The Supermodular Stochastic Ordering

Margaret Meyer Bruno Strulovici*

March 30, 2011

Abstract

This paper uses the stochastic dominance approach to compare the interdependence of random vectors. One multivariate distribution dominates another according to the supermodular stochastic ordering if it yields a higher expectation than the other for all supermodular objective functions. We show that this ordering is equivalent to one distribution's being derivable from another by a sequence of particular bivariate, marginal-preserving transformations, and develop methods for determining whether such a sequence exists. When random vectors result from common and idiosyncratic shocks, we provide non-parametric sufficient conditions for supermodular dominance. Moreover, we characterize the orderings corresponding to supermodular objective functions that are also increasing or symmetric. For the symmetric case, we compare reward distributions generated by lotteries and tournaments. Applications to welfare economics, matching markets, insurance, and finance are discussed.

Keywords: Interdependence, Correlation, Copula, Supermodular, Ultramodular, Concordance, Tournament, Mixture, Majorization.

*Email addresses: margaret.meyer@nuffield.ox.ac.uk b-strulovici@northwestern.edu. First draft: February 8, 2010. We are grateful for valuable comments from David Cox, Ian Jewitt, Paul Klemperer, and seminar participants at the University of Wisconsin-Madison, Université de Montréal, University of Oxford Department of Statistics, the Risk, Uncertainty, and Decision Conference, and the Third Transatlantic Theory Workshop.

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. The stochastic dominance approach expresses attitudes towards interdependence 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 each 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 dependence among claims in a portfolio of insurance policies or among assets in a financial institution's portfolio; and v) systemic risk in financial and macroeconomic systems. Section 9 also briefly shows how our approach permits a non-parametric comparison of copulas.

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. This equivalence breaks down for three or more dimensions (Joe, 1990, and Müller and Scarsini, 2000).[1] In general, the supermodular ordering is strictly stronger than the combination of upper-orthant and lower-orthant

[1] Hu, Xie, and Ruan (2005) have shown that this equivalence continues to hold in three dimensions in the special case of Bernoulli random vectors.

dominance.

Focusing on the case where random vectors have supports on a finite lattice, we characterize the supermodular ordering for an arbitrary number of dimensions. Section 4 proves (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." Our elementary transformations play a role similar to the mean-preserving spreads defined by Rothschild and Stiglitz (1970) for univariate distributions to capture the notion of increased 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 all but two 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 4.1 shows how our definition of elementary transformations allows a very simple constructive proof of the known characterization of the supermodular ordering for bivariate distributions. Our constructive proof is based on the observation that, for any pair of bivariate distributions with identical marginals, if we allow elementary transformations to have weights of arbitrary sign, then there is a unique weighted sequence of such transformations that converts one distribution into the other. For pairs of distributions f, g in three or more dimensions, even with our restrictive definition of elementary transformations, there are many weighted sequences of such transformations that convert one distribution into the other. How, then, can we determine whether g dominates f according to the supermodular ordering? In Section 4.2, we describe two different methods. One method is to formulate a linear program 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. An alternative method, based on Minkowski's and Weyl's representation theorems for polyhedral cones, 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, where each mechanism or policy generates a multivariate distribution, and the set of mechanisms to be compared is large.

In some applications, it is natural to focus on objective functions that are symmetric. In Meyer and Strulovici (2011), we study the *symmetric supermodular ordering*, which corresponds to one distribution's generating a higher expected value than another for all symmetric supermodular objective functions. In particular, we show that two distributions are ranked according to the symmetric supermodular ordering if and only if the "symmetrized" versions of the distributions are ranked according to the supermodular ordering. In this paper (Section 5), we use that characterization to show that the symmetric supermodular ordering for any number of dimensions and l points in the support of each dimension is equivalent to a closely related ordering on an $l-1$-dimensional support. For n-dimensional random vectors representing n independent lotteries, we identify in Theorem 5 in Section 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 7 studies the special case of multivariate distributions generated as follows: first, a common shock determines probability distributions for each random variable. Then, all random variables are drawn independently from those distributions. The resulting multivariate distribution is a *mixture* of conditionally independent 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 independent random variables. Specifically, we provide in Theorem 6 sufficient conditions for two 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 5.

Section 8 extends our approach of using duality results for polyhedral cones to characterize a range of other stochastic orders, such as the increasing supermodular ordering and the supermodular and componentwise convex ordering.

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 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 an arbitrary 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, as an implication of Theorem 5 in Section 6, 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 when individual-level 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, our constructive methods for checking supermodular dominance, described in Section 4.2 and Appendix D, can be used to compare 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.

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, the recent financial crisis has stimulated interest in the development of measures of interdependence for the components of the financial system as a whole and not just for individual assets. Adrian and Brunnermeier (2009) and Beale et al (2011) for example, study interdependence among the returns of 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

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 with the following partial order: $z \leq v$ if and only if $z_i \leq v_i$ 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.

For any $z \in L$, let $z + e_i$ denote the element v of L, whenever it exists, such that $v_j = z_j$ for all $j \in N \setminus \{i\}$ and v_i is the smallest element of L_i greater than but not equal to z_i. For example, if $L = \{0,1\}^2$, $(0,0) + e_1 = (1,0)$ 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 $\mathbb{R}^d$:

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

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:

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

4 The Stochastic Supermodular Ordering

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

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

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

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

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

If two distributions f and g are such that $g = f + \alpha t^z_{i,j}$ for some $\alpha \geq 0$, then we say that g is obtained from f by an elementary transformation with weight α. The α-weighted elementary transformation raises the probability of nodes z and $z + e_i + e_j$ by the common amount α, reduces the probability of nodes $z + e_i$ and $z + e_j$ by the same amount, and leaves unchanged the probability assigned to all other nodes in L. Intuitively, such transformations increase the degree of interdependence of a multivariate distribution, as for some pair of components i and j, they make jointly high and jointly low realizations more likely, while making realizations where one component is high and the other low less likely. Furthermore, they raise interdependence without altering the marginal distribution of any component.

If, for example, $L = \{0, 1, 2\}^2$, there are four elementary transformations, corresponding to the four values of z , namely $(0, 0)$, $(1, 0)$, $(0, 1)$, and $(1, 1)$, such that $z + e_i + e_j$ is well defined. For $L = \{0, 1\}^3$, there are six elementary transformations, one corresponding to each face of the unit cube. 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 $L = \{0, 1\}^3$) and ii) affect values only at four *adjacent* points in the lattice, z, $z+e_i$, $z+e_j$, and $z+e_i+e_j$ (as illustrated by the example of $L = \{0,1,2\}^2$).

Using our notation for elementary transformations, (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:

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

If $f \prec_{SPM} g$, 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) *$f \prec_{SPM} g$ 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)), this equivalence means that $f \prec_{SPM}$ g if and only if $g - f \in \mathcal{T}^C$, which is the result we sought. ■

Observe that since any elementary transformation $t \in \mathcal{T}$ leaves the marginal distributions unchanged, it is an immediate implication of Theorem 1 that if $f \prec_{SPM} g$, then f and g have identical marginal distributions.

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

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

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

Proof. Suppose that on L, $f \prec_{SPM} g$. 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 $\{\alpha'_t\}_{t \in \mathcal{T}(M)}$. Applying Theorem 1 again then yields than on domain M, $f^M \prec_{SPM} g^M$. ■

Thus, if g is more interdependent than 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.

The problem of using Theorem 1 to determine, given a pair of distributions f and g, whether or not $f \prec_{SPM} g$ is greatly facilitated by two aspects of our approach. The first is our restriction to a *finite* support L. The second is our restriction that elementary transformations affect only two of the n dimensions and affect values at only adjacent points in the lattice. These two restrictions make it 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 result, whose proof is in the Appendix.

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

For 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 $L = \{0,1\}^3$, $t_{13}^{(0,0,0)} = t_{13}^{(0,1,0)} - t_{23}^{(1,0,0)} + t_{23}^{(0,0,0)}$.

4.1 Two Dimensions

Theorem 1 tells us that, given two distributions f, g, determining whether $f \prec_{SPM} g$ 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, given our definition of elementary transformations, this determination 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 $t_{i,j}^z$ 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 $z \in L_{(l_1-1),(l_2-1)}$,

$$\begin{aligned}\delta(z) &= \alpha(z)t^z(z) + \alpha(z-e_1)t^{(z-e_1)}(z) + \alpha(z-e_2)t^{(z-e_2)}(z) + \alpha(z-e_1-e_2)t^{(z-e_1-e_2)}(z) \\ &= \alpha(z) - \alpha(z-e_1) - \alpha(z-e_2) + \alpha(z-e_1-e_2), \qquad (8)\end{aligned}$$

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

A simple inductive process allows us to solve the equations (8) 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 $t^{(0,0)}$, (8) 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 $t^{(0,0)}$, (8) 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 $f \prec_{\mathcal{SPM}} g$ if and only if every elementary transformation has a nonnegative weight in the decomposition, it follows that for two dimensions,

$$f \prec_{\mathcal{SPM}} g \quad \Leftrightarrow \quad G(z) - F(z) \geq 0 \quad \forall z \in L. \qquad (9)$$

Note that (9) 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

$$G(z) - F(z) \geq 0 \quad \forall z \in L \quad \Leftrightarrow \quad \overline{G}(z) - \overline{F}(z) \geq 0 \quad \forall z \in L. \qquad (10)$$

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

$f \prec_{CONC} g$, 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 (9) and (10) we can conclude that

$$f \prec_{SPM} g \quad \Leftrightarrow \quad f \prec_{CONC} g. \tag{11}$$

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 $f \prec_{CONC} g$ implies $f \prec_{SPM} g$ by defining a notion of a simple "correlation increasing" transformation.[2] 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 $\alpha_{t^z} \equiv \alpha(z) = G(z) - F(z)$.

Now

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

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

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

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.

4.2 Constructive Methods for Comparing Distribution Interdependence

For three or more dimensions, how can one determine whether $f \prec_{\mathcal{SPM}} g$? We provide two answers to this question, which exploit Theorem 1's dual characterization of the ordering as well as Theorem 3's result that all elementary transformations as defined in (3) are extreme. Details of these methods are provided in the Appendix.

One approach 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 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 second 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 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, where each mechanism generates a multivariate distribution, and the set of mechanisms or policies 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 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.

5 The 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 4) how it relates to the supermodular ordering, and provide some characterization results (Propositions 1 and 2). Theorem 5 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,

$$\theta^{symm}(z) = \frac{1}{n!} \sum_{\sigma \in \Sigma(n)} \theta(\sigma(z)), \tag{12}$$

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. We can now state the following useful result, from Meyer and Strulovici (2011):

THEOREM 4 *Given a pair of distributions f, g defined on L, $f \prec_{SSPM} g$ if and only if $f^s \prec_{SPM} g^s$.*

Theorem 4 states that one can characterize the symmetric supermodular order in terms of the supermodular order applied to symmetric distributions.

This theorem is 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 individual 1 got the high prize and 2 the low prize, or vice versa. Theorem 4 will be very helpful in interpreting the application of Theorem 5 in Section 6 to the welfare comparison of reward distributions generated by tournaments and independent lotteries.

5.1 Binary Variables, n Dimensions

Suppose each of n random variables is binary, so $L = \{0, 1\}^n$. If the objective function is symmetric, then only the number of 1's, $c(x) = \sum_{i=1}^{n} I_{\{x_i=1\}}$, contained in any vector $x \in L$ 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 $\tilde{L}^1$. We prove this using 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 8.3). Since their dual cones are equivalent, it follows therefore that symmetric supermodular functions on $L = \{0,1\}^n$ are equivalent to convex functions on $\tilde{L}^1 = \{0,1,\ldots,n\}$. We have therefore proved:

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

5.2 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 symmetric function $w : L \to \mathbb{R}$ 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 1, 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 $y + 2e_k \in \tilde{L}^{l-1}$, $\tilde{w}(y) + \tilde{w}(y+2e_k) \geq 2\tilde{w}(y+e_k)$.[3] 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^y_k on $\tilde{L}^{l-1}$ such that

$$t^y_k(y) = t^y_k(y+2e_k) = 1 \qquad t^y_k(y+e_k) = -2, \tag{13}$$

and $t^y_k(z) = 0$ for all other $z \in \tilde{L}^{l-1}$. Proposition 1 can now be generalized to:

Proposition 2 *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 2 is proved by showing that w is symmetric and supermodular on L if and only if $\tilde{w}$

[3]See Section 9 for more detail.

is supermodular and componentwise-convex on $\tilde{L}^{l-1}$.[4] To show this, we use the dual approach and show that any "supermodular" elementary transformation $t^x_{i,j}$ 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 (13) characterizing componentwise-convexity. For transformations $t^x_{i,j}$ 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 (13). For transformations $t^x_{i,j}$ 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).

Propositions 1 and 2 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.

6 Lotteries, Tournaments, and 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 i^{th} 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

[4] 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 $\mathbb{R}^{l-1}$, $\tilde{w}(y \wedge z) + \tilde{w}(y \vee z) \geq \tilde{w}(y) + \tilde{w}(z)$.

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 stochastically ordered 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[5] the k^{th} column vector of $\bar{B}$. That is, $A \succ_{CCM} B$ if for each $k \in \{1, \ldots, m\}$ and for each $l \in \{1, \ldots, n\}$

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

with equality holding for $l = 1$.

THEOREM 5 *Let A and B be two $n \times m$ row-stochastic matrices such that, for each $j \leq m$, the j^{th} 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 5. Recall that Theorem 4 showed that $f \prec_{SSPM} g$ if and only if $f^s \prec_S g^s$. 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. Theorem 5 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.[6]

To illustrate this interpretation of Theorem 5, 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., all rows and columns sum to 1. A "tournament" is a mechanism that allocates, according to some

[5]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)).

[6]Hu and Yang (2004, Theorem 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 supermodularly 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 5 where the rows of the matrix B are all identical.

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 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 5 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 5 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, over m possible levels, on the ith of n tasks, and suppose that output levels are independently distributed across tasks. Suppose that the production function is symmetric and supermodular in the output levels on the different tasks, reflecting the fact that task outputs are, respectively, identically valued and complementary. Two row-stochastic matrices with matching column sums then describe two different production settings in which the distribution of output, averaged across all tasks, is the same. Theorem 5 then identifies conditions under which expected production is higher in one setting than the other 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 1 and Theorem 5. With binary output levels on the tasks, a benefit function for the principal that is a convex function of the total number of successes is a symmetric supermodular function of the vector of task outputs. The effort allocation determines an $n \times 2$ row-stochastic matrix, the second 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 $m = 2$, Theorem 5 implies that, holding total effort fixed, if one effort allocation yields a vector of success probabilities that majorizes the vector yielded by 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).[7]

We have examples showing that Theorem 5 does not hold if we relax either the assumption that A is stochastically ordered or that $A \succ_{CCM} B$.

Theorem 5 has the following useful corollary, which is proved in the Appendix.

[7]Bond and Gomes's results follow from a result due to Karlin and Novikoff (1963), which is the special case of Theorem 5 when $m = 2$.

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.

7 Aggregate vs. Idiosyncratic Shocks

This section considers random vectors that are the result of aggregate and idiosyncratic shocks. To each variable X_r is associated a $q \times l$ stochastic matrix A_r. Each row of A_r represents a probability distribution for the variable X_r on the support $\{0, 1, \ldots, l-2\}$. The vector $(X_1, \ldots, X_n)$ is constructed as follows. First, a row index $\iota \in \{1, \ldots, q\}$ is drawn randomly, according to the uniform distribution.[8] Then, each variable X_r is independently drawn from the ι^{th} row A_r.[9]

We provide sufficient conditions for two mixture distributions to be ranked according to the supermodular ordering. These conditions have a natural interpretation as a non-parametric ordering of the relative size of aggregate vs. idiosyncratic shocks. In finance and insurance contexts, 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 (Cousin and Laurent, 2008). In macroeconomics, the relative importance of aggregate vs. sectoral shocks affects variation and covariation of output levels (Foerster, Sarte, and Watson, 2011).

[8]Uniform distribution is just for convenience, the argument extends easily for other distributions of ι, for example by replicating rows of the matrices.

[9]For the special case where all the matrices A_r are the same, Shaked (1977) defines random variables in this manner as "positively dependent by mixture."

Our goal is to establish the following result. Suppose that $(A_1, \ldots, A_n)$ and $(B_1, \ldots, B_n)$ are two sets of stochastic matrices, generating the random vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$, respectively. For any matrix A, we let $\bar{A}$ denote the cumulative matrix of A (see Section 6).

THEOREM 6 *Suppose that*

- *For each r, the rows of A_r are stochastically ordered.*
- $A_r \succ_{CCM} B_r$.[10]

Then,

$$(X_1, \ldots, X_n) \succeq_{SPM} (Y_1, \ldots, Y_n).$$

$A_r \succ_{CCM} B_r$ implies 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 $Y_i \geq k$: the marginal distribution of X_i is the same as the marginal distribution of Y_i.

Just as for Theorem 5, we have examples showing that Theorem 6 does not hold if we relax either the assumption that A is stochastically ordered or that $A \succ_{CCM} B$.[11]

The theorem is proved in its full generality in Appendix C. For expositional simplicity, we focus in the main text on the case where the matrices $\{A_r\}_{r\in\{1,\ldots,n\}}$ ($\{B_r\}_{r\in\{1,\ldots,n\}}$) are independent or r, so that the random vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ both have symmetric distributions.

A distribution on the support $\{0, \ldots, l-1\}$ may be 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 $\bar{p}^1, \ldots, \bar{p}^n$, respectively. 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 3 *If w is supermodular, u^w is supermodular and componentwise convex.*

[10] See Section 6 for the definition of the Cumulative Column Majorization Criterion.

[11] 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 (2011) 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 6 where the rows of the matrix B are all identical.

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

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

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 $r, s \in \{0, \ldots, l-1\}$.*

Proof. 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,[12] 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 Proposition 3, 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. ■

The proof of Theorem 6, which is in Section C 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 $k \in \{1, l-1\}$,*

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

[12] 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.

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

The function $u = u^w$ is polynomial in $\bar{p}$ and hence twice differentiable. Moreover, it is componentwise convex and supermodular from Proposition 3, 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 6 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 6 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 $\bar{A}$ 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.

8 Increasing Supermodular Ordering, Ultramodular Ordering, and Other Difference-Based Orderings

This section generalizes Section 4 to study *difference-based orderings*, which have a particular linear structure. Using duality, we show how the order based on any two difference-based orders is characterized by elementary transformations from both of the initial orders. We exploit this to study the increasing supermodular ordering and the ultramodular ordering.

8.1 Combined Properties of Objective Functions

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 7 (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. ■

8.2 Increasing Supermodular Ordering

To accommodate the introduction of new types of elementary transformations, let $\mathcal{T}(\mathcal{S})$ denote the set of elementary transformations characterizing the set $\mathcal{S}$ of supermodular functions.

A function w on L is increasing if for any $x \in L$ and i such that $x + e_i \in L$, $w(x + e_i) \geq w(x)$. Let $\mathcal{M}$ denote the set of increasing functions on $\mathcal{L}$. For any $x \in L$ and i such that $x + e_i \in L$, let m_i^x denote the function on L such that $m_i^x(x) = -1$, $m_i^x(x + e_i) = 1$, and m_i^x vanishes everywhere else. Let $\mathcal{T}(\mathcal{M})$ denote the set of all such functions. One may easily check that

$$w \in \mathcal{M} \Longleftrightarrow w \cdot m \geq 0 \qquad \forall m \in \mathcal{T}(\mathcal{M}).$$

First-order stochastic dominance for distributions on L is defined by

$$g \succeq_{FOSD} f \Longleftrightarrow w \cdot g \geq w \cdot f \qquad \forall w \in \mathcal{M}.$$

It is easy to adapt the proof of Theorem 1 to show that $g \succeq_{FOSD} f$ if and only if there exist nonnegative coefficients $\{\beta_m\}_{m\in\mathcal{T}(\mathcal{M})}$ such that

$$g = f + \sum_{m\in\mathcal{T}(\mathcal{M})} \beta_m m. \tag{14}$$

The *increasing supermodular ordering* (denoted $\succeq_{ISPM}$) is defined as follows.

$$g \succeq_{ISPM} f \Longleftrightarrow w \cdot g \geq w \cdot f \qquad \forall w \in \mathcal{S} \cap \mathcal{M}.$$

Since the functions w are now required to be increasing, $g \succeq_{ISPM} f$ does not imply that g and f have identical marginals. Rather, as can be seen by taking, for each $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, l_i - 1\}$, $w(z) = I_{\{z_i \geq k\}}$, which is increasing and supermodular, $g \succeq_{ISPM} f$ implies that each marginal distribution of g dominates the corresponding marginal distribution of f according to first-order stochastic dominance. Theorem 8 below shows that, if one corrects for the differences in marginal distributions, then checking increasing supermodular dominance is equivalent to checking supermodular dominance.

To simplify notation, assume that $L_i = \{0, 1, \ldots, l_i - 1\}$. Given two distributions f and g with $\delta \equiv g - f$, let γ denote the function on L such that $\gamma(x)$ vanishes everywhere except on the set L_0 of x's that have at most one positive component, and where, for any $i \in \{1, \ldots n\}$ and $k \in \{1, 2, \ldots, l_{i-1}\}$,

$$\gamma(ke_i) = Pr(Y_i = k) - Pr(X_i = k) = \sum_{z: z_i = k} \delta(z),$$

and

$$\gamma(0) = \delta(0) - \sum_{z \notin L_0} \delta(z).$$

It is easy to check that for all i and k, including $k = 0$,

$$\sum_{z: z_i = k} \gamma(z) = \sum_{z: z_i = k} \delta(z). \tag{15}$$

Intuitively, if $g \succeq_{ISPM} f$, then γ captures the first-order stochastic dominance of g over f, while $g - f - \gamma$ captures the supermodular dominance of g over f.

Theorem 8 (Increasing Supermodular Ordering) *The following statements are equivalent:*

1) $g \succeq_{ISPM} f$*,*

2) There exist nonnegative coefficients $\{\alpha_t\}_{t \in \mathcal{T}(\mathcal{S})}$*,* $\{\beta_m\}_{m \in \mathcal{T} \mathcal{M}}$ *such that*

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

3) There exist nonnegative coefficients $\{\alpha_t\}_{t \in \mathcal{T}(\mathcal{S})}$*,* $\{\beta_m\}_{m \in \mathcal{T}(\mathcal{M}}$ *such that*

a) $\gamma = \sum_{m \in \mathcal{T}(\mathcal{M})} \beta_m m$*, and* *b)* $g = f + \gamma + \sum_{t \in \mathcal{T}(\mathcal{S})} \alpha_t t$

4) $w \cdot (g - f - \gamma) \geq 0$ *for all supermodular* w*, and for each* i *the* i^{th} *marginal distribution of* g *dominates the* i^{th} *marginal of* f *according to first-order stochastic dominance.*

Proof. The equivalence of 1) and 2) is an immediate consequence of Theorem 7. The equivalence of 3) and 4) follows from Theorem 1, the definition of γ, and the decomposition result in (14). It is obvious that 3) implies 2). We now show that 1) implies 4). For any $w \in \mathcal{S}$, let

$$w^0(z) = w(z) - \sum_{i=1}^{n} w(z_i e_i) + (n-1)w(0).$$

Clearly, $w^0(z_i e_i) = 0$ for all i, and hence $w^0 \cdot \gamma = 0$. Moreover, w^0 is supermodular, since it is the sum of supermodular functions, and w^0 is increasing, since for any $z \in L$ and i such that $z + e_i \in L$, we have

$$w^0(z + e_i) - w^0(z) \geq w^0((z_i + 1)e_i) - w^0(z_i e_i) = 0$$

by supermodularity of w^0.

Letting $\delta = g - f$, 1) implies, therefore, that $w^0 \cdot \delta \geq 0$ and hence $w^0 \cdot (\delta - \gamma) \geq 0$. Finally, exploiting (15) several times, we have

$$\begin{aligned}(w - w^0) \cdot (\delta - \gamma) &= \sum_z (\delta(z) - \gamma(z)) \left(\sum_{i=1}^{n} w(z_i e_i) - (n-1)w(0) \right) \\ &= \sum_{i=1}^{n} \sum_{k=0}^{l_i - 1} w(ke_i) \sum_{z : z_i = k} (\delta(z) - \gamma(z)) = 0, \quad (16)\end{aligned}$$

which shows that $w \cdot (\delta - \gamma) \geq 0$. Finally, taking, for each $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, l_i - 1\}$, $w(z) = I_{\{z_i \geq k\}}$, 1) implies

$$\sum_{z : z_i \geq k} g(z) \geq \sum_{z : z_i \geq k} f(z),$$

which proves the second part of 4). ■

8.3 Componentwise Convex Ordering

Some applications require more properties of objective functions than just 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 consider the case of objective functions that are supermodular and componentwise convex. The case of supermodular, componentwise concave objective functions can be analyzed similarly.

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, suppose that $L_i = \{0, 1, \ldots, l_i - 1\}$ for each dimension i: points in the support are equally spaced. 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

$$t_i^x(x) = t_i^x(x + 2e_i) = 1 \quad \text{and} \quad t_i^x(x + e_i) = -2, \tag{17}$$

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.[13] As is easily checked, these elementary transformations entirely characterize componentwise convex functions, that is:

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

Theorem 9 (Componentwise Convex Ordering) *$f \prec_{\mathcal{X}} g$ if and only if there exist nonnegative coefficients θ_t, $t \in \mathcal{T}(\mathcal{X})$, such that*

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

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

For the supermodular ordering, we showed in Section 4.1 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

[13] If for some i the points in L_i are not equally spaced, the definition (17) 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.

arbitrary signs. For the componentwise-convex ordering, the case of one dimension is special in an analogous sense. 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. Given this uniqueness, it follows from Theorem 9 that $f \prec_{\mathcal{X}} g$ if and only if the weight on every elementary transformation in the decomposition is nonnegative.

For multidimensional distributions, determining whether g dominates f according to the componentwise convex ordering requires combining Theorem 9 with the analog of one of the constructive methods described in Section 4.2 for the supermodular ordering.

8.4 Ultramodular Ordering

Theorem 7 also yields a characterization of the ultramodular ordering. A function is *ultramodular* if it is supermodular and componentwise convex (see Marinacci and Montrucchio, 2005). The negative of an ultramodular function is *inframodular*, i.e. submodular and componentwise concave.

Our motivation for characterizing the ultramodular ordering is two-fold. First, Propositions 2 and 3 showed that ultramodular functions arise naturally in the characterization of the supermodular ordering in specific environments. Second, Müller and Scarsini (2011) have argued that inframodularity is the natural multivariate generalization of univariate concavity, in that it captures "fear of loss and not just aversion to randomness" (p. 2). Analogously, then, ultramodularity can be seen as the natural multivariate generalization of univariate convexity.

Corollary 2 *Let $\mathcal{U}$ denote the set of ultramodular objective functions. $f \prec_{\mathcal{U}} g$ if and only if there exists a nonnegatively weighted sequence of elementary transformations of either type t_i^x (defined in (17)) or type $t_{i,j}^x$ (defined in (3)) that, added to f, yield g.*

Müller and Scarsini (2011) have derived a similar characterization for the set of inframodular functions. 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.

We now use the methods of Section 8.2 to characterize the increasing ultramodular order. Say that $g \succeq_{IUM} f$ if $w \cdot g \geq w \cdot f$ for all functions w that are increasing and ultramodular.

For any $i \in \{1, \ldots, n\}$, the function $w(z) = a + bz_i$, where a and $b > 0$ are constants, is increasing and ultramodular. Therefore if Y, X are distributed according to g, f, respectively, $g \succeq_{IUM} f$

implies that for each i, $E[Y_i] \geq E[X_i]$. Theorem 10 below shows that, if one corrects for the differences in univariate means, then checking increasing ultramodular dominance is equivalent to checking ultramodular dominance.

Continue to take $L_i = \{0, 1, \ldots, l_i - 1\}$. Given f, g with $\delta \equiv g - f$, define the function λ on L such that λ vanishes everywhere except on the set of points $\{0, e_1, \ldots, e_n\}$, and

$$\lambda(e_i) = \sum_{k=0}^{l_i - 1} k \sum_{z : z_i = k} \delta(z) \quad \forall i \qquad \text{and} \qquad \lambda(0) = -\sum_{i=1}^{n} \lambda(e_i). \tag{18}$$

Intuitively, if $g \succeq_{IUM} f$, then λ captures the difference in the univariate means, while $g - f - \lambda$ captures the ultramodular dominance of g over f.

Theorem 10 (Increasing Ultramodular Ordering) *The following statements are equivalent*

1) $g \succeq_{IUM} f$

2) λ *is decomposable into a nonnegative weighted sum of transformations* $m_i \in \mathcal{T}(\mathcal{M})$, $i \in \{1, \ldots, n\}$, *where each* m_i *transfers weight from 0 to* e_i, *and* $g - f - \lambda$ *is decomposable into a nonnegative weighted sum of elementary transformations of either type* t_i^x *(defined in (17)) or type* $t_{i,j}^x$ *(defined in (3)).*

3) $w \cdot (g - f - \lambda) \geq 0$ *for all ultramodular functions* w, *and* $E[Y_i] \geq E[X_i]$ *for all* $i \in \{1, \ldots, n\}$.

Proof. That 3) implies 2) follows from Corollary 2 and the definition of λ. That 2) implies 1) is obvious. To show that 1) implies 3), define, for any ultramodular w,

$$w^0(z) = w(z) - w(0) - \sum_{i=1}^{n} z_i [w(e_i) - w(0)].$$

Clearly, $w^0(e_i) = 0$ for all i and $w^0(0) = 0$, and hence $w^0 \cdot \lambda = 0$. Moreover, w^0 is ultramodular, since it is the sum of ultramodular functions, and w^0 is increasing, since for any $z \in L$ and i such that $z + e_i \in L$, we have

$$w^0(z + e_i) - w^0(z) \geq w^0((z_i + 1)e_i) - w^0(z_i e_i) \geq w^0(e_i) - w^0(0) = 0,$$

where the first inequality follows from supermodularity and the second from componentwise convexity of $w^0(z)$.

Letting $\delta = g - f$, 1) implies, therefore, that $w^0 \cdot \delta \geq 0$ and hence $w^0 \cdot (\delta - \lambda) \geq 0$. Finally,

$$(w - w^0) \cdot (\delta - \lambda) = \sum_z (\delta(z) - \gamma(z)) \left(w(0) + \sum_{i=1}^{n} z_i [w(e_i) - w(0)] \right)$$
$$= \sum_{i=1}^{n} (w(e_i) - w(0)) \sum_{k=0}^{l_i - 1} k \sum_{z : z_i = k} (\delta(z) - \lambda(z)) = 0, \quad (19)$$

where the final equality follows from (18). Hence $w \cdot (\delta - \gamma) \geq 0$. Finally, taking, for each $i \in \{1, \ldots, n\}$ $w(z) = a + bz_i$, where a and $b > 0$ are constants, 1) implies $E[Y_i] \geq E[X_i]$, proving the second part of 3). ■

9 Relation to Copulas

An increasingly popular way to think about interdependence across random variables is via the concept of copula. 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. A function C is a *copula* if it is the joint distribution of n uniform random variables.

Theorem 11 (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*

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

Theorem 1 implies that for two distributions to be comparable according to the supermodular ordering, they must have identical marginals. From Sklar's representation, this implies that their copulas are defined on the same domain (which is finite, in our context), since the range of the functions F_i's is identical for the two distributions. We can therefore apply our analysis to copulas defined on the lattice $L = \times_i L_i$ such that L_i consists of all elements in the range of F_i.[14]

As a result, our characterization of the supermodular ordering permits a non-parametric comparison of interdependence across copulas.

[14] Decancq (2009) starts directly with copulas to study interdependence of multivariate distributions via stochastic orderings.

Appendices

A Proof of Theorem 3

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

$$t^* = \sum_{s \in \mathcal{T} \setminus \{t^*\}} \alpha_s s. \tag{20}$$

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 (20) 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. ■

B Proof of Theorem 5 and Its Corollary

The proof of Theorem 5 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

$$0 \leq \left(\sum_{j=k}^{m} \beta_{1j}\right)\left(\sum_{j=k}^{m} \beta_{2j}\right) - \left(\sum_{j=k}^{m} \alpha_{1j}\right)\left(\sum_{j=k}^{m} \alpha_{2j}\right), \tag{21}$$

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 $\bar{\beta}$,

$$0 \leq \bar{\beta}_{1k}\bar{\beta}_{2l} - \bar{\alpha}_{1k}\bar{\alpha}_{2l} + \bar{\beta}_{1l}\bar{\beta}_{2k} - \bar{\alpha}_{1l}\bar{\alpha}_{2k}. \tag{22}$$

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 (21) *for each k, and* (22) *for each $k \neq l$.*

Proof. Since all ε_j's simplify in the above β sums except for ε_k, Condition (21) 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 (22) 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 5. 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 $A \prec_{SSPM} B^{so}$.

B.1 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.[15] Such a construction is given by Hardy et al. (1952). At each step, the last column of the resulting matrix is stochastically ordered, 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 $\bar{A} \prec_{SSPM} \bar{B}^{so}$.

B.2 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

[15] 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 $\bar{A}$.

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 5

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 $i \leq i_1 = \lfloor np_1 \rfloor$, $a_{i_1+1,1} = np_1 - i_1$, 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 $i \leq i_1$ 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.

C Proof of Theorem 6

The proof proceeds in three steps. We first establish the result for the case of symmetric distributions (i.e., A_r and B_r are independent of r) and when B is stochastically ordered (Step 1). We then generalize it to the case where B is not stochastically ordered (Step 2). Finally, we prove it for the case of asymmetric distributions A_r, B_r depend on r.

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.

C.1 Step 1: $\bar{B}$ stochastically ordered

We first assume for now that $\bar{B}$ is also stochastically ordered, so that $\bar{b}_{ik}$ is increasing in i. Here and throughout, we exclude the first column of ones that could conventionally appear in cumulative probability matrices.

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

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

We will simply say that $\bar{B}$ is "strictly monotonic."

Let k denote the smallest 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 $\tilde{k} \leq k$, whose columns majorize $\bar{B}$'s, and such that $\bar{A} \succ_{SPM} C$.*

Proof.

Let C solve the optimization problem

$$\inf_E \sum_{i\geq 2} \sum_{j\geq i} e_{j,k} \tag{23}$$

subject to the following constraints:

1. E has entries in $[0, 1]$.
2. E satisfies row monotonicity (i.e., entries of E are decreasing in the column index),
3. E is stochastically ordered (i.e., entries of E are increasing in the row index),
4. E dominates $\bar{B}$ according to the cumulative column criterion (i.e., each column of E's majorizes the corresponding column of $\bar{B}$),
5. $\bar{A}$ dominates E according to the stochastic supermodular ordering
6. $E_{\tilde{k}} = \bar{B}_{\tilde{k}}$ 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 (23) is continuous. Therefore, its minimum is reached by some C. We show that C_k is equal to $\bar{B}_k$, 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 constraints of the minimization problem (23). Such change only affects the $i + 1$ partial sum of (23), and decreases it by an amount ε, which will contradict the assumption that C minimizes (23).

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, $\bar{k}$is equal to the number of columns of C.

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

If $\bar{k} > k$, 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 $\bar{k}$.[16]

D is stochastically ordered provided that $c_{i,k} + \varepsilon \leq c_{i+1,k} - \varepsilon$, which holds for all $\varepsilon \leq \chi/2$. This claim 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

$$\sum_{j \geq i+1} d_{j,\tilde{k}} \geq \sum_{j \geq i+1} \bar{b}_{j,\tilde{k}} \tag{24}$$

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,

[16] If $\bar{k}$ equals the number of columns of C, 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.

we have

$$\sum_{j\geq i+2} d_{j,\tilde{k}} = \sum_{j\geq i+2} c_{j,\tilde{k}} \geq \sum_{j\geq i+2} \bar{b}_{j,\tilde{k}} \tag{25}$$

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 (25) implies (24).

Lemma 2 implies that $C \succ_{SPM} D$. 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 (23) starting from row $i+1$ and left other partial sums unaffected. This concludes the proof of the lemma. ■

To conclude the proof of Theorem 6 when $\bar{B}$ is strictly monotonic, we apply the above lemma inductively on k.

If $\bar{B}$ is not strictly monotonic, we approximate $\bar{A}$ and $\bar{B}$ by a sequence of cumulative matrices $A(n), B(n)$ that are strictly increasing with $\chi_n = 1/n$, that converge to $\bar{A}$ and $\bar{B}$ as $n \to \infty$, and such that A_n majorizes B_n. The previous analysis shows 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 this approximating sequence exists for n large enough, we scale down the entries of $\bar{A}$ and $\bar{B}$ by a factor $\delta_n = 1 - (p+q)/n$ where $p \times q$ are matrix dimensions of $\bar{A}$ and $\bar{B}$,[17] and add $e_{i,j} = \frac{1}{n}(i+j)$ to the result. The matrices thus constructed, $A(n)$ and $B(n)$, are strictly increasing with factor $1/n$ and have entries less than 1. Moreover, one may easily check that $A(n)$ majorizes $B(n)$, since the scaling and addition operations do not affect the comparison of partial sums.

C.2 Step 2: $\bar{B}$ not stochastically ordered

In general, $\bar{B}$ is not stochastically ordered. Let $\bar{B}^{so}$ denote the stochastically ordered version of $\bar{B}$, 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: the entries of $\bar{B}^{so}$ in 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

[17]Recall that we have excluded the first column of ones that may appear in cumulative matrices

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 5 to convert $\bar{B}^{so}$ to $\bar{B}$ by the sequence of pairwise row transformations performed in Appendix B, and by applying Lemma 2 at each step to show that each transformation results in a matrix that is lower according to the supermodular order. One must ensure that each step preserves row monotonicity, because Lemma 2 applies only to rows that satisfy this constraint. Consider the first stage of the conversion from $\bar{B}^{so}$ to $\bar{B}$, which consists in a series 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 stage. 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 the 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.

C.3 Step 3: Asymmetric Distributions

The key observation is that the function v^w defined in Lemma 1 of Proposition 3 is supermodular and componentwise convex.

We have

$$Ew(X_1, \ldots, X_n) = \sum_{i=1}^{q} \frac{1}{q} v^w(\bar{\alpha}^1_i, \ldots, \bar{\alpha}^n_i)$$

and

$$Ew(Y_1, \ldots, Y_n) = \sum_{i=1}^{q} \frac{1}{q} v^w(\bar{\beta}^1_i, \ldots, \bar{\beta}^n_i),$$

where $\bar{\alpha}^r_i$ ($\bar{\beta}^r_i$) denotes the i^{th} row of $\bar{A}_r$ ($\bar{B}_r$). For each r, let $\bar{\beta}^r_i)^{so}$ denote i^{th} row of the stochastically ordered version of $\bar{B}$, which we denoted $\bar{B}^{so}$.

Replicating the argument used for the symmetric case, and using that the function v^w is supermodular and componentwise convex, we have

$$\sum_{i=1}^{q} v^w(\bar{\alpha}_i^1, \bar{\alpha}_i^2, \ldots, \bar{\alpha}_i^n) \geq v^w((\bar{\beta}^1)_i^{so}, \bar{\alpha}_i^2, \ldots, \bar{\alpha}_i^n) \tag{26}$$

$$\geq v^w((\bar{\beta}^1)_i^{so}, (\bar{\beta}^2)_i^{so}, \ldots, \bar{\alpha}_i^n) \tag{27}$$

$$\geq \cdots \tag{28}$$

$$\geq v^w((\bar{\beta}^1)_i^{so}, (\bar{\beta}^2)_i^{so}, \ldots, (\bar{\beta}^n)_i^{so}) \tag{29}$$

where, at each step, the stochastic ordering of the rows is preserved.

Finally,

$$v^w((\bar{\beta}^1)_i^{so}, (\bar{\beta}^2)_i^{so}, \ldots, (\bar{\beta}^n)_i^{so}) \geq v^w(\bar{\beta}_i^1, \bar{\beta}_i^2, \ldots, \bar{\beta}_i^n).$$

Indeed, one may use the algorithm described in Section B.2 to convert $\bar{B}_r^{so}$ into $\bar{B}_r$ for all r simultaneously. Supermodularity and componentwise convexity of v^w ensures that the inequality above holds.

D Constructive Methods: Linear Programming and Double Description Method

D.1 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[18] linear program (A):

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

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

subject to

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

Theorem 12 (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 $E^+\alpha^* = \delta^+$. ■

D.2 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 $f \prec_{\mathcal{SPM}} g$. 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 $f \prec_{\mathcal{SPM}} g$ 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.

D.3 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:

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

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.

References

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

Adrian, T., and Brunnermeier, M.K. (2009) "CoVar", Fed. Res. Bank of N.Y. Staff Report 348.

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

Avis, D., and Fukuda, K. (1982) "A Pivoting Algorithm for Convex Hulls and Vertex Enumeration of Arrangements and Polyhedra," *Discrete and Computational Geometry*, Vol. 8, pp. 295–313.

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

Beale, N., Rand, D., Battey, H., Croxson, K., May, R., and Nowak, M. (2010) "Individual and Systemic Risk: The Regulator's Dilemma," Working Paper, Harvard University.

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

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

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

Bourguignon, F., and Chakravarty, S.R. (2002) "Multi-dimensional poverty orderings," DELTA Working Paper, 2002-22.

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

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

Decancq, K. (2009) "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. (2005) *Actuarial Theory for Dependent Risks: Measures, Orders, and Models*, Wiley.

Epstein. L., Tanny, S. (1980) "Increasing Generalized Correlation: A Definition and Some Economic Consequences," *Canadian Journal of Economics*, Vol. 13, pp. 16–34.

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

Fishburn, P. C., Lavalle, I. H. (1995) "Stochastic Dominance on Unidimensional Grids," *Mathematics of Operations Research*, Vol 20, pp. 513–525.

Foerster, A.T., Sarte, P.-D. G., and Watson, M.W. (2011) "Sectoral vs. Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, Vol. 119, pp. 1-38.

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

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

Gajdos, T., Maurin, E. (2004) "Unequal Uncertainties and Uncertain Inequalities: An Axiomatic Approach," *Journal of Economic Theory*, Vol. 116, pp. 93–118.

Goldman, A. J. 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, pp. 19-39.

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

Hennessy, D. A., Lapan, H. E. (2003) "A Definition of 'More Systematic Risk' with Some Welfare Implications," *Economica*, Vol. 70, pp. 493–507.

Hu, T., Xie, C., and Ruan, L (2005) "Dependence Structures of Multivariate Bernoulli Random Vectors," *Journal of Multivariate Analysis*, Vol. 94, pp. 172–195.

Hu, T., Yang, J. (2004) "Further developments on sufficient conditions for negative dependence of random variables," *Statistics and Probability Letters*, Vol. 66, pp. 369–381.

Joe, H. (1990) "Multivariate Concordance," *Journal of Multivariate Analysis*, Vol. 35, pp. 12–30.

Jogdeo, K. (1978) "On a Probability Bound of Marshall and Olkin," *Annals of Statistics*, pp. 232–234.

Karlin, S., and Novikoff, A. (1963) "Generalized Convex Inequalities," *Pacific Journal of Mathematics*, Vol. 13, pp. 1251–1279.

Kroll, Y., and Davidovitz, L. (2003) "Inequality Aversion versus Risk Aversion," *Economica*, Vol. 70, pp. 19–29.

Levy, H., and Paroush, J. (1974) "Toward Multivariate Efficiency Criteria," *Journal of Economic Theory*, Vol. 7, pp. 129–142.

Lorentz, G.G. (1953) "An Inequality for Rearrangements," *American Mathematical Monthly*, Vol. 60, pp. 176–179.

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

Marinacci, M., and Montrucchio, L. "Ultramodular Functions," *Mathematics of Operations Research*, Vol. 30, pp. 311–332.

McAfee, R. P. (2002) "Coarse Matching," *Econometrica*, Vol. 70, pp. 2025–34.

Meyer, M.A. (1990) "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.

Meyer, M. A., and Mookherjee, D. (1987) "Incentives, Compensation and Social Welfare," *Review of Economic Studies*, Vol. 45, pp. 209–226.

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

Meyer, M.A., and Strulovici, B. (2011) "Increasing Interdependence of Multivariate Distributions," Working Paper 1523, Ctr. for Math. Studies in Econ. and Mgt. Science, Kellogg School of Management.

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

Müller, A., and Scarsini, M. (2000) "Some Remarks on the Supermodular Order," *Journal of Multivariate Analysis*, Vol. 73, pp. 107–119.

Müller, A., and Scarsini, M. (2001) "Stochastic Comparison of Random Vectors with a Common Copula," *Mathematics of Operations Research*, Vol. 26, pp. 723–740.

MÜLLER, A., AND SCARSINI, M. (2011) "Fear of Loss, Inframodularity, and Transfers," *Journal of Economic Theory*, forthcoming.

MÜLLER, A., AND STOYAN, D. (2002) *Comparison Methods for Stochastic Models and Risks*, Wiley.

PATTON, A. (2009) "Are 'Market Neutral' Hedge Funds Really Market Neutral?," *Review of Financial Studies*, Vol. 22, pp. 2495–2530.

PRAT, A. (2002) "Should a Team Be Homogenous?," *European Economic Review*, Vol. 46, pp. 1187–1207.

ROTHSCHILD, M. AND STIGLITZ, J. E. (1970) "Increasing risk: I. A definition," *Journal of Economic Theory*, Vol. 2, pp. 225–243.

SHAKED, M. (1977) "A Concept of Positive Dependence for Exchangeable Random Variables," *Annals of Statistics*, Vol. 5, pp. 505–515.

SHAKED, M., AND SHANTHIKUMAR, J.G. (1997) "Supermodular Stochastic Orders and Positive Dependence of Random Vectors," *Journal of Multivariate Analysis*, Vol. 61, pp. 86–101.

SHIMER, R., AND SMITH, L. (2000) "Assortative Matching and Search," *Econometrica*, Vol. 68, pp. 343–369.

SKLAR, A. (1959) "Fonctions de répartition à n dimensions et leurs marges," *Publication de l'Institut de Statistique de l'Université de Paris*, Vol, 8, pp. 229–231.

TCHEN, A.H. (1980) "Inequalities for Distributions with Given Marginals," *Annals of Probability*, Vol. 8, pp. 814–827.

TOPKIS, D.M. (1968) *Ordered Optimal Solution*, Doctoral Dissertation, Stanford University.

TOPKIS, D.M. (1978) "Minimizing a Submodular function on a Lattice," *Operations Research*, Vol. 26, pp. 305–331.

TORRISI, F., AND BAOTIC, M. (2005) "CDDMEX - Matlab interface for the CDD solver." Available at http://control.ee.ethz.ch/hybrid/cdd.php