+1$ components (see Bertsimas and Tsitsiklis, Theorem 2.4). That is,

$$\bar{w}(x) = \sum_{y\in Y(x)} \lambda^*_y w(y).$$

From (ii), this implies that $\bar{w}(x) \geq w(x)$, which concludes the proof.[18] ■

Theorem 12 allows us to characterize the convex order in terms of a set of elementary transformations. For each subset $\chi = \{x_0, \ldots, x_n\} \subset L$ of $n+1$ elements and weights $\alpha \in \Delta_n$ such that $y = \sum_{i=0}^{n} \alpha_i x_i \in L \setminus \chi$, let $t(\chi, \alpha)$ denote the function on L such that $t(x_i) = \alpha_i$ for $0 \leq i \leq n$, $t(y) = -1$, and $t(x) = 0$ for $x \in L \setminus (\chi \cup \{y\})$, and let $\mathcal{T}_x$ denote the set of all such transformations. Let $\mathcal{C}_x$ denote the set of convex-extensible functions on L. Proceeding as for Theorem 1 and using Theorem 12, we get the following result:

THEOREM 13 (CONVEX ORDERING) *$f \prec_{\mathcal{C}_x} g$ if and only if there exist nonnegative coefficients $\{\alpha_t\}$, $t \in \mathcal{T}_x$, such that*

$$g = f + \sum_{t\in\mathcal{T}^x} \alpha_t t.$$

[18] The result can also be proved by adapting the approach of Kiselman (2005), by showing that the epigraph of w in $\mathbb{Z}^n \times \mathbb{R}$ is $\mathbb{Z}^n \times \mathbb{R}$ convex. With this approach, Carathéodory's theorem is used to reduce the number of convex combinations entering the characterization.

Theorem 3 showed that for the set of elementary transformations defined in equation (3) corresponding to the supermodular ordering, none of the transformations is redundant. An analogous result does not hold for the convex order. For example, consider for $L = \{0,1,2\}^2$ the 3-point convex combination where $(0,0)$ and $(2,0)$ receive weight $1/4$ and $(1,2)$ receives weight $1/2$. The resulting barycenter is $(1,1)$. In this case, however, the convex combination can be decomposed into two simpler ones, one putting weights $1/2$ on $(1,2)$ and $(1,0)$, and the other putting weights $1/2$ on $(0,0)$ and $(2,0)$. In terms of elementary transformations, we have

$$t(\{(0,0),(2,0),(1,2)\},(.25,.25,.5)) = t(\{(1,2),(1,0)\},(.5,.5))+\frac{1}{2}t(\{(0,0),(2,0)\},(.5,.5)).$$

Therefore, some "elementary transformations" in $\mathcal{T}_x$ are redundant.

For the class of supermodular and convex objective functions, Theorem 11 implies that $f \prec_{\mathcal{S}\cap\mathcal{C}_x} g$ if and only if g can be obtained by adding to f a positive sum of elementary transformations from $\mathcal{T}(\mathcal{S})$ and $\mathcal{T}_x$. In this case, redundancy is even more severe. In fact, preliminary investigation suggests, for the case of two dimensions, that one can dispense with all elementary transformations based on 3-point convex combinations.

10 Relation to Copulas

An increasingly popular way to think about interdependence across random variables is the concept of copulas. A common view is that copulas capture interdependence by separating marginal distributions from joint distributions. This view is based on Sklar's seminal theorem, which we recall here. For simplicity let us say that C is a *copula* if it is the joint distribution of n uniform random variables.

THEOREM 14 (SKLAR, 1959) *Let F be any distribution function of n variables, with marginals $F_1,\ldots,F_n$. There exists a copula C such that*

$$F(x_1,\ldots,x_n) = C(F_1(x_1),\ldots,F_n(x_n)).$$

Suppose that copulas indeed contain all interesting information about interdependence. There still remains the task of comparing copulas for different distributions. If the comparison of two joint distributions depends only on their copulas, how should one compare these copulas? A natural idea, followed by Decancq (2007) is to apply the stochastic supermodular ordering to copulas rather than to the distributions themselves.

Our analysis challenges the use of copulas for comparing interdependence. Firstly, recall that Theorem 1 implies that for two distributions to be comparable according to the supermodular ordering, they *must have identical marginals.* Therefore, the apparent gain provided by copulas to abstract from differing marginal distributions disappears when interdependence comparisons are based on the supermodular ordering.

Secondly, the use of copulas can only increase the complexity of the comparison. With discrete support, there is an uncountable infinity of copula representations for any distribution F. The only constraint (other than the usual conditions for any function to be a copula) is that copulas must coincide on the range of values of the marginal distributions. This point can be illustrated by the simplest example: suppose that $L = L_2$, i.e. L consists of a one-dimensional two-point support, and that $F(0) = 1/4$. Then, any nondecreasing function $C : [0,1] \to [0,1]$ such that $C(1/4) = 1/4$, $C(0) = 0$ and $C(1) = 1$ provides a representation of F in Sklar's theorem. It is generally impossible to reconstruct a distribution from its copula. To illustrate, suppose in the previous example that $C(x) = 0$ for $x < 1/4$, $C(x) = 1/4$ for $1/4 \leq x < 1/2$, $C(x) = 1/2$ for $1/2 < x < 1$ and $C(1) = 1$. One could mistakenly infer that there are three points in the support of F, since the copula has three jumps. Or, if one already knows the initial distribution has a two-point support, how to determine which value of $1/4$ or $1/2$ corresponds to $F(0)$? One could impose the rule of picking a particular copula that is constant between any two values in the range of F, but then the copula coincides with F, except that the domain is scaled by the values of marginal distributions. Therefore, even with this rule, copulas do not offer any advantage compared to working with the initial distributions. In conclusion, the use of copulas should be rejected because i) distributions can only be compared if they have identical marginals, so that advantage of copulas disappears, and ii) copulas are only well defined on the range of values of marginal distributions, and contain no other useful information. To compare copulas according to the stochastic supermodular ordering, one has to essentially reconstruct the initial joint distribution.

11 Conclusion

12 References

Acemoglu, D., Dahleh, M., Lobel, I., and Ozdaglar, A.E., "Bayesian Learning in Social Networks", NBER Working Paper No. W14040, May 2008.

Acemoglu, D., Ozdaglar, A.E., and ParandehGheibi, A., "Spread of (Mis)Information in Social Networks", M.I.T. Economics W.P. No. 09-15, May 2009.

Adler, M. D., and Sanchirico, C. W., "Inequality and Uncertainty: Theory and Legal Applications", University of Pennsylvania Law Review, 2006, 155, 279.

Athey, S., "Characterizing Properties of Stochastic Objective Functions", unpublished paper, Sept. 2000.

Atkinson, A. B., and Bourguignon, F., "The Comparison of Multi-Dimensioned Distributions of Economic Status", Review of Economic Studies, 1982, 49, 183-201.

Avis, D., and Fukuda, K., "A Pivoting Algorithm for Convex Hulls and Vertex Enumeration of Arrangements and Polyhedra", Discrete and Computational Geometry, 1992, 8, 295-313.

Avis, D., and Bremner, D., "How Good Are Convex Hull Algorithms?", 11^{th} Annual ACM Symposium on Computational Geometry, Vancouver, B.C., 1995, pp. 20–28.

Ben-Porath, E., Gilboa, I., and Schmeidler, D., "On the Measurement of Inequality under Uncertainty", Journal of Economic Theory, 1997, 75, 194-204.

Bertsimas, D., and Tsitsiklis, J. *Introduction to Linear Optimization*, Athena Scientific, Nashua, N.H., 1997.

Boland, P. J., and Proschan, F., "Multivariate Arrangement Increasing Functions with Applications in Probability and Statistics", Journal of Multivariate Analysis, 1988, 25, 286-298.

Bond, P., and Gomes, A., "Multitask Principal-Agent Problems: Optimal Contracts, Fragility, and Effort Misallocation", Journal of Economic Theory, 2009, 144, 175-211.

Bourguignon, F., and Chakravarty, S.R., "Multi-dimensional poverty orderings", DELTA W.P. 2002-22.

de Castro, L.I., "Affiliation and Dependence in Economic Models", mimeo, Oct. 2009.

Chew, S.H., and Sagi, J., "An Inequality Measure for Uncertain Allocations", mimeo, 2010.

Christofides, T.C., and Vaggelatou, E., "A Connection between Supermodular Ordering and Positive/Negative Association", Journal of Multivariate Analysis, 2004, 88, 138-151.

Choi, S., Gale, D., and Kariv, S., "Behavioral Aspects of Learning in Social Networks: An Experimental Study", Advances in Applied Microeconomics, 2005, 13.

Choi, S., Gale, D., and Kariv, S., "Social Learning in Networks: A Quantal Response Equilibrium Analysis of Experimental Data", W.P., May 2009.

Cousin, A., and Laurent, J.-P., "Comparison Results for Exchangeable Credit Risk Portfolios", Insurance: Mathematics and Economics, 2008, 42, 1118-1127.

Decancq, K., "Copula-Based Measurement of Dependence between Dimensions of Well-Being", HEDG Working Paper 09/32.

Denuit, M., Dhaene, J., Goovaerts, M.J., and Kaas, R., Actuarial Theory for Dependent Risks: Measures, Orders, and Models, 2005, Wiley.

Dhaene, J., and Goovaerts, M.J., "Dependency of Risks and Stop-Loss Order", ASTIN Bulletin, 1996, 26, 201-212.

Epstein. L., and Tanny, S., "Increasing Generalized Correlation: A Definition and Some Economic Consequences", Canadian Journal of Economics, 1980, 13, 16-34.

Esary, J.D., Proschan, F., and Walkup, D.W., "Association of Random Variables, with Applications", Annals of Mathematical Statistics, 1967, 44, 1466-1474.

Fernandez, R., and Gali, J., "To Each According to? Markets, Tournaments, and the Matching Problem with Borrowing Constraints", Review of Economic Studies, 1999, 66, 799-824.

Fishburn, P. C., and Lavalle, I. H., "Stochastic Dominance on Unidimensional Grids", Mathematics of Operations Research, 1995, 20, 513-525.

Foerster, A.T., Sarte, P.-D. G., and Watson, M.W., "Sectoral vs. Aggregate Shocks: A Structural Factor Analysis of Industrial Production", NBER W.P. 14389, Oct. 2008.

Frostig, E., "Comparison of Portfolios which Depend on Multivariate Bernouilli Random Variables with Fixed Marginals", Insurance: Mathematics and Economics, 2001, 29, 319-331.

Fukuda, K., "Frequently Asked Questions in Polyhedral Computation", mimeo, Swiss Federal Institute of Technology, 2004.

Fukuda, K., and Prodon, A., "Double description method revisited." In Combinatorics and Computer Science, volume 1120 of Lecture Notes in Computer Science, pages 91–111. Springer, 1996.

Gajdos, T., and Maurin, E., "Unequal Uncertainties and Uncertain Inequalities: An Axiomatic Approach", Journal of Economic Theory, 2004, 116, 93-118.

Goldman, A. J. and Tucker, A. W. "Polyhedral convex cones." In Linear inequalities and related systems, Annals of Mathematics Studies 38, eds. H. W. Kuhn and A. W. Tucker. Princeton University Press, 1956, pp. 19-39.

Golub, B., and Jackson, M.O., "Naive Learning in Social Networks: Convergence, Influence and the Wisdom of Crowds", American Economic Journal: Microeconomics, 2010, 2, 112-149.

Hardy, G.H., Littlewood, J.E., and Polya, G., *Inequalities*, Cambridge: Cambridge University Press, 2nd edition, 1952.

Hennessy, D. A., and Lapan, H. E., "A Definition of 'More Systematic Risk' with Some Welfare Implications", Economica, 2003, 70, 493-507.

Hu, T., Müller, A., and Scarsini, M., "Some Counterexamples in Positive Dependence", Journal of Statistical Planning and Inference, 2004, 153-158.

Hu, T., and Wu, Z., "On Dependence of Risks and Stop-Loss Premiums", Insurance: Mathematics and Economics, 1999, 24, 323-332.

Hu, T., Xie, C., and Ruan, L, "Dependence Structures of Multivariate Bernoulli Random Vectors", Journal of Multivariate Analysis, 2005, 94, 172-195.

Hu, T., and Yang, J., “Further developments on sufficient conditions for negative dependence of random variables”, Statistics and Probability Letters, 2004, 66, 369-381.

Joe, H., “Multivariate Concordance”, Journal of Multivariate Analysis, 1990, 35, 12-30.

Jogdeo, K., “On a Probability Bound of Marshall and Olkin”, Annals of Statistics, 1978, 232-234.

Karlin, S., and Novikoff, A., “Generalized Convex Inequalities”, Pacific Journal of Mathematics, 1963, 13, 1251-1279.

Kiselman, C., “Convex Functions on Discrete Sets”, Combinatorial Image Analysis, 2005, Springer.

Kroll, Y., and Davidovitz, L., “Inequality Aversion versus Risk Aversion”, Economica, 2003, 70, 19-29.

Levy, H., and Paroush, J., “Toward Multivariate Efficiency Criteria”, Journal of Economic Theory, 1974, 7, 129-42.

Lorentz, G.G., “An Inequality for Rearrangements”, American Mathematical Monthly, 1953, vol. 60, 176-179.

Luenberger, D., *Vector Space Optimization*, Wiley, New York, 1969.

Lu, T.-Y., and Zhang, Y., “Generalized Correlation Order and Stop-Loss Order”, Insurance: Mathematics and Economics, 2004, 35, 69-76.

Marinacci, M., and Montrucchio, L., “Ultramodular Functions”, Mathematics of Operations Research, 2005, 30, 311-332.

McAfee, R. P., “Coarse Matching”, Econometrica, 2002, 70, 2025-34.

Meyer, M.A., “Interdependence in Multivariate Distributions: Stochastic Dominance Theorems and an Application to the Measurement of Ex Post Inequality under Uncertainty”, Nuffield College Disc. Paper No. 49, April 1990.

Meyer, M. A., and Mookherjee, D., “Incentives, Compensation and Social Welfare”, Review of Economic Studies, 1987, 45, 209-226.

Meyer, M. A., and Rothschild, M., “(In)efficiency in Sorting and Matching”, mimeo, Oxford, 2004.

Meyer, M.A., and Strulovici, B., “Increasing Interdependence of Multivariate Distributions”, 2010, available at http://faculty.wcas.northwestern.edu/ bhs675/IIMD1.pdf.

Motzkin, T. S., Raiffa, H., Thompson, G. L. , and Thrall, R. M.. “The Double Description Method”. In H. W. Kuhn and A. W. Tucker Ed., Contributions to the Theory of Games II, 1953, 51-73.

Müller, A., and Scarsini, M., “Some Remarks on the Supermodular Order”, Journal of Multivariate Analysis, 2000, 73, 107-119.

Müller, A., and Scarsini, M., “Stochastic Comparison of Random Vectors with a Common Copula”, Mathematics of Operations Research, 2001, 26, 723-740.

Müller, A., and Scarsini, M., “Fear of Loss, Inframodularity, and Transfers”, mimeo, March, 2010.

Müller, A., and Stoyan, D., Comparison Methods for Stochastic Models and Risks, 2002, Wiley.

Murota, K., “Discrete Convex Analysis”, Mathematical Programming, 1998, 83, 313-371.

Murota, K., *Discrete Convex Analysis”*, SIAM Monographs on Discrete Mathematics and Applications, Philadelphia, PA, 2003.

Patton, A., “Are ‘Market Neutral’ Hedge Funds Really Market Neutral?”, Review of Financial Studies, 2009, 22, 2495-2530.

Prat, A., “Should a Team Be Homogenous?”, European Economic Review, 2002, 46, 1187-1207.

Rothschild, M. and Stiglitz, J. E. “Increasing risk: I. A definition”, Journal of Economic Theory, 1970, 2, 225-243.

Rüschendorf, L., “Comparison of Multivariate Risks and Positive Dependence”, Journal of Applied Probability, 41, 2004, 391-406.

Shaked, M., “A Concept of Positive Dependence for Exchangeable Random Variables”, Annals of Statistics, 1977, 5, 505-515.

Shaked, M., and Shanthikumar, J.G., "Supermodular Stochastic Orders and Positive Dependence of Random Vectors", Journal of Multivariate Analysis, 1997, 61, 86-101.

Shaked, M., and Tong, Y.L., "Some Partial Orderings of Exchangeable Random Variables by Positive Dependence", Journal of Multivariate Analysis, 1985, 17, 333-349.

Shimer, R., and Smith, L., "Assortative Matching and Search", Econometrica, 2000, 68, 343-369.

Sklar, A. "Fonctions de répartition à n dimensions et leurs marges", Publ. Inst. Statist. Univ. Paris, 1959, 8, 229-231.

Tchen, A.H., "Inequalities for Distributions with Given Marginals", Annals of Probability, 1980, 8, 814-827.

Topkis, D.M., *Ordered Optimal Solution*, Doctoral Dissertation, Stanford University, 1968.

Topkis, D.M., "Minimizing a Submodular function on a Lattice", Operations Research, 1978, 26, 305-331.

Yi, Z., and Weng, C., "On the Correlation Order", Statistics and Probability Letters, 2006, 76, 1410-1416.

Appendices

A Appendix: Implementation of the Double Description Method

B Appendix: Proof of Proposition 2

C Appendix: Proof of Theorem 6 and Its Corollary

The proof of Theorem 6 is based on the following two lemmas.

Lemma 3 *Suppose that $X \prec_{SSPM} Y$ are two-dimensional and that Z is a p-dimensional (p arbitrary) random vector independent of X and Y. Then $(X, Z) \prec_{SSPM} (Y, Z)$.*

Proof. We need to check that $Ew(X, Z) \leq Ew(Y, Z)$ for all w symmetric and supermodular. For each z in $\mathbb{R}^p$, let $r(z) = Ew(X, z)$ and $s(z) = Ew(Y, z)$. For each z, the function $w(\cdot, Z)$ is symmetric and supermodular in its two arguments, and so $X \prec_{SSPM} Y$ implies that $r(z) \leq s(z)$ for all z. Taking expectations with respect to Z (and using independence of Z) then shows the result. ■

Let X and Y be two-dimensional random vectors generated by $2 \times m$-matrices A and B, respectively. Suppose that

$$B = A + \sum_{k=2}^{m} \varepsilon_k E_k,$$

where $\varepsilon_k \geq 0$ and E_k is the matrix with zeros everywhere except for columns $k-1$ and k, where it is defined by

$$(E_k)_{1,k-1} = (E_k)_{2,k} = -1$$

and

$$(E_k)_{1,k} = (E_k)_{2,k-1} = 1$$

Intuitively, B is putting, for each pair of consecutive prizes, less probability on the second individual (row) getting the lower of the two prizes and more weight on him getting the better one. Given this, one would expect that B is more equal than A if A was treating individual one (first row) better than the second one.

This intuition is captured by the lemma to follow. With two dimensions the symmetric supermodular ordering is characterized by the following symmetric supermodular functions:

$$w^k(X) = 1_{X_1 \geq c_k, X_2 \geq c_k}$$

for each $k \geq 2$ and, for $k \neq l$ greater than 2,

$$w^{kl}(X) = 1_{X_1 \geq c_k, X_2 \geq c_l} + 1_{X_1 \geq c_l, X_2 \geq c_k},$$

where $c_1 < c_2 < \cdots < c_m$ is an arbitrary vector of indices decreasing with prize values (so that the first prize has the lowest index, etc.). The reason why indices are greater than 2 is that for $k = 1$ the indicator-based conditions above are always satisfied, since all prizes have indices above c_1. For each k, $Ew^k(X) \leq Ew^k(Y)$ is equivalent to

$$0 \leq \left(\sum_{j=k}^{m} \beta_{1j}\right)\left(\sum_{j=k}^{m} \beta_{2j}\right) - \left(\sum_{j=k}^{m} \alpha_{1j}\right)\left(\sum_{j=k}^{m} \alpha_{2j}\right), \tag{24}$$

where α's and β's are the entries of matrices A and B, respectively. Similarly, for each $k \neq l$ greater than 2, $E^{kl}(X) \leq Ew^{kl}(Y)$ is equivalent to, using the more compact notation of cumulative matrices $\bar{A}$ and $\bar{B}$ with entries $\bar{\alpha}$ and $\bar{\beta}$,

$$0 \leq \bar{\beta}_{1k}\bar{\beta}_{2l} - \bar{\alpha}_{1k}\bar{\alpha}_{2l} + \bar{\beta}_{1l}\bar{\beta}_{2k} - \bar{\alpha}_{1l}\bar{\alpha}_{2k}. \tag{25}$$

LEMMA 4 *Suppose that for each $k \in \{2, \ldots m\}$,*

$$\sum_{j=k}^{m} \alpha_{2j} \geq \sum_{j=k}^{m} \alpha_{1j} + \varepsilon_k.$$

Then, α and β satisfy (24) *for each k, and* (25) *for each $k \neq l$.*

Proof. Since all ε_j's simplify in the above β sums except for ε_k, Condition (24) becomes, after simplification,

$$\varepsilon_k \left[\sum_{j=k}^{m} \alpha_{2j} - \left(\sum_{j=k}^{m} \alpha_{1j} + \varepsilon_k\right)\right],$$

which is nonnegative by assumption. For each $k \neq l$ greater than 2, Condition (25) is proved as follows. Since by construction $\bar{\beta}_{1k} = \bar{\alpha}_{1k} + \varepsilon_k$ and $\bar{\beta}_{2k} = \bar{\alpha}_{2k} - \varepsilon_k$ for all $k \geq 2$, therefore the condition simplifies to

$$0 \leq \varepsilon_k \left[\bar{\alpha}_{2l} - (\bar{\alpha}_{1l} + \varepsilon_l)\right] + \varepsilon_l \left[\bar{\alpha}_{2k} - (\bar{\alpha}_{1k} + \varepsilon_k)\right],$$

both terms of which are nonnegative by assumption. ■

When α and β represent probability distributions, the conclusion of Lemma 4 is that $X \prec_{SSPM} Y$.

COROLLARY 3 *Suppose that α and β consist of probability vectors satisfying the assumption of Lemma 4. Then,*

$$X \prec_{SSPM} Y.$$

The reason for stating Lemma 4 and its corollary separately is that we wish to apply Lemma 4to intermediary transformations of matrices A and B whose rows do not necessarily represent probability distributions, as will be clear from the final proof of this section. The corollary simply states how the conclusion of the Lemma should be interpreted in our context, when A and B consist of probability distributions. The condition in Lemma 4 implies that the one-dimensional distribution generated by the second row of A assigns lower prizes (in the fist-order stochastic dominance sense) than the one generated by the first row of A, and is strictly stronger than that, since the FOSD inequalities must hold by more than ε_k for each k.

We can now conclude the proof of Theorem 6. We first show that $\bar{A} \prec_{SSPM} \bar{B}^{so}$ and then that $\bar{B}^{so} \prec_{SSPM} \bar{B}$, where $\bar{B}^{so}$ is the matrix obtained from $\bar{B}$ by reordering each of its column from the smallest to the greatest element. This will then prove the result, by transitivity. Notice that $\bar{B}^{so}$ is essentially a stochastic reordering of the matrix B so as to systematically put more probability of lower prizes to high index individuals. With this interpretation, it is not surprising that $\bar{B}^{so} \prec_{SSPM} \bar{B}$. Since A is already assumed to be stochastically ordered the comparison assumed on A and B carries over to a comparison between A and B^{so}, an so it is not surprising either that $A \prec_{SSPM} B^{so}$.

Proof that $\bar{A} \prec_{SSPM} \bar{B}^{so}$. We use the following algorithm: We start by transforming the last column of $\bar{A}$ into the last column of $\bar{B}$ by applying to $\bar{A}$ a sequence of elementary transformations $\varepsilon_m E_m$ of the type described in Lemma 4, only involving the last column of $\bar{A}$ and only one pair of rows at each time, and such that, after each step, the resulting matrix is still stochastically ordered.[19] Such a construction is given by Hardy et al. (1952). At each step, the last column of the resulting matrix is stochastically ordered,

[19]In terms of A, these transformations involve only the last two columns of A. Note that E_m's have no impact on cumulative sums for $k < m$ so they only affect $\bar{A}$ through its last column. For convenience, we state the result in terms of the cumulative matrix $\bar{A}$.

and remaining columns are untouched, so Lemma 4 can be applied. Lemma 4 combined with Lemma 3 ensures that at each step the new matrix SSPM dominates the previous and, by transitivity, $\bar{A}$. Once the last column of $\bar{A}$ has been transformed into that $\bar{B}^{so}$, one proceed to do the same for the second to last column of $\bar{A}$, etc. Once the second column has been transformed, the resulting matrix is $\bar{B}^{so}$ itself, which shows by transitivity, that $\bar{A} \prec_{SSPM} \bar{B}^{so}$.

Proof that $\bar{B}^{so} \prec_{SSPM} \bar{B}$. Columns of $\bar{B}^{so}$ and $\bar{B}$ have the same entries, only in a different order, since $\bar{B}^{so}$'s entries are increasing with the row index, for fixed columns. Without loss of generality, reset the entries in each column of $\bar{B}^{so}$ as $1, 2, \ldots, n$, with the same correspondence for $\bar{B}$. The goal is to find an algorithm that rearranges these entries to match $\bar{B}$'s. Resetting entries is for convenience only in order to emphasize the workings of the algorithm. In practice, the elementary transformations used will match actual entries of $\bar{B}^{so}$. Starting from the last row, n, of $\bar{B}^{so}$, whose entries are equal to n after relabeling, we will move these 'n'-labeled entries upwards, gradually, so as to position them as in $\bar{B}$. We will do this by a sequence of entry permutations between rows n and i for i starting from $n-1$ until i reaches 1. We will do this so that, at each step i, the rows above n remain stochastically ordered, and the n^{th} row remains stochastically higher than rows above i. This guarantees that applications of Lemma 4, at each step, is valid and so that the transformed matrix always SSPM dominates the previous one and, by transitivity, $\bar{B}^{so}$. Thus, starting with rows n and $n-1$, flip entries of $\bar{B}^{so}$ for each column j in which $\bar{B}_{nj} \neq n$. The result is that some entries of in the last row of $\bar{B}^{so}$ are now equal to $n-1$, while entries in its $(n-1)$ row are equal to n, for exactly those columns where $\bar{B}_{nj} \neq n$. The result is that now the n and $n-1$ rows of $\bar{B}^{so}$ are no longer stochastically ordered, but both rows still dominate all rows with indices less than $n-2$. The next step is to flip entries between the n and $n-2$ rows of the resulting matrix, for columns where its n^{th}-row entry does not match that of $\bar{B}$. As a result, the n^{th} row now contains (possibly) entries labeled '$n-2$' while the $n-2$ two contains $n-1$ entries. Notice that, i) the n, $n-1$, and $n-2$ rows still dominate all rows with indices less than $n-3$, and ii) the $n-1$ row dominates the $n-2$ row. The reason for the last point is that the $n-2$ row inherited an $n-1$ only if the $n-1$ row inherited an n entry. Proceeding systematically by decreasing the row index each time, the result is that the n^{th} row now has the same entries as $\bar{B}$'s, and that the first $n-1$ rows of the resulting matrix are still stochastically ordered. Applying next to the $n-1$ row what was done to the n row, we can transform it into the $n-1$ row of $\bar{B}$ while preserving at each step the stochastic ordering of the first $n-2$ rows and guaranteeing that the $n-1$ row dominates rows with which it has not yet been flipped. Applying this larger algorithmic loop to each row, in decreasing index order, eventually transforms $\bar{B}^{so}$ into $\bar{B}$ through a sequence of steps that increase in the SSPM sense, which proves the result. ■

Proof of the Corollary to Theorem 6

The matrix A generating (among all row-stochastic matrices with matching column sums) the worst distribution with respect to SSPM dominance is constructed as follows. For any real number x, let $\lfloor x \rfloor$ denote the largest integer below x. Set $a_{i,1} = 1$ for all $i \le i_1 = \lfloor np_1 \rfloor$, $a_{i_1+1,1} = np_1 - i_1$, and $a_{i,1} = 0$ for all $i > \lfloor i_1 + 1 \rfloor$. This assignment maximizes the entries of the low-index rows of the first column, subject to A's row-stochasticity constraint and to the sum of entries in the first column being equal to np_1. Put differently, the first column of A, seen from top down, majorizes all vectors with entries less than one and summing to np_1. Remaining vectors are defined similarly: the second column vector is the vector that majorizes all vectors that respect A's row-stochasticity and the summing up to p_2. Precisely, set $a_{i,2} = 0$ for all $i \le i_1$ since these rows already have ones in the first column, $a_{i_1+1,2} = \min\{1 - a_{i_1+1,1}, np_2\}$. The first argument of the minimizer expresses the constraint that the row sum cannot exceed one, and the second argument that entries in the second column cannot exceed np_2. Finally, set $a_{i,2} = 1$ for all i's between $i_1 + 1$ and $i_2 = i_1 + \lfloor np_2 - a_{i_1+1,2} \rfloor$, and $a_{i_2+1,2} = np_2 - a_{i_1+1,2} - (i_2 - i_1)$. Thus, after completing the $i_1 + 1$ row with whatever probability remains after setting $a_{i_1+1,1}$, one sets entries below equal to 1 subject to the column sum being less than np_2, and put whatever fraction remains in the next entry below. Remaining columns are constructed similarly.

By construction, A is stochastically ordered, as is easily checked. Moreover, given any row-stochastic matrix B with the same column sums as A, it is intuitive and easy to check that $A \succ_{CCM} B$ since A puts as much weight as possible in the first columns of the first rows and, equivalently, in the last columns of the last row. Precisely, for any column k and row l, the sum of entries in A over all columns with index above k and rows with index above l is maximal, subject to row-stochasticity and column-sum constraints.

D Appendix: Proof of Theorem 7

Let $\bar{A}$ and $\bar{B}$ denote cumulative-probability matrices (i.e., entries are increasing with the column index and less than one) of equal dimensions. By assumption $\bar{A}$ stochastically ordered means that $\bar{a}_{ik}$ is increasing in i. Finally, suppose that, for each k, the column vector $\bar{A}_k$ majorizes the column vector $\bar{B}_k$.

We wish to show that $X \succ_{SPM} Y$ or, abusing notation, $\bar{A} \succ_{SPM} \bar{B}$, where $\bar{A}$ and $\bar{B}$ are the cumulative matrices generating X and Y respectively. We break the proof in several steps.

D.1 $\bar{B}$ stochastically ordered

Assume for now that $\bar{B}$ is also stochastically ordered, so that $\bar{b}_{ik}$ is increasing in i.

Further suppose for now that $\bar{B}$ has strictly monotonic entries across row and column indices and let

$$\chi = \min_{i,j}\{\bar{b}_{i+1,j} - \bar{b}_{i,j}, \bar{b}_{i,j} - \bar{b}_{i,j+1}\} > 0.$$

Here and throughout, we exclude the first column of ones that could conventionally appear in cumulative probability matrices. We will simply say that $\bar{B}$ is "strictly monotonic."

Let k denote the largest column index such that $\bar{A}_k \neq \bar{B}_k$.

Lemma 5 *There exists a cumulative-probability matrix C that is stochastically ordered, such that $C_{\tilde{k}} = \bar{B}_{\tilde{k}}$ for all $\tilde{k} \geq k$, whose columns majorize $\bar{B}$'s, and such that $\bar{A} \succ_{SPM} C$.*

Proof.

We proceed by contradiction. Let C solve the optimization problem

$$\inf_E \sum_{i\geq 2} \sum_{j \geq i} e_{j,k} - \bar{b}_{j,k} \tag{26}$$

subject to the following constraints:

1. E satisfies row monotonicity (i.e., entries of E are decreasing in the column index),
2. E is stochastically ordered (i.e., entries of E are increasing in the row index),
3. E dominates $\bar{B}$ according to the cumulative column criterion (i.e., each column of E's majorizes the corresponding column of $\bar{B}$),
4. $\bar{A}$ dominates E according to the stochastic supermodular ordering

5. $E_{\tilde{k}} = \bar{B}_{\tilde{k}}$ for all $\tilde{k} < k$.

The set of E's satisfying these five constraints is compact (as a closed, bounded subset of a finite dimensional space) and nonempty (since $\bar{A}$ belongs to it), and the objective (26) is continuous. Therefore, its minimum is reached by some C. Moreover, the objective is nonnegative for all E since E_k majorizes $\bar{B}_k$. Finally, the minimum is equal to zero if and only if C_k is equal to $\bar{B}_k$. We now show that this last property indeed holds, which will prove the lemma.

Suppose by contradiction that $C_k \neq \bar{B}_k$.

Since C_k majorizes $\bar{B}_k$ and $C_k \neq \bar{B}_k$, there exists a row i such that

- $c_{i,k} \leq \bar{b}_{i,k}$
- $c_{i+1,k} \geq \bar{b}_{i+1,k}$
- One of the previous two inequalities is strict.

We will show that it is possible to increase $c_{i,k}$ by some small amount ε, and decrease $c_{i+1,k}$ by the same amount, while satisfying all 5 constraints of the minimization problem (26). Such change only affects the $i+1$ partial sum of (26), and decreases it by an amount ε, which contradicts the assumption that C minimizes (26).

First, we observe that

- $c_{i,k} \leq c_{i+1,k} - \chi$, from the previous two inequalities.
- $c_{i,k} \leq c_{i,k-1} - \chi$, since $c_{i,k} \leq \bar{b}_{i,k} \leq \bar{b}_{i,k-1} - \chi = c_{i,k-1} - \chi$.
- $c_{j,\tilde{k}} = \bar{b}_{j,\tilde{k}}$ for all $\tilde{k} < k$ and all j.

Therefore, it is possible to strictly increase $c_{i,k}$, up to an amount χ, without violating row-monotonicity of row i.

Let $\bar{k}$ denote the largest column index such that $c_{i+1,\tilde{k}} = c_{i+1,k}$ for all $\tilde{k} \in [k, \bar{k}]$. Possibly, $\bar{k} = p$, where p is number of columns of the matrices $\bar{A}$ and $\bar{B}$.

If $\bar{k} = k$, it means that one can decrease $c_{i+1,k}$ without violating row-monotonicity of row $i+1$.

Now consider the harder case where $\bar{k} > k$.

Define the matrix D that is identical to C for all rows other than i and $i+1$ and for all columns outside of $[k, \bar{k}]$, and such that

- $d_{i,\tilde{k}} = c_{i,\tilde{k}} + \varepsilon$
- $d_{i+1,\tilde{k}} = c_{i+1,\tilde{k}} - \varepsilon = c_{i+1,k} - \varepsilon$

for all $\tilde{k} \in [k, \bar{k}]$ where ε is some positive constant that we will determine later.

By construction, if ε is small enough, this transformation respects row-monotonicity for rows i and $i+1$: For i, this comes from the earlier observation that $c_{i,k} \leq c_{i,k+1} - \chi$. For $i+1$, this comes from the definition of $\bar{k}$.[20]

Furthermore, we now check that D is still stochastically ordered, provided that ε is such that $c_{i,k} + \varepsilon \leq c_{i+1,k} - \varepsilon$, which holds for all $\varepsilon \leq \chi/2$. This is clearly true for all columns outside of $[k, \bar{k}]$, where D is identical to C. For any column $\tilde{k} \in [k, \bar{k}]$, notice that

$$d_{i,\tilde{k}} \leq d_{i,k} = c_{i,k} + \varepsilon \leq c_{i+1,k} - \varepsilon = d_{i+1,\tilde{k}},$$

which shows the result.

Finally, the columns of D still majorize those of B. For this, we only need to check that

$$\sum_{j \geq i+1} d_{j,\tilde{k}} \geq \sum_{j \geq i+1} \bar{b}_{j,\tilde{k}} \tag{27}$$

for all $\tilde{k} \in [k, \bar{k}]$. All other majorization inequalities hold trivially since D has the same relevant partial sums as C for columns outside of $[k, \bar{k}]$ and for row indices other than $i+1$. By construction, we have

$$\sum_{j \geq i+2} d_{j,\tilde{k}} = \sum_{j \geq i+2} c_{j,\tilde{k}} \geq \sum_{j \geq i+2} \bar{b}_{j,\tilde{k}} \tag{28}$$

Moreover,

$$d_{i+1,\tilde{k}} = c_{i+1,k} - \varepsilon \geq \bar{b}_{i+1,k} - \varepsilon \geq \bar{b}_{i+1,\tilde{k}}$$

where the last inequality holds for $\varepsilon \leq \chi$. Combining this with (28) implies (27).

[20] If $\bar{k} = p$, we note that, necessarily, $c_{i+1,k} \geq \bar{b}_{i,k} + \chi > 0$, so we can indeed decrease the entries of C's $i+1$-row by an amount ε without creating negative entries.

Such transformation is such that $C \succ_{SPM} D$, from Lemma 2. By transitivity, this implies that $\bar{A} \succ_{SPM} D$.

This contradicts the hypothesis that C was minimizing the objective, since the ε reduction has strictly improved the partial sum of (26) starting from row $i+1$ and left other partial sums unaffected. This proves the lemma. ■

To conclude the proof of Theorem 7, for $\bar{B}$ strictly monotonic, we apply the above lemma inductively on k.

If $\bar{B}$ is not strictly monotonic, we perturb $\bar{A}$ and $\bar{B}$ by taking a limit of cumulative matrices A_n, B_n with $\chi_n = 1/n$ converging to $\bar{A}, \bar{B}$ and such that A_n majorizes B_n, and apply the previous analysis to show that

$$A_n \succ_{SPM} B_n,$$

for each n. Taking the limit as n goes to infinity then shows the result.

To show that a perturbation of order $1/n$ is feasible, for n large enough, simply scale down entries of $\bar{A}$ and $\bar{B}$ by a factor $\delta_n = 1 - (p+q)/n$ (where $p \times q$ are matrix dimensions (recall that we have excluded the first column of ones that may appear in cumulative matrices) and add $e_{i,j} = \frac{1}{n}(i+j)$ to A and B. This ensures that A_n and B_n are strictly increasing with factor $1/n$ with entries less than 1. Moreover, A_n majorizes B_n.

Thus we have shown the result when $\bar{B}$ is stochastically ordered.

D.2 General Case

In general, $\bar{B}$ is not stochastically ordered. Let $\bar{B}^{so}$ denote the stochastically ordered version of $\bar{B}$, i.e., the matrix whose k^{th} column consists of the entries of the k^{th} column of $\bar{B}$, ordered from the smallest to the largest. It is easy to check that $\bar{B}^{so}$ is also row monotonic, i.e., that the entries of $\bar{B}^{so}$ of any given row are decreasing in the column index. Indeed, the i^{th} entry of $\bar{B}^{so}_k$ is the i^{th} smallest entry in the column $\bar{B}_k$. Since $\bar{B}$ is row monotonic by assumption, that entry must be larger than the i^{th} smallest entry in the column $\bar{B}_{k+1}$, which is the i^{th} entry of $\bar{B}^{so}_{k+1}$. Applying the previous analysis to $\bar{A}$ and $\bar{B}^{so}$, one concludes that $\bar{A} \succ_{SPM} \bar{B}^{so}$. Therefore, we will be done if we show that $\bar{B}^{so} \succ_{SPM} \bar{B}$.

To see this we exploit the algorithm used in the proof of Theorem 6 to convert $\bar{B}^{so}$

to $\bar{B}$ by the sequence of pairwise row transformations performed in Section C, and by applying Lemma 2 at each step to show that each transformation results in a lower matrix that is lower according to the supermodular order. The only difficulty is to ensure that each step preserves row monotonicity. Indeed, Lemma 2 only applies to rows that satisfy this constraint. Consider the first induction step of the conversion from $\bar{B}^{so}$ to $\bar{B}$, which consists of pairwise transformations between the n^{th} row of $\bar{B}^{so}$ and its i^{th} row for i decreasing from $n-1$ to 1. Let D^i denote the matrix resulting from each of these transformations, and let $D = D^1$ denote the resulting matrix at the end of this first induction step. The submatrix of D where the last row has been removed is the stochastic ordering of the submatrix of $\bar{B}$ where the last row has been removed. In particular, it satisfies row monotonicity. Moreover, the j^{th} row of D^i equals that of D for $j \geq i$ and $j \neq n$, and equals that of $\bar{B}^{so}$ for $j < i$. All rows for $j < n$ satisfy row monotonicity. There remains to show that n^{th} row of D^i is satisfies row-monotonicity, for each i. The n^{th} entry d^i_{nk} of the column D^i_k consists of the i^{th} largest entry δ_{ik} of $\bar{B}^{so}_k$, if the entry d_{nk} of D_k is smaller than δ_{ik}, and to d_{nk} otherwise. Now consider any two consecutive columns $k-1$ and k. We must show that $d^i_{n,k} \leq d^i_{n,k-1}$. If $d^i_{n,k} = d_{nk}$, then we use that $d_{nk} \leq d_{n,k-1} \leq d_{i,k-1}$. If $d^i_{n,k} = \delta_{i,k}$, then we use that $\delta_{i,k} \leq \delta_{i,k-1} \leq d^i_{n,k-1}$. This establishes row monotonicity and, therefore, the applicability of the Lemma 2.