# Beyond Correlation: Measuring Interdependence Through Complementarities

Margaret Meyer Bruno Strulovici*

April 17, 2015

**Abstract**

To assess the overall effect of individual shocks or the conjunction of welfare indicators such as health and education, economists need to measure and compare the interdependence of random variables. However, statistical measures like correlation do not capture an economically-grounded notion of interdependence. This paper studies interdependence through the prism of complementarities, as captured by supermodular objective functions. We characterize the *supermodular ordering* via a minimal class of interdependence-increasing transformations and extend the characterization when objective functions are also monotonic or symmetric. We provide constructive methods for checking whether given distributions are ranked according to the supermodular ordering, and sufficient conditions for comparing random vectors generated by common and idiosyncratic shocks or by heterogeneous lotteries. Applications to welfare economics, committee decisions, and systemic risk are developed.

*Keywords:* Interdependence, Supermodularity, Correlation, Copula, Mixture, Majorization, Tournament. *JEL Codes: D63, D81, G11, G22*

## 1 Introduction

The interdependence of random variables is of central interest to economists: It determines the macroeconomic consequences of firm-level shocks, the solvency of insurance companies protecting large numbers of households, and the price of financial derivatives whose payoffs depend on the

*We are grateful for comments from David Cox, Ian Jewitt, Paul Klemperer, and seminar participants at Cambridge, Université de Montréal, Oslo, Oxford Economics, Oxford Statistics, QMUL, Stanford (conference in honor of Paul Milgrom), Warwick, Wisconsin-Madison, Yale, the RUD Conference, the Transatlantic Theory Workshop, and ESSET Gerzensee. This project was first presented at the RUD (2008) conference in Oxford under the title "Increasing Interdependence of Multivariate Distributions", and a first draft was circulated in February 2010, entitled "The Supermodular Stochastic Ordering." Strulovici gratefully acknowledges financial support from the NSF (Grant No. 1151410) and the Alfred P. Sloan Foundation. Meyer: Nuffield College and Dept. of Economics, Oxford University, Oxford OX1 1NF, UK, and CEPR, margaret.meyer@nuffield.ox.ac.uk; Strulovici: Dept. of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA, b-strulovici@northwestern.edu.

return of many assets. Interdependence also affects welfare measures based on multiple indicators like health, education and income, as well as the assessment of inequality in populations subject to individual income risk. Comparisons of the quality of noisy matching procedures and comparisons of the alignment of preferences among members of a group can also be viewed as comparisons of the interdependence of random variables.

Unfortunately, common measures of interdependence, such as correlation, are often inadequate: just like variance does not capture risk well outside of Gaussian distributions or quadratic objective functions, correlation is too weak a concept to capture interdependence well, except under analogous restrictive assumptions. Moreover, correlation-based rankings can be reversed depending on how data is aggregated.[1] At the other extreme, rankings of interdependence based on the concepts of affiliation and association are also problematic, because they are too demanding.[2] Finally, while copulas are a useful tool for extracting information about the dependence between random variables, whatever their marginal distributions, one still must choose a method of ranking copulas.

This paper studies an ordering of interdependence based on *supermodular* objective functions. Supermodular functions are used to capture complementarity, a concept closely related to interdependence.[3] Indeed, a frequent reason why economists care about interdependence in the first place stems from the presence of complementarities. To see these links clearly, consider the case of two variables: a function $w$ is defined to be supermodular if $w(x', y') + w(x, y) \geq w(x, y') + w(x', y)$ whenever $x' \geq x$ and $y' \geq y$. Hence for a supermodular function, the effect of increasing two arguments together exceeds the sum of the effects of increasing each argument separately. If $w$ is supermodular, the expectation $E[w(X, Y)]$ with respect to random variables $X$ and $Y$ increases, holding marginal distributions fixed, as $X$ and $Y$ become more interdependent: raising the probability that both variables are high or both low raises $E[w(X, Y)]$ by more than the reduction from lowering the probability of one being high and the other low. We will say that one multivariate distribution displays greater interdependence than another in the sense of the supermodular ordering if the expectation of all supermodular functions is higher under the former distribution than under the latter.

Supermodularity is a pervasive property in economic analysis: it characterizes production functions in matching contexts (as well as in manufacturing systems), deprivation functions in the

[1] We provide an example in Section 2.

[2] These issues are discussed by Genest and Verret (2002) and Meyer and Strulovici (2012).

[3] See Milgrom and Roberts (1990, 1995) for this connection in the context of manufacturing. More recently, Dziewulski and Quah (2014) develop a test for the supermodularity of production functions. Their analysis builds on two characterizations from the present paper.

assessment of multidimensional inequality, ex post welfare functions in the presence of income risk, aggregate loss functions in finance and actuarial science, and estimator functions (such as OLS) in econometrics. Supermodular functions are also ubiquitous in comparative statics analysis. Section 5 explores in detail the application of the supermodular ordering to these areas of economics. These applications, in conjunction with our characterization results, demonstrate the power and tractability of the supermodular ordering as a tool for comparing the interdependence of random variables in a wide range of economic contexts.

Our first formal contribution is to characterize the supermodular ordering, for multivariate distributions with an arbitrary number of dimensions, in terms of *elementary transformations*. Theorem 1 shows that one multivariate distribution is dominated by another according to the supermodular ordering if and only if one can go from the former to the latter by a sequence of two-dimensional elementary transformations that increase the probability of homogeneous outcomes (those in which the two variables involved in the transformation are both lower or both higher) and reduce the probability of heterogeneous ones. Each transformation affects the probabilities of only four adjacent points (a "square") in the support of the distributions. Proposition 2 shows that our set of elementary transformations is minimal, in that any proper subset fails to characterize the supermodular ordering.[4] Our characterization theorem is new because it holds for random vectors of arbitrary dimension, and it involves a minimal set of transformations.[5]

These transformations are reminiscent of Rothschild and Stiglitz's (1970) mean-preserving spreads. Indeed, our transformations may be described as "marginal-preserving alignments". While the structure of our theorem characterizing the supermodular ordering in terms of marginal-preserving alignments parallels the structure of Rothschild and Stiglitz's theorem for the univariate

[4]This property is useful for the constructive methods described below.

[5]For the special case of bivariate distributions, Levy and Paroush (1974), Epstein and Tanny (1980), and Tchen (1980) have shown that the supermodular ordering is equivalent to the combination of upper- and lower-"orthant" dominance, and the latter two papers also provided a characterization in terms of elementary transformations. With three or more dimensions, the supermodular ordering is a strictly stronger ordering than the combination of upper- and lower-orthant dominance (see Joe (1990), Müller and Scarsini (2000), and Meyer and Strulovici (2012)). Promislow and Young (2005) study the cone of supermodular functions on lattices but do not provide any characterization like ours. Giovagnoli and Wynn (2008) briefly mention without proof a characterization result based on a much larger set of transformations than ours, involving more than two dimensions and non-adjacent points. In work subsequent to ours (and acknowledged as such), Müller (2013) contains a similar characterization: "The study of mass transfer principles as described above has recently found increasing interest in the economics literature in the context of comparing multivariate risks, see, e.g., [109, 318, 334]. Indeed, the basic principle that is used in this paper has already been used in [318, 334] for the special cases of supermodular ordering and inframodular ordering." Here, '318' refers to a 2011 version of the present paper entitled "The Supermodular Stochastic Ordering."

convex ordering, there is an important difference: our set of elementary transformations is minimal, whereas theirs is not.[6]

For many applications, the choice of a particular support is somewhat arbitrary. For example, when comparing multivariate empirical distributions of attributes such as income, health, and education (see Section 5.4), the distributions depend on the way the data for each attribute has been aggregated into discrete categories. One very appealing property of the supermodular ordering, that follows directly from Theorem 1, is that it is robust to coarsening of the support (aggregation), as well as to any weakly monotonic transformation of coordinates. For example, if one joint distribution of income and life expectancy dominates another according to the supermodular ordering when income is described by brackets of 1000 dollars and life expectancy by brackets of five years, supermodular dominance continues to hold if brackets of 5,000 dollars and ten years are used, and also holds if the coarsening is uneven, so that, for example, brackets become wider near the top of the distribution.[7] In contrast to the supermodular ordering, the ranking of bivariate distributions according to the linear correlation coefficient is not robust to weakly monotonic transformations of coordinates and, a fortiori, not robust to coarsening of the support.

A potential explanation for the popularity of correlation as a measure of interdependence, despite its lack of economically-grounded justification and its non-robustness to coarsening, is its ease of use. Fortunately, the supermodular ordering is also easy to work with, and we provide a set of methods to show this. These methods come from operations research; to our knowledge, they have not been employed to characterize stochastic orderings. First, we provide an algorithm to test, given a specific pair of distributions, whether one supermodularly dominates the other, by testing for the existence of a sequence of elementary transformations taking one distribution to the other: our characterization says that supermodular dominance holds if and only if such a sequence exists. The algorithm can easily be formulated as a linear program. The procedure is similar to the one used to check the existence of a solution to Afriat inequalities to test whether consumption choice data is consistent with utility maximization, and it is equally easy to implement. Second, we develop an algorithm that, for any given finite lattice, generates a *minimal* set of inequalities characterizing

[6]For the support $\{0, 1, 2, 3\}$, consider the mean-preserving spread (MPS) which adds probability mass $\epsilon$ to outcomes 0 and 3 and removes mass $\epsilon$ from outcomes 1 and 2. This MPS can be decomposed into the sum of two MPSs, the first (resp., second) of which adds mass $\epsilon$ to outcomes 0 and 2 (resp., 1 and 3) and removes mass $2\epsilon$ from outcome 1 (resp., 2). For this support, these two MPSs constitute a minimal set.

[7]Uneven coarsenings of data are common and clearly relevant: since the exact thousand-dollar value of an income above a million dollars, say, is unimportant, violations of interdependence based on such detail seem equally unimportant.

the supermodular ordering for *all* pairs of distributions on that lattice. This algorithm cannot be reduced to a linear program. It is based on the "double description" method, which is used to switch between alternative representations of polyhedral cones and which has recently been implemented in standard computer languages. It is worth emphasizing that our *minimal* characterization of the supermodular ordering, mentioned earlier, has the practical benefit of greatly simplifying the complexity of our algorithms. While it is beyond the scope of the present paper to provide an explicit application of our methods to empirical data, we do provide in the Appendix the code for our characterization algorithm based on the double description method and an illustration of its use.

In many applications – for example, in welfare economics – the objective function is not only supermodular but also increasing in its arguments. Theorem 3 demonstrates that the comparison of two distributions according to the *increasing* supermodular ordering can be decomposed into two steps: First compare marginals according to first-order stochastic dominance, and then compare the joint distributions, *after* correcting to ensure identical marginals, according to supermodular dominance. It is also often natural to focus on symmetric objective functions – welfare economics is again a good example, since symmetry captures a form of anonymity across citizens. Fortunately, imposing this restriction on objective functions adds no complication to the analysis: Proposition 3 shows 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.

We also provide sufficient conditions for random vectors arising in natural economic environments to be ranked according to the supermodular ordering. Section 3 studies the class of *mixture distributions*, representing random vectors generated by both common and idiosyncratic shocks. We model such distributions as follows: First, a common shock determines, for each random variable, the probability distribution from which it will be drawn. Then, each of the random variables is drawn independently from the distribution determined by the realization of the common shock. The resulting multivariate distribution is a mixture of conditionally independent random variables. 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). Intuitively, for mixture distributions, the "more important" the common shock relative to idiosyncratic shocks, the "more interdependent" the random variables should be. While in simple parameterized settings, it is easy to formalize and confirm this intuition, two questions arise when considering more general settings. First, how can "greater relative importance" of the common shock be formalized? Second, how can greater interdependence of the resulting conditionally independent variables be assessed? Our Theorem 4 answers both questions. We use the supermodular ordering to compare interdependence, and we present easily checkable sufficient conditions on the structure of mixture distributions for two such distributions to be comparable according to the supermodular ordering. Our sufficient conditions thus provide a useful non-parametric ordering of the relative importance of common vs. idiosyncratic shocks for mixture distributions.

Section 4 studies the class of $n$-dimensional random vectors representing $n$ independent lotteries and focuses on the case where the objective function defined on the outcomes of these lotteries is symmetric, as with an ex post welfare function. Theorem 5 provides sufficient conditions for symmetric supermodular dominance for random vectors within this class. These conditions capture the idea that one set of lotteries is less heterogeneous than another, holding fixed the average of the lotteries. In fact, at a mathematical level these conditions take the same form as the conditions in Theorem 4 capturing the idea of "lower relative importance" of the common shock. Moreover, given Proposition 3, the conditions identified in Theorem 5 are also sufficient conditions for comparing the degree of negative interdependence in symmetrized versions of independent distributions.[8]

Section 5 develops a range of economic applications of the supermodular ordering. Section 5.1 shows how Theorem 5 can be applied to make comparisons of ex post inequality in the presence of uncertainty. Section 5.2 uses the ordering to analyze how changes in the degree of alignment of the preferences of committee members affect equilibrium search and voting behavior. Section 5.3 applies the symmetric supermodular ordering to examine how systemic risk in banking networks depends on the structure of the interconnections among banks. More briefly, Section 5.4 discusses the role of supermodular objective functions in comparisons of i) multidimensional distributions of economic status; ii) the "richness" of datasets for prediction and parameter estimation; and iii) the efficiency of two-sided or many-sided matching mechanisms in the presence of frictions.

[8] The more heterogeneous are a set of independent lotteries ordered by first-order stochastic dominance, holding fixed their average, the more negatively interdependent is the symmetrized version of their joint distribution.

## 2 Setting and Characterization Results

For any fixed $n$, we consider distributions over a finite, $n$-dimensional lattice $\mathcal{L}$ constructed as follows. The $i^{th}$ variable takes values in a totally ordered set $(\mathcal{L}_i, \leq_i)$ with $m_i < \infty$ elements. $\mathcal{L}$ is defined as the Cartesian product $\times_i \mathcal{L}_i$, endowed with the usual partial order: $x \leq y$ if and only if $x_i \leq_i y_i$ for all $i \in \mathcal{N} \equiv \{1, \ldots, n\}$. Each $\mathcal{L}_i$ is order-isomorphic to a finite subset of $\mathbb{R}$ and the reader may without loss think of $\mathcal{L}$ as a (possibly uneven) finite lattice built on a hyperrectangle of $\mathbb{R}^n$. For any $x \in \mathcal{L}$, let $x + e_i$ denote the element $y$ of $\mathcal{L}$, whenever it exists, such that $y_j = x_j$ for all $j \in \mathcal{N} \setminus \{i\}$ and $y_i$ is the smallest element of $\mathcal{L}_i$ greater than but not equal to $x_i$. For example, if $\mathcal{L} = \{0, 1\}^2$, $(0, 0) + e_1 = (1, 0)$ and $(1, 0) + e_2 = (0, 0) + e_1 + e_2 = (1, 1)$.

**Vectorial structure.** Labeling arbitrarily the $d = \prod_{i=1}^n m_i$ elements (or "nodes") of $\mathcal{L}$, one may view each real-valued function defined on $\mathcal{L}$ as a vector of $\mathbb{R}^d$, where each coordinate is the value of the function evaluated at a specific node of $\mathcal{L}$. In particular, a multivariate distribution whose support is contained in $\mathcal{L}$ may be represented as an element of the unit simplex $\Delta_d$ of $\mathbb{R}^d$.

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

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

To any class $\mathcal{W}$ of functions on $\mathcal{L}$ corresponds an ordering of multivariate distributions:

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

We will be particularly interested in the orderings generated by supermodular, increasing supermodular, and symmetric supermodular objective functions.

**Supermodular functions and elementary transformations.** For any $x, y \in \mathcal{L}$, let $x \wedge y$ and $x \vee y$ respectively denote the component-wise minimum (or "meet") and component-wise maximum (or "join") of $x$ and $y$.[9] A function $w$ is *supermodular* (on $\mathcal{L}$) if $w(x \wedge y) + w(x \vee y) \geq w(x) + w(y)$ for all $x, y \in \mathcal{L}$, and *submodular* if $-w$ is supermodular. Let $\mathcal{S}$ denote the set of supermodular functions. The **supermodular ordering** (denoted $\prec_{\mathcal{S}PM}$) is the ordering defined by (1) for the class $\mathcal{S}$. For random vectors $X$ and $Y$ with distributions $f$ and $g$ and cumulative distributions $F$ and $G$, respectively, we will use the expressions $X \prec_{\mathcal{S}PM} Y$, $f \prec_{\mathcal{S}PM} g$, and $F \prec_{\mathcal{S}PM} G$ interchangeably.

To characterize the supermodular ordering, we introduce a class of elementary transformations capturing "increasing interdependence". For any $x \in \mathcal{L}$ such that $x + e_i + e_j \in \mathcal{L}$, let $t^x_{i,j}$ denote

[9] Explicitly, $(x \wedge y)_i = \min\{x_i, y_i\}$ and $(x \vee y)_i = \max\{x_i, y_i\}$ for all $i \in \mathcal{N}$.

the function defined on $\mathcal{L}$ by

$$t_{i,j}^x(x) = t_{i,j}^x(x + e_i + e_j) = 1, \qquad t_{i,j}^x(x + e_i) = t_{i,j}^x(x + e_j) = -1, \tag{2}$$

and $t_{i,j}^x(y) = 0$ for all other $y \in \mathcal{L}$. We call $t_{i,j}^x$ an *elementary transformation* on $\mathcal{L}$, and let $\mathcal{T}$ denote the set of all elementary transformations.

If distributions $f$ and $g$ are such that $g = f + \alpha t_{i,j}^x$ for some $\alpha \geq 0$, then we say that $g$ is obtained from $f$ by an elementary transformation with weight $\alpha$. The $\alpha$-weighted elementary transformation raises the probability of nodes $x$ and $x + e_i + e_j$ by the common amount $\alpha$, reduces the probability of nodes $x + e_i$ and $x + e_j$ by the same amount, and leaves unchanged the probability of all other nodes in $\mathcal{L}$. 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. Thus, our elementary transformations could alternatively be described as "marginal-preserving alignments".

To illustrate, consider the $3\times3$ lattice $\mathcal{L} = \{0, 1, 2\}^2$. There are four elementary transformations, corresponding to $x = (0, 0)$, $(1, 0)$, $(0, 1)$, and $(1, 1)$. For the $2 \times 2 \times 2$ lattice $\mathcal{L} = \{0, 1\}^3$, there are six elementary transformations, one corresponding to each face of the unit cube. Note that each elementary transformation affects only *two* of the $n$ dimensions (as illustrated by the example of $\mathcal{L} = \{0, 1\}^3$), and it affects values only at four *adjacent* points in the lattice, $x$, $x + e_i$, $x + e_j$, and $x + e_i + e_j$ (as illustrated by $\mathcal{L} = \{0, 1, 2\}^2$).[10]

**Theorem 1 (Characterization of the 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* $\mathbb{R}^d$*,*

$$g = f + \sum_{t\in\mathcal{T}} \alpha_t t. \tag{3}$$

Since any elementary transformation $t \in \mathcal{T}$ leaves the marginal distributions unchanged, Theorem 1 implies that $f$ and $g$ must have identical marginal distributions whenever $f \prec_{SPM} g$. In addition, the theorem also implies that distributions that are comparable according to the supermodular ordering are essentially characterized by their covariance matrix, in the following sense.

[10]Our marginal-preserving alignments are broadly analogous to Rothschild and Stiglitz's (1970) mean-preserving spreads. However, as defined by Rothschild and Stiglitz, a mean-preserving spread can alter the probabilities of four arbitrarily distant points. With a discrete support, the analog to our restriction that elementary transformations affect only *two* dimensions and *adjacent* points would be the restriction that mean-preserving spreads affect the probabilities of only *three adjacent* points.

**Proposition 1** *Given random vectors $X$ and $Y$ with distributions $f$ and $g$, respectively, if $f \prec_{SPM} g$ and, for all $i \neq j$, $Cov(X_i, X_j) = Cov(Y_i, Y_j)$, then $f = g$, that is, $X$ and $Y$ are identically distributed.*

For many applications, the choice of a particular support is somewhat arbitrary. For example, when comparing multivariate empirical distributions of attributes such as income, health, and education (see Section 5.4), the distributions depend on the way the data for each attribute has been aggregated into discrete categories. One very appealing property of the supermodular ordering, that follows directly from Theorem 1, is that it is robust to coarsening of the support (aggregation), as well as to any weakly monotonic transformation of coordinates.[11] To see this, suppose that each point $(x_1, \ldots, x_n) \in \mathcal{L}$ is transformed to $(r_1(x_1), \ldots, r_n(x_n))$, for some set of nondecreasing functions $\{r_i\}_i$, and denote the transformed support by $\mathcal{L}^r$. In the special case where all $\{r_i\}_i$ are strictly increasing, there is a one-to-one mapping between elementary transformations on $\mathcal{L}$ and elementary transformations on $\mathcal{L}^r$; if instead $x$ and $x + e_i$ (or $x$ and $x + e_j$) are transformed into the same point in $\mathcal{L}^r$, then the transformation $t^x_{i,j}$ is mapped into the zero function on $\mathcal{L}^r$. Hence, it follows from Theorem 1 that for $\{r_i\}_i$ nondecreasing, $(X_1, \ldots, X_n) \prec_{SPM} (Y_1, \ldots, Y_n)$ implies $(r_1(X_1), \ldots, r_n(X_n)) \prec_{SPM} (r_1(Y_1), \ldots, r_n(Y_n))$. Moreover, for $\{r_i\}_i$ strictly increasing, the reverse implication holds as well.[12]

## 2.1 Constructive Methods for Applying the Characterization in Theorem 1

Two aspects of our approach greatly facilitate the use of Theorem 1 to determine, given a pair of distributions $f$ and $g$, whether $f \prec_{SPM} g$. The first is our restriction to a *finite* support $\mathcal{L}$.[13] The second is our restriction that elementary transformations, defined in (2), affect only two of the $n$ dimensions and affect values at only adjacent points in the lattice. These two restrictions make it straightforward, either manually or algorithmically, to list the entire set $\mathcal{T}$ of elementary

[11] In contrast, the ranking of bivariate distributions according to the linear correlation coefficient is not robust to weakly monotonic transformations of coordinates and, a fortiori, not robust to coarsening of the support. To see this, for $L = \{l, m, h\}^2$, where $l < m < h$, let $(Y_1, Y_2)$ have distribution $g$, where $g(l, m) = g(m, l) = g(h, h) = \frac{1}{3}$, and let $(X_1, X_2)$ have distribution $f$, where $f(l, l) = f(m, h) = f(h, m) = \frac{1}{3}$. Then $corr(Y_1, Y_2) > (<)\, corr(X_1, X_2)$ if $\frac{(l+h)}{2} > (<)\, m$. This in turn implies that if $L$ is coarsened by combining the realizations $l$ and $m$ in each dimension, then $corr(Y_1, Y_2) > corr(X_1, X_2)$, while if instead $m$ and $h$ are combined, then $corr(Y_1, Y_2) < corr(X_1, X_2)$.

[12] An alternative proof of the first implication uses the fact that for $w(x_1, \ldots, x_n)$ supermodular and $\{r_i\}_i$ nondecreasing, $w(r_1(x_1), \ldots, r_n(x_n))$ is also supermodular. See Shaked and Shanthikumar (1997, Theorem 2.2).

[13] Theorem 6 (Section 6.1) may also be used, in conjunction with Theorem 1, to compare distributions on a continuous support using our techniques, as long as the distributions have a continuous density.

transformations on any given $\mathcal{L}$. In fact, our set of elementary transformations is minimal, in the following sense:

**Proposition 2** *All elements of $\mathcal{T}$ are extreme rays of the convex cone $\mathcal{C}(\mathcal{T})$ generated by $\mathcal{T}$.*

This proposition says that dropping any elementary transformation from our set destroys the characterization provided by Theorem 1.

**Comparing Two Given Distributions**

From Theorem 1, $f \prec_{\mathcal{SPM}} g$ if and only there exist nonnegative coefficients $\{\alpha_t\}_{t\in\mathcal{T}}$ such that $g - f = \sum_{t\in\mathcal{T}} \alpha_t t$. For a given 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 $T = |\mathcal{T}|$ denote the number of elementary transformations on $\mathcal{L}$, and let $E$ denote the $d \times T$-matrix whose columns are the $d$-dimensional vectors consisting of all elementary transformations of $\mathcal{L}$. Theorem 1 can be re-expressed as follows: $f \prec_{\mathcal{SPM}} g$ if and only if there exists $\alpha \in \mathbb{R}^T$ nonnegative such that $E\alpha = g - f$. Let $\delta^+$ denote the vector of $\mathbb{R}^d$ whose $i^{th}$ component equals $|(g-f)_i|$ and $E^+$ 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 linear program:[14]

$$\min_{(\alpha,\beta)\in\mathbb{R}^T\times\mathbb{R}^d} \sum_{i=1}^{d} \beta_i \qquad \text{subject to} \qquad E^+\alpha + \beta = \delta^+, \quad \alpha \geq 0, \quad \beta \geq 0. \tag{4}$$

**Theorem 2 (Pairwise Comparison)** *The linear program (4) 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 $(\alpha^*, \beta^* = 0)$ is any solution of (4).*

**Explicit Characterization of the Supermodular Ordering via a Set of Inequalities**

To compare many distributions, for example as part of a larger optimization problem, it is convenient to generate once and for all, for a given support, an explicit characterization of the supermodular ordering. Given any finite support $\mathcal{L}$, we present a method for generating such a representation in the form of a finite list of inequalities that are satisfied by the vector $g - f$ if and only if $f \prec_{\mathcal{SPM}} g$.

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

Recall that, by definition, $f \prec_{SPM} g$ if $g - f$ makes a nonnegative scalar product with all supermodular functions on $\mathcal{L}$, seen as vectors of $\mathbb{R}^d$. This condition can be reduced to a finite set of linear inequalities by exploiting the geometric properties of $\mathcal{S}$. $\mathcal{S}$ is a convex cone such that $w$ is supermodular (i.e., belongs to $\mathcal{S}$) if and only if it makes a nonnegative scalar product with each of the $T$ elementary transformations on $\mathcal{L}$ as defined by (2). In matrix form, $\mathcal{S} = \{w \in \mathbb{R}^d : Aw \geq 0\}$, where $A = E'$ is the $T \times d$ matrix whose rows consist of all elementary transformations. Since $Aw \geq 0$ describes a finite set of linear inequalities, $\mathcal{S}$ is a polyhedral cone, and $A$ is called the *representation matrix* of $\mathcal{S}$. The Minkowski-Weyl Theorem (Ziegler, 1997) states that a cone is polyhedral if and only if it has a finite number of extreme rays. In our context, this theorem implies that to any $T \times d$ representation matrix $A$ corresponds a *generating matrix* $R$, with $d$ rows and a finite number of columns, such that

$$Aw \geq 0 \quad \Longleftrightarrow \quad w = R\lambda \quad \text{for some vector } \lambda \geq 0.$$

The columns of the matrix $R$ are the finite set of extreme rays of the cone $\mathcal{S}$. The stochastic supermodular ordering is thus entirely determined by the extreme rays of $\mathcal{S}$, in that

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

The Minkowski-Weyl Theorem thus proves the existence, for any finite support $\mathcal{L}$, of a finite list of inequalities, one corresponding to each extreme ray of $\mathcal{S}$, that entirely characterize the supermodular ordering on $\mathcal{L}$.

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), provides an algorithm to go back and forth between the descriptions of a polyhedral cone in terms of its representation matrix $A$ and its generating matrix $R$. In our context, the representation matrix $A$, determined by the set of elementary transformations defined by (2), is straightforward to compute and generate in a program, for any support $\mathcal{L}$. By applying Fukuda's implementation of the double description method, we have developed an algorithm that provides, given any $\mathcal{L}$, the list of inequalities, one corresponding to each extreme supermodular function, that characterize the supermodular ordering on $\mathcal{L}$. We have computed these inequalities for a range of problems that are intractable by hand. In the Appendix, we provide the code for our algorithm and, for illustration, the set of inequalities that it yields when $\mathcal{L} = \{0, 1\}^4$.

The fact that our set of elementary transformations is minimal has the practical benefit of greatly simplifying the complexity of our algorithm. The complexity can be further reduced by aggregating data into coarser categories (coarsening the support), and as discussed above, aggregation of data (coarsening) preserves the supermodular ordering. Thus, with an appropriate degree of coarsening, the double description method can be used to achieve a tractable comparison of distributions according to the supermodular ordering.

## 2.2 The Increasing Supermodular Ordering

In many economic settings, we want to compare multivariate distributions not just with respect to interdependence but also with respect to the levels of the random variables. A function $w$ on $\mathcal{L}$ is *increasing* if for any $x \in \mathcal{L}$ and $i$ such that $x + e_i \in \mathcal{L}$, $w(x + e_i) \geq w(x)$. Let $\mathcal{I}$ denote the set of increasing functions on $\mathcal{L}$. For any $x \in \mathcal{L}$ and $i$ such that $x + e_i \in \mathcal{L}$, let $\tau_i^x$ denote the function on $\mathcal{L}$ such that $\tau_i^x(x) = -1$, $\tau_i^x(x + e_i) = 1$, and $\tau_i^x$ vanishes everywhere else. Let $\mathcal{U}$ denote the set of all such functions. One may easily check that $w$ belongs to $\mathcal{I}$ if and only if $w \cdot \tau \geq 0$ for all $\tau \in \mathcal{U}$. First-order stochastic dominance for distributions on $\mathcal{L}$ is defined by

$$f \prec_{FOSD} g \Longleftrightarrow w \cdot f \leq w \cdot g \qquad \forall w \in \mathcal{I}. \tag{5}$$

It is easy to adapt the proof of Theorem 1 to show that $f \prec_{FOSD} g$ if and only if there exist nonnegative coefficients $\{\beta_\tau\}_{\tau \in \mathcal{U}}$ such that

$$g = f + \sum_{\tau \in \mathcal{U}} \beta_\tau \tau. \tag{6}$$

The **increasing supermodular ordering** (denoted $\prec_{\mathcal{I}SPM}$) is defined as follows:

$$f \prec_{\mathcal{I}SPM} g \Longleftrightarrow w \cdot f \leq w \cdot g \qquad \forall w \in \mathcal{S} \cap \mathcal{I}.$$

In contrast to $f \prec_{\mathcal{S}PM} g$, $f \prec_{\mathcal{I}SPM} g$ does not imply that $f$ and $g$ have identical marginals. Rather, $f \prec_{\mathcal{I}SPM} g$ implies that each marginal distribution of $f$ is dominated by the corresponding marginal distribution of $g$ according to first-order stochastic dominance: this can be seen by taking, for each $i \in \mathcal{N}$ and each $k_i \in \mathcal{L}_i$, $w(z) = I_{\{z_i \geq k_i\}}$, which is both increasing and supermodular.

Theorem 3 below demonstrates that comparison of two distributions according to the increasing supermodular ordering can be decomposed into a two-step comparison, first comparing the marginals according to first-order stochastic dominance and then comparing the joint distributions, *after* correcting to ensure identical marginals, according to supermodular dominance.

To simplify notation, assume that $\mathcal{L}_i = \{0, 1, \ldots, m_i - 1\}$ (as explained just before Section 2.1, this labeling of values is without loss of generality). Given two distributions $f$ and $g$ with $\delta \equiv g - f$, define the function $\gamma$ on $\mathcal{L}$, to correct for differences in the marginals of $f$ and $g$, as follows. Let

$\gamma(z)$ vanish everywhere except on the set $\mathcal{L}_0$ of $z$'s that have at most one positive component, and for any $i \in \mathcal{N}$ and $k \in \{1, 2, \ldots, m_{i-1}\}$, let

$$\gamma(ke_i) = Pr(Y_i = k) - Pr(X_i = k) = \sum_{z:z_i=k} \delta(z). \tag{7}$$

Finally, let $\gamma(0, 0, \ldots, 0)$ be such that $\sum_{z\in\mathcal{L}_0} \gamma(z) = 0$. Since $\sum_{z\in\mathcal{L}} \delta(z) = \sum_{z\in\mathcal{L}}(g(z) - f(z)) = 0$, it follows from (7) that for all $i$ and $k$, including $k = 0$,

$$\sum_{z:z_i=k} \gamma(z) = \sum_{z:z_i=k} \delta(z). \tag{8}$$

Equation (8) ensures that $f + \gamma$ has the same marginal distributions as $g$, so $f + \gamma$ and $g$ can potentially be compared according to $\prec_{\mathcal{SPM}}$.[15] At the same time, $\gamma$ contains all the information needed to determine whether the marginals of $g$ first-order stochastically dominate the marginals of $f$.

**Theorem 3 (Increasing Supermodular Ordering)** *The following statements are equivalent:*

*1)* $f \prec_{\mathcal{ISPM}} g$.

*2) There exist nonnegative coefficients* $\{\alpha_t\}_{t\in\mathcal{T}}$, $\{\beta_\tau\}_{\tau\in\mathcal{U}}$ *such that*

*a)* $\gamma = \sum_{\tau\in\mathcal{U}} \beta_\tau \tau$, *and*

*b)* $g = f + \gamma + \sum_{t\in\mathcal{T}} \alpha_t t$.

*3) For each* $i$, *the* $i^{th}$ *marginal distribution of* $f$ *is dominated by the* $i^{th}$ *marginal distribution of* $g$ *according to first-order stochastic dominance, and for all supermodular* $w$, $w\cdot(f+\gamma) \leq w\cdot g$.

It follows from Theorem 3 that when comparisons are restricted to pairs of distributions with identical marginals, the increasing supermodular ordering and the supermodular ordering are equivalent.

### 2.3 The Symmetric Supermodular Ordering

Symmetric objective functions play an important role in many economic applications. For example, if the objective function is an ex post welfare function, imposing symmetry amounts to assuming ex post anonymity across individuals. In finance and insurance contexts, losses may be evaluated

[15]Strictly speaking, we are assessing whether for all supermodular $w$, $w \cdot g \geq w \cdot (f + \gamma)$; this way of expressing greater interdependence in $g$ than in $f + \gamma$ is valid whether or not all elements of the vector $f + \gamma$ lie in $[0, 1]$.

according to a convex function of the total loss across all assets or all insurance policies. Any convex function of the sum of losses is both symmetric and supermodular in the individual losses. The symmetric supermodular ordering, characterized in this section, will be used in the applications developed in Sections 4, 5.1, and 5.3.

A lattice $\mathcal{L} = \times_{i=1}^{n} \mathcal{L}_i$ is symmetric if $\mathcal{L}_i = \mathcal{L}_j$ for all $i \neq j$. A real-valued function $f$ on a symmetric lattice $\mathcal{L}$ is *symmetric on* $\mathcal{L}$ if $f(x) = f(\sigma(x))$ for all $x \in \mathcal{L}$ and permutations $\sigma$.

Given two distributions $g$ and $f$ on a symmetric lattice $\mathcal{L}$, $g$ dominates $f$ according to the **symmetric supermodular ordering**, written $f \prec_{SSPM} g$, if and only if $w \cdot f \leq w \cdot g$ for all symmetric supermodular functions $w$ on $\mathcal{L}$. For any function $f$ defined on a symmetric lattice $\mathcal{L}$, the *symmetrized version of* $f$, denoted $f^{symm}$, is defined by

$$f^{symm}(x) = \frac{1}{n!} \sum_{\sigma \in \Sigma(n)} f(\sigma(x)), \tag{9}$$

where $\Sigma(n)$ is the set of all permutations of $\mathcal{N}$. If $w$ is a supermodular function, then $w^{symm}$ is supermodular. The following result, proved in Meyer and Strulovici (2012, Section 2.3), shows that one can characterize the symmetric supermodular ordering in terms of the supermodular order applied to symmetrized distributions.

**Proposition 3** *Given distributions $f, g$ defined on a symmetric lattice, $f \prec_{SSPM} g$ if and only if $f^{symm} \prec_{SPM} g^{symm}$.*

To go further, we simplify notation once again by relabeling the points in the support so that $\mathcal{L} = \{0, 1, \ldots, m-1\}^n$. For $x \in \mathcal{L}$ and $k \in \{1, \ldots, m-1\}$, define $\bar{c}^k(x) = \sum_{i=1}^{n} I_{\{x_i \geq k\}}$ and $\bar{c}(x) = (\bar{c}^1(x), \ldots, \bar{c}^{m-1}(x))$. $\bar{c}^k(x)$ 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 $\tilde{L}^{m-1}$, an $(m-1)$-dimensional subset of $\{0, 1, \ldots, n\}^{m-1}$. Since all permutations of $x \in \mathcal{L}$ correspond to the same vector $\bar{c}(x)$, it follows that $w$ is symmetric if and only if it can be written as $w(x) = \phi(\bar{c}(x))$, for some $\phi$ defined on $\tilde{L}^{m-1}$.

The result below shows that for any number of dimensions $n$, the symmetric supermodular ordering of random vectors $X$ and $Y$ on $\mathcal{L}$ is equivalent to an ordering of the derived random vectors $\bar{c}(X)$ and $\bar{c}(Y)$ on $\tilde{L}^{m-1}$. To state the result, we need the following definition. A function $\phi$ on $\tilde{L}^{m-1}$ is *componentwise-convex* if for any $y \in \tilde{L}^{m-1}$ and $k = \{1, 2, \ldots, m-1\}$ such that $y + 2e_k \in \tilde{L}^{m-1}$, $\phi(y) + \phi(y + 2e_k) \geq 2\phi(y + e_k)$.

**Proposition 4** *For random vectors $X$ and $Y$ distributed on $\mathcal{L} = \{0, 1, \ldots, m-1\}^n$, $X \prec_{SSPM} Y$ if and only if $E\phi(\bar{c}(X)) \leq E\phi(\bar{c}(Y))$ for all supermodular and componentwise-convex functions $\phi$*

*defined on $\tilde{L}^{m-1}$. In the special case where $m = 2$, $X \prec_{SSPM} Y$ if and only if $E\phi(\sum_{i=1}^n I_{\{X_i=1\}}) \leq E\phi(\sum_{i=1}^n I_{\{Y_i=1\}})$ for all convex functions $\phi$ defined on $\{0, 1, \ldots, n\}$.*

In the special case $m = 2$, each component of the random vectors $X$ and $Y$ has a binary support $\{0, 1\}$. In this case, whatever the dimension of $X$ and $Y$, Proposition 4 shows that comparison of $X$ and $Y$ according to the symmetric supermodular ordering reduces to comparison of $\sum_{i=1}^n I_{\{X_i=1\}}$ and $\sum_{i=1}^n I_{\{Y_i=1\}}$ according to the well-understood univariate convex ordering, which is equivalent to the ordering of greater riskiness studied by Rothschild and Stiglitz (1970).

More generally, Proposition 4 is useful because, even as the dimension $n$ of the underlying random vectors $X$ and $Y$ increases, the dimension of the derived random vectors $\bar{c}(X)$ and $\bar{c}(Y)$ remains fixed at $m - 1$. We will use this proposition in Section 5.3, where we apply the symmetric supermodular ordering to compare systemic risk for different networks of financial linkages across banks.

# 3 Aggregate vs. Idiosyncratic Shocks

In economics, particularly macroeconomics and finance, the interdependence of random variables often arises from the presence of aggregate shocks or common factors. This section focuses on mixture distributions, representing random vectors generated by both aggregate and idiosyncratic shocks, and provides non-parametric sufficient conditions for one such random vector to display more interdependence, in the sense of the supermodular ordering, than another. The following example will help to motivate our analysis.

**Example 1** Let the random vector $X$ be such that $X_i = \theta + \varepsilon_i$, where $\theta$ and $\{\varepsilon_i\}_{i\in\mathcal{N}}$ are all independent and have binomial distributions $B(\eta_\theta, p)$ and $B(\eta_\varepsilon, p)$, respectively, with $p \in (0, 1)$ and $\eta_\varepsilon = \eta - \eta_\theta$. An increase in $\eta_\theta$ raises each pairwise covariance $Cov(X_i, X_j)$ while leaving the marginal distribution of each $X_i$ unchanged. Theorem 4 below can be used to show, for $p$ rational, that raising the importance of the common shock $\theta$ by increasing $\eta_\theta$ makes the random variables $(X_1, \ldots, X_n)$ more supermodularly dependent. In addition, if we set $\eta_\theta = \lambda_\theta/p$, $\eta_\varepsilon = \lambda_\varepsilon/p$ and let $p$ go to zero while holding $\lambda_\theta$ and $\lambda_\varepsilon$ fixed, the limiting distributions of $\theta$ and $\{\varepsilon_i\}_{i\in\mathcal{N}}$ are Poisson with parameters $\lambda_\theta$ and $\lambda_\varepsilon$, respectively. Hence this example also implies that for Poisson distributed random variables $X_i = \theta + \varepsilon_i$, when $\lambda_\theta$ increases, holding $\lambda_\theta + \lambda_\varepsilon$ (and hence the marginal distribution of each $X_i$) fixed, $(X_1, \ldots, X_n)$ become more supermodularly dependent.

A similar result was known for $X_i = \theta + \varepsilon_i$ when $\theta$ and $\varepsilon$ are normally distributed: increasing the variance of $\theta$, while leaving the variance of each $X_i$ unchanged, makes the random vector $X$ more supermodularly dependent.[16] This result can also be recovered from our example, using the fact that normal distributions may be obtained as the limit of binomial distributions. However, even with the additive structure $X_i = \theta + \varepsilon_i$, for *arbitrary* distributions an increase in each $Cov(X_i, X_j)$ does not generally make $X$ more supermodularly dependent.[17]

The process generating a random vector as a mixture distribution can be decomposed into two steps: first, the realization of the "aggregate" shock selects, for each component of the random vector, one distribution out of many possible ones; second, for each component, independently, an outcome is drawn from the distribution randomly selected.[18] We consider all mixture distributions, with the restriction that the outcomes of each step can take finitely many values and the mixture weights (in the distribution of the aggregate shock) are rational. With this restriction, we can represent any random vector with a mixture distribution in the following manner.

To each variable $X_r$, $r \in \mathcal{N}$, is associated a $q \times m_r$ row-stochastic matrix $A(r)$, where each row of $A(r)$ represents a probability distribution for the variable $X_r$ on some finite support with $m_r$ values. The vector $(X_1, \ldots, X_n)$ is constructed as follows. First, a row index $i \in \{1, \ldots, q\}$ is drawn randomly, according to a uniform distribution over the $q$ possible values. This step represents the realization of the aggregate shock.[19] Then, each variable $X_r$ is independently drawn from the distribution described by the $i^{th}$ row of $A(r)$. This step represents the realization of the idiosyncratic shocks. The unconditional marginal distribution of each $X_r$ is described by the (equally-weighted) average of the rows of $A(r)$. Without loss of generality, we take the support of each random variable $X_r$ to be $\{1, \ldots, m_r\}$.

For the representation of mixture distributions described above, greater importance of the aggregate shock relative to the idiosyncratic shocks should correspond, for each matrix $A(r)$, to the rows being more different from one another, holding the average of the rows of each $A(r)$, and hence the unconditional distribution of each $X_r$, fixed.

[16] This follows from Müller and Scarsini (2000)'s result that for random vectors $Y$ and $Z$ with multivariate normal distributions, the condition $Cov(Y_i, Y_j) \leq Cov(Z_i, Z_j)$ for all $i \neq j$, coupled with identical marginal distributions, implies that $Y \prec_{SPM} Z$.

[17] See the example in Section A.1 of the Appendix.

[18] In the statistics literature, the distributions described below are often referred to as unidimensional latent variable models (Holland and Rosenbaum, 1986).

[19] The analysis can easily be extended to accommodate non-uniform distributions of the aggregate shock, by appropriate replications of the rows of the matrix, as long as these distributions have finite supports and rational weights.

The following terminology and notation will be useful to formalize this idea. A matrix $A$ is *row-stochastic* if each row represents a probability distribution. For any $q \times m$ matrix $A$, the entries of the (upper) *cumulative-sum matrix* $\bar{A}$ of $A$ are defined by $\bar{A}_{i,j} = \sum_{k=j}^{m} A_{i,k}$. Thus, $\bar{A}_{i,j}$ is decreasing in $j$. If $A$ is row-stochastic, the first column of $\bar{A}$ has all entries equal to 1. Clearly, there is a one-to-one mapping between row-stochastic matrices and their cumulative-sum equivalents.

A row-stochastic matrix $A$ is *stochastically ordered* if for each $k$, $\bar{A}_{i,k}$ is weakly increasing in $i$. This is equivalent to the requirement that for all $i \in \{2, \ldots, q\}$, the $i^{th}$ row of $A$ dominates the $(i-1)^{th}$ row in the sense of first-order stochastic dominance, so that high-index aggregate shocks are more likely to yield high outcomes for the variable $X$ generated by $A$. Given a row-stochastic matrix $A$, the *stochastically-ordered version* of $\bar{A}$, denoted $\bar{A}^{so}$, is the stochastically-ordered matrix obtained from $\bar{A}$ by reordering each of its columns from the smallest to the largest element. If $A$ is itself stochastically ordered, then $\bar{A}^{so} = \bar{A}$, and in this case we will use the expressions "$A$ is stochastically ordered" and "$\bar{A}$ is stochastically ordered" interchangeably.

Our ordering of matrices builds upon Hardy, Littlewood, and Polya's (1934, 1952) definition of majorization, which formalizes greater dispersion in the elements of a vector.

**Definition 1** *A vector* **a majorizes** *a vector* **b** *of identical dimension if i) the sums of the elements of* **a** *and* **b** *are equal, 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** *.*

We now present our ordering of matrices, which we term "cumulative column majorization", that formalizes the idea that the rows of a matrix $A$ are "more different" from one another than the rows of $B$ (holding the average of the rows fixed).

**Definition 2** *Given two row-stochastic matrices $A$ and $B$ of dimension $q \times m$, $A$ dominates $B$ according to the* **cumulative column majorization criterion***, denoted $A \succ_{CCM} B$, if for all $k \leq m$, the $k^{th}$ column vector of $\bar{A}$ majorizes the $k^{th}$ column vector of $\bar{B}$. Equivalently, $A \succ_{CCM} B$ if for for all $l \leq q$ and $k \leq m$, $\sum_{i=l}^{q} \bar{A}^{so}_{i,k} \geq \sum_{i=l}^{q} \bar{B}^{so}_{i,k}$, with equality holding for $l = 1$, for all $k \leq m$.*

Note that the definition of $A \succ_{CCM} B$ requires that $\bar{A}$ and $\bar{B}$ have equal column sums. Hence, if random variable $X$ is generated by matrix $A$ and random variable $Y$ by $B$, $A \succ_{CCM} B$ implies that the unconditional distributions of $X$ and $Y$ are identical.

The condition that $A \succ_{CCM} B$ says that, for each point in the support $\{1, \ldots, m\}$, the $q$-vector of upper cumulative probabilities corresponding to the $q$ possible conditional distributions (rows of the matrix) is more dispersed for matrix $A$ than for matrix $B$. Since the unconditional distributions of $X$ and $Y$ are the same, greater dispersion of the conditional distributions in $A$ than in $B$ implies that the aggregate shock is more important in the mixture distribution generated by $A$ than in that generated by $B$.

The main result of this section is the following theorem.

**Theorem 4** *Let $(A(1), \ldots, A(n))$ and $(B(1), \ldots, B(n))$ be two sets of row-stochastic matrices generating the random vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$, respectively, with for each $r \in \mathcal{N}$, $A(r)$ and $B(r)$ having dimension $q \times m_r$. Suppose that, i) for each $r \in \mathcal{N}$, $A(r)$ is stochastically ordered, and ii) for each $r \in \mathcal{N}$, $A(r) \succ_{CCM} B(r)$. Then $(X_1, \ldots, X_n) \succ_{\mathcal{S}PM} (Y_1, \ldots, Y_n)$.*

We have examples showing that the theorem does not hold if we drop either condition *i)* or *ii)*.[20] We conjecture that Theorem 4 can be extended to the case where the aggregate shock or the random vectors have continuous supports.[21]

The condition that for each $r$, $A(r) \succ_{CCM} B(r)$ says that the realization of the aggregate shock is relatively more informative about what the realizations of $\{X_r\}_{r \in \mathcal{N}}$ will be than about what the realizations of $\{Y_r\}_{r \in \mathcal{N}}$ will be. In the special case where the matrices $A(r)$ and $B(r)$ are both stochastically ordered, $A(r) \succ_{CCM} B(r)$ reduces to

$$\sum_{i=l}^{q} \sum_{j=k}^{m_r} A_{i,j}(r) = \sum_{i=l}^{q} \bar{A}_{i,k}(r) \geq \sum_{i=l}^{q} \bar{B}_{i,k}(r) = \sum_{i=l}^{q} \sum_{j=k}^{m_r} B_{i,j}(r) \quad \forall l \geq 2, k \geq 2, \tag{10}$$

coupled with the condition that $A(r)$ and $B(r)$ have matching column sums. Since higher values of $i$ correspond to higher realizations of the aggregate shock and higher values of $j$ to higher realizations

[20] Jogdeo (1978) showed that for any stochastically ordered row-stochastic matrices $\{A(r)\}$, the distribution of $(X_1, \ldots, X_n)$ generated from them displays association, a widely-used dependence concept defined in Esary, Proschan, and Walkup (1967). It follows from this and Theorem 2 of Meyer and Strulovici (2012) that the distribution of $(X_1, \ldots, X_n)$ 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 4 where for each $r$, the matrix $B(r)$ consists of $q$ identical rows.

[21] If sequences of random vectors $\{X_n\}$ and $\{Y_n\}$ satisfy $X_n \succ_{\mathcal{S}PM} Y_n$ for all $n$ and respectively converge in law to $X$ and $Y$, then $X \succ_{\mathcal{S}PM} Y$. To handle, say, an aggregate shock that was uniformly distributed on $[0, 1]$, the strategy would be to construct sequences of matrices $\{A(r)_n\}$ and $\{B(r)_n\}$, representing finer and finer discrete uniform distributions of the aggregate shock, and to apply Theorem 4 to the sequences of random vectors $\{X_n\}$ and $\{Y_n\}$ generated by these matrices. For the continuous analogues of the matrices $A(r)$ and $B(r)$, it is straightforward to define the continuous analogue of condition *i)* in Theorem 4, and the definition of cumulative column majorization can be replaced with a notion of cumulative column Lorenz dominance. One would then need to show that given these conditions on the continuous analogues of $A(r)$ and $B(r)$, each pair of discretizations $A(r)_n$ and $B(r)_n$ satisfies the conditions of Theorem 4.

of $X_r$, the condition in (10) can be read as saying that the matrix $A(r)$ dominates $B(r)$ in the sense of "upper-orthant dominance".[22]

**Example 2** Consider the $n$-dimensional random vectors $X$, $Y$, $Z$, and $V$ with symmetric mixture distributions on support $\mathcal{L} = \{1, 2, 3\}^n$, generated by the $2\times 3$ matrices $A$, $B$, $C$, and $D$, respectively:

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad B = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix} \quad C = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \quad D = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

The rows of each matrix have the same arithmetic average, $(\frac{1}{4}\ \frac{1}{2}\ \frac{1}{4})$, which represents the common marginal distribution of each $X_r$, $Y_r$, $Z_r$, and $V_r$. $A$, $B$, and $C$ are stochastically ordered, so in each, the first (second) row unambiguously corresponds to a low (high) realization of the aggregate shock. $D$, however, is not stochastically ordered. It is easily checked that $A \succ_{CCM} B$, $A \succ_{CCM} C$, and $A \succ_{CCM} D$. These conditions formally capture the fact that in $A$, the distribution of the variables conditional on the low (high) realization of the aggregate shock is more concentrated on low (high) values, compared to any of $B$, $C$, and $D$. Hence Theorem 4 implies that for any $n$, $(X_1, \ldots, X_n)$ dominates $(Y_1, \ldots, Y_n)$, $(Z_1, \ldots, Z_n)$, and $(V_1, \ldots, V_n)$ according to $\succ_{SPM}$.[23]

Theorem 4 has potential applications in several areas of economics. In finance and insurance contexts, the supermodular ordering is useful for comparing the degree of dependence among asset returns or insurance claims in a portfolio.[24] In contrast to the approach taken by Epstein and Tanny (1980) and Patton (2009), who compare only bivariate distributions, we can, by focusing on mixture distributions, compare interdependence according to the supermodular ordering for

[22] Athey and Levin (2001) compared information structures (joint distributions of signal and state of the world) for "monotone decision problems". For the special case where both $A(r)$ and $B(r)$ are stochastically ordered, the partial ordering $A(r) \succ_{CCM} B(r)$ is formally very similar to the partial ordering on information structures that Athey and Levin showed to correspond to preference by all decision-makers with payoff functions supermodular in the state and the action. Both orderings have the interpretation that one set of (first-order) stochastically ordered conditional distributions is more dispersed than the other.

[23] For symmetric mixture distributions generated from two-row matrices and for any $n \geq 2$, we can show that the pair of conditions in Theorem 4 are necessary as well as sufficient for $X \succ_{SPM} Y$. Example 2 illustrates this result. $B$ and $C$ cannot be ranked according to $\succ_{CCM}$, so it follows from the necessity of the $\succ_{CCM}$ condition that $Y$ and $Z$ cannot be ranked according to $\succ_{SPM}$. In fact, because the third column of $\bar{C}$ majorizes (strictly) the third column of $\bar{B}$, we can deduce that for $w(x) = I_{\{x_1 \geq 3, x_2 \geq 3\}}$, $Ew(Z) > Ew(Y)$, and because the second column of $\bar{B}$ majorizes (strictly) the second column of $\bar{C}$, we can deduce that for $w(x) = I_{\{x_1 \geq 2, x_2 \geq 2\}}$, $Ew(Y) > Ew(Z)$. Moreover, even though $D \succ_{CCM} B$, because $D$ is not stochastically ordered, it follows that $V$ does not supermodularly dominate $Y$; this can be checked by taking $w(x) = I_{\{x_1 \geq 3, x_2 \geq 2\}}$.

[24] See Müller and Stoyan (2002) and Denuit, Dhaene, Goovaerts, and Kaas (2005).

portfolios with any number of distinct components. Mixture distributions are increasingly used by financial economists to model positively dependent risks in a portfolio, but our theorem yields supermodular dominance results for a wider class of such distributions than previous analyses (e.g. Cousin and Laurent, 2008). Macroeconomists seeking to understand the sources of variation in aggregate production are naturally interested in the interdependence of output levels across sectors. Hennessy and Lapan (2003) have in fact proposed using the supermodular ordering to make such comparisons of "systematic risk". In the spirit of our mixture distribution analysis, Foerster, Sarte, and Watson (2011) have empirically explored how the relative importance of aggregate vs. sectoral shocks affects the covariation of output levels across sectors and hence the volatility of overall output. In a similar spirit, theoretical analyses of coordination games have used mixture distributions to examine how changes in the degree of interdependence in agents' information affect the volatility of aggregate behavior (Myatt and Wallace, 2012). Theorem 4 provides a flexible method for generating or modeling distributions that are comparable according to the supermodular ordering, by changing the relative importance of aggregate and idiosyncratic shocks.

# 4 Comparing Distributions Generated from Heterogeneous Lotteries

Let $(X_1, \ldots, X_n) \in \{0,1\}^n$ (resp., $(Y_1, \ldots, Y_n) \in \{0,1\}^n$) denote the outcomes of $n$ independent Bernoulli trials, where the probability of success (outcome=1) on trial $i$ is $a_i$ (resp., $b_i$). If $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$, so the expected number of successes is the same for the random vector $X$ as for $Y$, what can be said about the relative variability of the distributions of $c(X)$ and $c(Y)$?[25] Karlin and Novikoff (1963) showed that if $(a_1, \ldots, a_n)$ majorizes $(b_1, \ldots, b_n)$, then $c(X) \prec_{\mathcal{CX}} c(Y)$.

To develop an intuition for why a less dispersed vector of success probabilities generates greater variability of the total number of successes, consider the case where $n = 2$, $(a_1, a_2) = (1, 0)$, and $(b_1, b_2) = (\frac{3}{4}, \frac{1}{4})$. Then $c(X) = 1$ with probability 1, while $c(Y)$ takes the values $\{0, 1, 2\}$ with probabilities $\{\frac{3}{16}, \frac{5}{8}, \frac{3}{16}\}$.

Propositions 3 and 4, combined with Karlin and Novikoff's result, imply that if $(a_1, \ldots, a_n)$ majorizes $(b_1, \ldots, b_n)$, then i) $(X_1, \ldots, X_n) \prec_{\mathcal{SSPM}} (Y_1, \ldots, Y_n)$ and ii) the symmetrized version of the distribution of $X$ is dominated by the symmetrized version of the distribution of $Y$ according to the supermodular ordering.

In what follows, let $X' = (X'_1, \ldots, X'_n)$ denote the random vector whose distribution matches

[25] As in Section 2.3, $c(\cdot)$ denotes the count function.

the symmetrized distribution of the random vector $X$, and define $Y'$ similarly. In the example above, the distribution of $(X_1', X_2')$ places probability $\frac{1}{2}$ on $(1,0)$ and $(0,1)$, while that of $(Y_1', Y_2')$ places probability $\frac{5}{16}$ on $(1,0)$ and $(0,1)$ and probability $\frac{3}{16}$ on $(1,1)$ and $(0,0)$. These two joint distributions have identical (uniform) marginals on $\{0,1\}$. Clearly, $(X_1', X_2') \prec_{SPM} (Y_1', Y_2')$, since the distribution of $Y'$ is obtained from that of $X'$ by an elementary transformation of size $\frac{3}{16}$. Moreover, whereas the distribution of $(Y_1', Y_2')$ displays some negative dependence, the distribution of $(X_1', X_2')$ displays perfect negative dependence. Finally, note that had we started with a uniform vector of success probabilities for the independent trials, then the resulting multivariate outcome distribution would have been symmetric, so even after symmetrization it would have displayed independence.

The example illustrates that lower dispersion in the vector of success probabilities corresponds not only to higher variability of the total number of successes, but also to symmetric supermodular dominance of the $n$-dimensional outcome distribution. Furthermore, when an independent distribution on $\{0,1\}^n$ with unequal marginals is symmetrized, the symmetrized version displays negative interdependence, and is more negatively interdependent the more different from one another are the marginals of the original, independent distribution.

This section focuses on multivariate distributions representing the outcome of $n$ independent lotteries with an *arbitrary* finite support, exploring the connections between lower dispersion in the (marginal) distributions of the independent lotteries, the symmetric supermodular ordering on the joint distribution of lottery outcomes, and the degree of negative interdependence in the symmetrized versions of these joint distributions. Given two sets of $n$ independent lotteries, Theorem 5 provides sufficient conditions for their outcome distributions to be comparable according to the symmetric supermodular ordering, or equivalently, for the degree of negative interdependence of the symmetrized versions of their outcome distributions to be comparable according to the supermodular ordering. We show below how Theorem 5 can be used to compare different production designs in the presence of complementarity among tasks and in Section 5.1 how it can be used to compare ex post inequality of reward schemes under uncertainty.

We consider random vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ generated by $n \times m$ row-stochastic matrices $A$ and $B$, respectively, as follows: the $i^{th}$ row of $A$ (resp. $B$) represents the marginal distribution of $X_i$ (resp. $Y_i$) on support $\{1, \ldots, m\}$, and the $\{X_i\}$ (resp. $\{Y_i\}$) are independent.[26] Just as above we compared sets of $n$ independent Bernoulli trials with the same average success

[26] The choice of support $\{1, \ldots, m\}$ for each $X_i$ and $Y_i$ is without loss of generality, since the symmetric supermodular ordering is invariant to monotonic coordinate changes that preserve the symmetry of the lattice.

probability, here we compare sets of $n$ independent lotteries with the same average distribution over the $m$ prizes. This constraint translates into the requirement that for each $j$, the $j^{th}$ column of $A$ has the same sum as the $j^{th}$ column of $B$.

Denote by $(X_1', \ldots, X_n')$ and $(Y_1', \ldots, Y_n')$ the random vectors whose distributions match the symmetrized distributions of $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$, respectively. The common marginal distribution of the $\{X_i'\}$ is the average of the rows of matrix $A$. Hence, requiring that the matrices being compared have matching column sums implies that the common marginal distribution of the $\{X_i'\}$ is identical to that of the $\{Y_i'\}$.

In the Bernoulli example above, dispersion of the $n$-vector of success probabilities was captured by majorization. For $n$ lotteries with $m$-point supports, represented by the $n$ rows of a matrix, our cumulative column majorization ordering defined in Section 3 formalizes the notion of greater dispersion in the lotteries, holding their average fixed.

Theorem 5 below provides sufficient conditions for two sets of independent lotteries to be comparable according to the symmetric supermodular ordering. These sufficient conditions are very closely related to the sufficient conditions for supermodular dominance of mixture distributions identified in Theorem 4, and the techniques for proving these theorems are likewise very similar.

**Theorem 5** *Let $A$ and $B$ be $n \times m$ row-stochastic matrices generating the independent random vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$, respectively. Let $(X_1', \ldots, X_n')$ and $(Y_1', \ldots, Y_n')$ have distributions matching the symmetrized distributions of $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$, respectively. Suppose that i) $A$ is stochastically ordered, and ii) $A \succ_{CCM} B$. Then $(X_1, \ldots, X_n) \prec_{SSPM} (Y_1, \ldots, Y_n)$ and $(X_1', \ldots, X_n') \prec_{SPM} (Y_1', \ldots, Y_n')$.*

We have examples showing that Theorem 5 does not hold if we drop either condition *i)* or *ii)*.[27]

As a first application of Theorem 5 (a second is developed in Section 5.1), we revisit Bond and Gomes's (2009) multi-task principal-agent model. Suppose that each row $i$ of $A$ and $B$ represents the distribution of performance, over $m$ possible levels, on one of $n$ tasks, and that performance levels are independently distributed across tasks. The production function is symmetric and supermodular in the performance levels on the different tasks, reflecting interchangeability and complementarity among tasks. A manager must choose how to allocate resources across the different tasks, thereby

[27] Hu and Yang (2004, Theorem 3.4) showed that for any stochastically ordered row-stochastic matrix $A$, the symmetrized version of the distribution of $X$ displays negative association (a concept of negative dependence defined in Joag-Dev and Proschan (1983)), which in turn implies that this symmetrized version is supermodularly dominated by its independent counterpart (the independent symmetric distribution with identical marginals). This latter result corresponds to the special case of Theorem 5 where the rows of the matrix $B$ are all identical.

shifting the distributions of performance, subject to a constraint on the average distribution over all tasks. Theorem 5 identifies conditions under which expected production is higher in one setting than the other for all symmetric supermodular production functions.

Bond and Gomes focus on binary outcomes for each task ($m = 2$). An agent chooses a level $e_i \in [\underline{e}, \overline{e}]$ of effort for each task $i$, incurring a total effort cost $\sum_{i=1}^{n} e_i$. The probability of success on task $i$ equals $e_i$, and the principal's benefit is assumed to be a convex function of the number of successes, so it is a symmetric supermodular function of the vector of binary task outcomes. For a given $\sum_{i=1}^{n} e_i$, Bond and Gomes show that the socially efficient allocation of this total effort involves equal effort on all tasks. However, the optimal contract rewarding the agent as a function of the number of successes may well induce the agent to exert minimal effort $\underline{e}$ on a subset of tasks and maximal effort $\overline{e}$ on the remainder. In this case, given the total effort exerted, the agent's effort allocation actually minimizes expected social surplus. Theorem 5 implies these conclusions about the best and worst allocations from a social perspective.[28]

More generally, Theorem 5 allows us to examine, for arbitrary $m$ and $n$, the existence, in the sense of the symmetric supermodular ordering, of a best and worst set of independent lotteries, holding fixed the average distribution over the prizes. Because the symmetric supermodular ordering is a partial ordering, one should not generally expect the existence of a best and a worst distribution. However, Proposition 8 (in the Appendix) shows that for the class of distributions considered here, a best and a worst set of lotteries do indeed exist.

# 5 Applications

## 5.1 Welfare and Inequality

### Ex Post Inequality in the Presence of Uncertainty

[28]The effort allocation determines an $n \times 2$ row-stochastic matrix, the second column of which is the vector of success probabilities, and holding the total effort fixed corresponds to fixing the column sums of the matrix. With two columns, 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 the vector of success probabilities from one effort allocation majorizes the vector from another, then the former allocation generates lower expected social surplus, for all symmetric supermodular benefit functions. (Bond and Gomes's conclusions also follow from Karlin and Novikoff's (1963) result for Bernoulli trials, discussed above). The final step is to observe that a vector of equal success probabilities is majorized by all vectors with the same total; and one in which all probabilities are either minimal or maximal ($\underline{e}$ or $\overline{e}$) majorizes all vectors with the same total.

When individual outcomes are uncertain, members of a group may be concerned, ex ante, about ex post inequality.[29] As argued by Meyer and Mookherjee (1987), an aversion (on the part of a group or a social planner) to ex post inequality can be formalized by adopting an ex post welfare function that is symmetric and supermodular in the realized utilities of the individuals. 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. This suggests that tournaments should be dominated, in the sense of the symmetric 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, 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. Theorem 5 can be applied to generalize this result to tournaments that are arbitrarily asymmetric across individuals.

With $n$ individuals and $n$ distinct prizes, a "tournament" reward scheme allocates each of the prizes to exactly one individual, and it is fully described by the probability it assigns to each of the $n!$ possible prize allocations. For welfare computations, a tournament may be summarized by a matrix $B$ that is bistochastic (both its columns and its rows sum to 1), where the $i^{th}$ row of $B$ describes individual $i$'s marginal distribution over the $n$ prizes. The more asymmetric the tournament is, the more disparate are the rows of the corresponding matrix $B$. Given any tournament, consider the associated reward scheme giving each individual the same marginal distribution as in the tournament, but which determines individual rewards independently. For any tournament, however asymmetric, Theorem 5 implies that expected ex post welfare under the tournament is less than or equal to expected ex post welfare under the independent joint distribution of rewards sharing the same set of marginals, for all symmetric and supermodular ex post welfare functions.[30]

[29] See Meyer and Mookherjee, 1987; Meyer, 1990; Ben-Porath et al, 1997; Gajdos and Maurin, 2004; Chew and Sagi, 2012; and Saito, 2013 . This concern is distinct from concerns about the mean and riskiness of rewards.

[30] For a symmetric tournament, the joint distribution of rewards is dominated according to the supermodular ordering by the independent joint distribution sharing the same set of marginals. To see why, when analyzing tournaments that are arbitrarily asymmetric, we need to impose symmetry of the ex post welfare function, consider the following tournament with $n = 3$: with probability $\frac{1}{2}$, prizes $h$, $m$, and $l$, where $h > m > l$, are allocated to individuals 1, 2, and 3, respectively, and with probability $\frac{1}{2}$, $h$, $m$, and $l$ are allocated to individuals 3, 1, and 2, respectively. In this tournament, the rewards to 1 and 2 are *positively* dependent, even though the rewards to 1 and 3 (as well as the rewards to 2 and 3) are negatively dependent. The positive dependence of the rewards to 1 and 2 implies that the tournament reward distribution is not supermodularly dominated by the corresponding independent distribution. When we impose symmetry of the ex post welfare function, in addition to supermodularity, we are

**Proposition 5** *For any number $n$ of individuals, given any tournament, the joint distribution of prizes under the tournament is dominated, according to the symmetric supermodular ordering, by the independent joint distribution sharing the same set of marginals.*

### 5.2 Search and Voting in Committees with Conflicting Interests

There are many contexts where it is of interest to assess the degree of alignment in the preferences or information of members of decision-making groups.[31] Modeling consensus-building in committees, Caillaud and Tirole (2007) study how the degree of interdependence of members' ex ante uncertain payoffs from a proposal affects the proposer's persuasion strategy. In a model of search and voting, Moldovanu and Shi (2013) examine how the degree of alignment in committee members' preferences affects equilibrium search and welfare. Both papers focus on the unanimity rule and impose restrictions on the payoff distributions: in Caillaud and Tirole, payoffs are binary, while Moldovanu and Shi focus on a single-parameter family of payoff functions. Here, we use the supermodular ordering as a non-parametric, $n$-dimensional ordering of interdependence in preferences and adapt and generalize Moldovanu and Shi's analysis of search and voting.

Job candidates are interviewed sequentially, without recall, by an $n$-person committee. The period-$t$ candidate has attribute vector $X_t = (X_{1t}, \ldots, X_{nt})$, where $X_t$ is i.i.d. across periods and has a known distribution. Committee member $i$'s utility equals $X_{it}$ if the period-$t$ candidate is hired (in which case search stops), and $i$ incurs search cost $c_i$ of evaluating attribute $i$ for each new candidate. We suppose initially that unanimous approval is required for a candidate to be hired, otherwise search continues. If $(Y_1, \ldots, Y_n) \sim g$, $(X_1, \ldots, X_n) \sim f$, and $(Y_1, \ldots, Y_n) \succ_{SPM} (X_1, \ldots, X_n)$, we will say that members' interests are *more aligned* when the values of the attributes are distributed according to $g$ than when they are distributed according to $f$.

In equilibrium, each member $i$ chooses a reservation level $z_i$ for attribute $i$, and the equilibrium reservation levels $(z_1, \ldots, z_n)$ satisfy the $n$ simultaneous equations

$$c_i = E\left[(X_i - z_i) I_{\{X_j \geq z_j \forall j\}}\right], \quad i = 1, \ldots, n. \tag{11}$$

Each member $i$ equates his cost of one more search with the expected gain from one more search, assessed relative to stopping now and obtaining $z_i$. Since search will stop next period if and only if

comparing the "average" degree of negative interdependence across the whole set of individuals. Equivalently, as Proposition 3 showed, we are comparing the interdependence of the symmetrized versions of the tournament reward distribution and of the independent joint distribution with the same marginals.

[31] See Boland and Proschan (1988) and Baldiga and Green (2013) on alignment of preferences, and Prat (2002) and Gendron-Saulnier and Gordon (2014) on alignment of information.

all members approve the next candidate, the expected gain to member $i$ depends on the reservation levels of the others via the factor $I_{\{X_j \geq z_j \forall j\}}$ multiplying $(X_i - z_i)$.

The key observation is that the gain to each member $i$ from one more search (square brackets in (11)) is supermodular in $(X_1, \ldots, X_n)$ for all $(z_1, \ldots, z_n)$. To see this, we rewrite this expression as $\prod_{j=1}^{n} r_j(X_j, z_j)$, where each $r_j(X_j, z_j)$ is nonnegative and increasing in $X_j$. Hence, as interests become more aligned, each member's expected gain from one more search increases. Since the right-hand side of (11) is decreasing in $z_i$, a greater alignment of interests implies that the optimal $z_i$ increases, for all $z_{-i}$. If, in particular, the committee is symmetric ($c_i = c$ for all $i$ and the distributions of attributes are symmetric across members), then a greater alignment implies that the common equilibrium reservation value increases: members become choosier.

To examine how the impact of greater alignment of interests depends on the voting rule, suppose now that a candidate is hired if and only if at least $m$ of the $n$ members vote to stop searching. For given $(z_1, \ldots, z_n)$, let $K(z_1, \ldots, z_n) = \{k | X_k \geq z_k\}$. The equilibrium reservation levels satisfy

$$c_i = E\left[(X_i - z_i) I_{\{|K| \geq m\}}\right], \quad i = 1, \ldots, n. \tag{12}$$

When unanimity is required to reject a candidate ($m = 1$), the expression in square brackets can be written as $(X_i - z_i) + |X_i - z_i| I_{\{X_j < z_j \forall j\}}$, which is again supermodular in $(X_1, \ldots, X_n)$, for all $(z_1, \ldots, z_n)$.[32] Consequently, the previous comparative statics still hold.

However, for voting rules intermediate between the two extremes (unanimity required for acceptance or unanimity required for rejection), the realized gain from one more search is not everywhere a supermodular function of the realized values of the attributes.[33] This failure of supermodularity can have a bite: we have examples with three members for which, under the simple majority rule, greater alignment in members' interests results in *lower* equilibrium reservation values: members become less choosy.

[32] It is the sum of two supermodular functions, the second of which is supermodular because it can be written as $\prod_{j=1}^{n} r_j(X_j, z_j)$, where each $r_j(X_j, z_j)$ is nonnegative and decreasing.

[33] To see why supermodularity can fail, observe that when two other committee members both switch their vote from "no" to "yes", this may be enough to hire a candidate such that $i$'s realized gain, $X_i - z_i$, is strictly negative, even when a switch by just one of the other two members would not be enough to get that candidate hired, in which case $i$'s realized gain would be 0.

### 5.3 Systemic Risk: Network Configuration and Interdependent Default

Financial economists, stimulated by the financial crisis, have been developing measures of "systemic risk", capturing the interdependence of the components of the financial system as a whole.[34]

This application shows that changes in the structures of financial linkages between banks can naturally lead to distributions of default risks that are ranked according to the symmetric supermodular ordering. We revisit the model developed by Allen, Babus, and Carletti (2012), who consider a particular diversification strategy of banks, asset-swapping, and examine how the pattern of asset swaps affects market outcomes and welfare. We generalize a stylized version of their model,[35] focusing on how different patterns of asset swaps (represented by different networks) generate multivariate distributions of bank failures with different degrees of interdependence.

Consider six banks and two networks of asset swaps, the "clustered" and the "unclustered". Each bank $i$ funds a project with return $\theta_i \in \{L, H\}$. The projects' returns are i.i.d. with $P(\theta_i = H) = p$. In the clustered network, banks 1,2, and 3 swap assets among themselves so that each of them holds an identical portfolio with return $Y_i' = \frac{1}{3}\sum_{j=1}^{3} \theta_j$ for $i \leq 3$, and similarly for banks 4,5, and 6. In the unclustered network, banks are arranged in a circle, and each bank swaps one-third of its assets with each of its two neighbors, yielding returns $X_i' = \frac{1}{3}(\theta_{i-1 \text{ mod } 6} + \theta_i + \theta_{i+1 \text{ mod } 6})$ for all $i$. The marginal distribution of each bank's return is the same in the two networks, but the form of the interdependence of bank returns differs. In the clustered network, banks in the same cluster have perfectly positively dependent returns, while those in different clusters have independent returns; in the unclustered network, by contrast, the dependence between a given bank's return and that of its neighbors is strongly positive (but imperfect), that between its own return and those of its neighbors' neighbors is weakly positive, and its return is independent of that of the remaining bank. Suppose a bank defaults (solvency status=0) if its return is less than or equal to some level $d \in [L, H)$, otherwise it is solvent (solvency status=1). Let banks' solvency statuses in the clustered network be described by $(Y_1, \ldots, Y_6) \in \{0, 1\}^6$, so $Y_i = I_{\{Y_i' > d\}}$, and in the unclustered network by $(X_1, \ldots, X_6) \in \{0, 1\}^6$, so $X_i = I_{\{X_i' > d\}}$.

[34]For example, Adrian and Brunnermeier (2009) and Acharya et al (2010) develop measures of association between negative events for an individual firm and negative events for the market. Beale et al (2011) study the interplay between diversification at the level of the financial institution, which lowers individual risk, and increasing similarity of institutions' portfolios, which raises systemic risk.

[35]Compared to our model, Allen et al (2012) restrict attention to the case where projects are equally likely to succeed and fail and where a bank defaults if and only if all three of the projects in its portfolio fail. Their model involves additional features, such as different maturities of debt, through which interdependence of banks' returns indirectly influences welfare.

We compare systemic risk in the two networks by using the symmetric supermodular ordering to compare the interdependence of the random vectors $(Y_1, \ldots, Y_6)$ and $(X_1, \ldots, X_6)$. Supermodularity of the "systemic cost function" $C(x_1, \ldots, x_6)$ reflects the judgment that the additional cost to the system from two bank defaults is higher than the sum of the marginal costs from each individual default, and symmetry reflects the fact that the banks in this setting are of equal size.[36][37] Proposition 4 can be applied to show the following result:

**Proposition 6** *For any probability of project success p and for any common failure threshold d for banks,* $(Y_1, \ldots, Y_6) \succ_{SSPM} (X_1, \ldots, X_6)$*. Hence for any supermodular and symmetric systemic cost function, expected systemic cost is higher under the clustered than under the unclustered network.*

The joint distributions of solvency statuses compared in Proposition 6, which have support $\{0,1\}^6$, are the coarsened (and translated) versions of the joint distributions of the actual bank returns, which have support $\{L, \frac{2L+H}{3}, \frac{L+2H}{3}, H\}^6$. Proposition 4 can also be applied to examine whether the distributions of actual returns under the clustered and unclustered networks can be ranked according to $\succ_{SSPM}$. In the Appendix, we show that this stronger result does not hold, by presenting a supermodular and symmetric systemic cost function defined on $\{L, \frac{2L+H}{3}, \frac{L+2H}{3}, H\}^6$ whose expectation is strictly higher under the unclustered network than under the clustered one.

### 5.4 Supermodular Objective Functions: Other Applications

#### Multidimensional Deprivation

The supermodular ordering is a useful tool for making comparisons of deprivation given data on multiple attributes, such as income, health, and education.[38] To compare multidimensional deprivation between two datasets (e.g., two countries, two time periods), one popular strategy is to aggregate across attributes to generate a deprivation measure for each individual and then sum these measures to obtain an aggregate deprivation measure for the whole dataset. Importantly, under this strategy, comparisons of deprivation depend upon i) whether the different dimensions are regarded

[36]By using a *symmetric* supermodular function for comparisons of expected systemic cost, we are comparing the "average" degree of interdependence across the whole set of banks.

[37]Since the marginal distribution of each bank's return, and hence of each bank's solvency status, is the same across the two networks, it is irrelevant whether or not we specifically restrict $C(x_1, \ldots, x_6)$ to be decreasing: when comparing multivariate distributions with identical marginals, the decreasing (symmetric) supermodular ordering is equivalent to the (symmetric) supermodular ordering. See the remark following Theorem 3.

[38]See Atkinson and Bourguignon, 1982, the Symposium in Honor of Amartya Sen in the *Journal of Public Economics*, Vol. 95, 2011, and the Symposium on Inequality and Risk in the *Journal of Economic Theory*, Vol. 147, 2012.

as complements or substitutes in the individual deprivation function and ii) the interdependence in the joint distributions of attributes in the two datasets.

According to the *intersection approach*, an individual is deemed "multidimensionally deprived" if, for each attribute $i$, his achievement $x_i$ falls below some threshold $z_i$ (see Alkire and Foster (2011) and Atkinson (2003)). This approach implies an individual deprivation function of the form $d(x_1, \ldots, x_n) = I_{\{x_i \leq z_i \ \forall i\}}$, which is supermodular, since it is a lower-orthant indicator function. Therefore, if one multidimensional distribution of achievements dominates another according to the supermodular ordering, the aggregate level of deprivation obtained by summing this deprivation measure over individuals is *higher* for the former distribution than for the latter, regardless of the thresholds. By contrast, the *union approach* classifies an individual as deprived if and only if there is at least one dimension $i$ in which $x_i \leq z_i$. The deprivation function is now *submodular*[39] and leads to the *opposite* result: higher interdependence in the multidimensional distribution of achievement levels, in the sense of the supermodular ordering, implies lower aggregate deprivation.

In the intersection approach, there is a complementarity among the different dimensions in the determination of individual deprivation. A natural generalization, which retains this complementarity, would make individual deprivation an increasing convex function of the number of dimensions in which $x_i$ falls below the threshold $z_i$:

$$d(x_1, \ldots, x_n) = \phi \left( \sum_{i=1}^{n} I_{\{x_i \leq z_i\}} \right), \tag{13}$$

where $\phi$ is increasing and convex. Similarly, a natural generalization of the union approach, which retains the substitutability among the different dimensions, would express individual deprivation in the form (13) where $\phi$ is increasing and concave. Our analysis easily extends to these deprivation functions.[40]

**Prediction and Parameter Estimation**

Consider the problem of making a prediction $\tilde{\theta}$ about the value of an unknown parameter $\theta$, to minimize the value of a loss function $L(\tilde{\theta} - \theta)$ that is convex and minimized at 0. The prediction

[39] The individual deprivation measure is $d(x_1, \ldots, x_n) = 1 - I_{\{x_i \geq z_i \ \forall i\}}$, which is a submodular function of $(x_1, \ldots, x_n)$, since the supermodular upper-orthant indicator function appears with a negative sign.

[40] In either case, we can regard the binary variables $x'_i \equiv I_{\{x_i \leq z_i\}}$ as coarsened versions of the original data. For $\phi$ convex (concave), the deprivation function in (13) is a symmetric supermodular (symmetric submodular) function of $(x'_1, \ldots, x'_n)$. Therefore, for a given vector of thresholds $(z_1, \ldots, z_n)$, aggregate deprivation will be lower in one population than another, for all deprivation measures in the class in (13) with $\phi$ convex (concave), if and only if the distribution of $(x'_1, \ldots, x'_n)$ in one population is more (less) interdependent, in the sense of the symmetric supermodular ordering, than in the other. Proposition 4 then shows that in this context, symmetric supermodular dominance is equivalent to univariate convex dominance for distributions of $\sum_{i=1}^{n} x'_i = \sum_{i=1}^{n} I_{\{x_i \leq z_i\}}$.

is based on some data $(X_1, \ldots X_n)$, where $X_i$ has distribution $F_i(\cdot|\theta)$, conditional on $\theta$. We focus, for this illustration, on the case in which the estimator $\tilde{\theta}$ is an affine function of the observed variables:[41] $\tilde{\theta} = \sum \kappa_i X_i$ for some nonnegative weights $\{\kappa_i\}_{i=1,\ldots,n}$.

The supermodular ordering can be used to compare the richness of various datasets, holding fixed the marginal distributions $F_i(\cdot|\theta)$. Intuitively, a dataset is richer if it comes from independent sources instead of closely related ones. Formally, let the datasets $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ be generated by joint distributions $F(\cdot|\theta)$ and $G(\cdot|\theta)$, respectively, and suppose that $F(\cdot|\theta) \prec_{SPM} G(\cdot|\theta)$ for all $\theta$. Then, since $L(\sum_i \kappa_i x_i - \theta)$ is supermodular in $(x_1, \ldots, x_n)$ for all $\theta$, we have that $E[L(\sum \kappa_i X_i - \theta)|\theta] \leq E^G[L(\sum \kappa_i Y_i - \theta)|\theta]$ for all nonnegative $\{\kappa_i\}$, for all convex $L$, and for all $\theta$. That is, for a given affine estimator $\tilde{\theta}$, the dataset $(X_1, \ldots, X_n)$ in which the observations are richer, in the sense of being less supermodularly dependent, generates a better prediction of the unknown parameter. From this it follows that the less supermodularly dependent dataset also generates a better prediction when the weights $\{\kappa_i\}$ can be chosen optimally according to the dataset.

**Matching**

The supermodular ordering is well suited for comparing the efficiency of two-sided or many-sided matching mechanisms when the outcomes of the matching process are subject to frictions. With production functions that are supermodular in the qualities of the different components of a match, efficient matching is perfectly assortative, corresponding to a perfectly positively dependent joint distribution of the random variables representing the qualities of each component. In the presence of noisy information, costly search, or credit constraints, perfectly assortative matching will generally not arise. In these settings, Theorem 1 and the constructive methods of Section 2.1 can be used to assess when one matching mechanism will generate higher expected surplus than another, for all supermodular production functions. While applications to two-sided matching problems have received some attention,[42] multi-dimensional applications remain largely unexplored.[43]

**Decision Making**

The application to committee decisions illustrated the relationship between increased interdependence and the comparative statics of decisions, showing how greater alignment in agents'

[41]While special, affine estimators are pervasive in statistics and econometrics. If, for example, $(X_1, \ldots, X_n)$ are exchangeable and have mean $\theta$, then $\tilde{\theta}$ will be the sample average of those variables.

[42]Fernandez and Gali (1999) use the known bivariate characterization of the supermodular ordering (Levy and Paroush, 1974) to compare the efficiency losses from markets and tournaments as allocative mechanisms in an economy with borrowing constraints. Meyer and Zeng (2013) employ the ordering to compare assignment mechanisms when qualities are ex ante uncertain and different mechanisms generate and use different information.

[43]One exception is Prat (2002), but he compares only a perfectly positively dependent joint distribution with an independent one.

preferences, as captured by supermodular dominance of the distribution of agents' payoffs, affected committee search and voting behavior. As another example, Gollier (2011) applies the supermodular ordering in the bivariate case to study how the efficient discount rate in an extended Ramsey-type model depends on the interdependence between initial consumption and the growth rate of consumption. A more systematic exploration of the role of the supermodular ordering in the comparative statics analysis of decisions should be a particularly fruitful area for future research.

# 6 Discussion: Continuous Support and Copulas

## 6.1 Continuous Support

The supermodular ordering on a continuous support can be characterized in terms of all its discrete coarsenings. For $F, G$ with continuous densities on $\mathcal{L} = \times_i [a_i, b_i]$, define the supermodular ordering on $\mathcal{L}$ as follows: $F \prec_{CSPM} G$ if and only if $E[w|F] \leq E[w|G]$ for all integrable supermodular functions on $\mathcal{L}$.

A finite coarsening $\tilde{\mathcal{L}}$ of $\mathcal{L}$ is defined by a finite partitioning $\tilde{\mathcal{L}}_i$ of each $\mathcal{L}_i$. The coarsened version of $F$ on $\tilde{\mathcal{L}}$ is the distribution $\tilde{F}$ such that for all $\tilde{x} \in \tilde{L}$, $\tilde{F}(\tilde{x})$ is the probability that $F$ puts on the on the cell (hyperrectangle) defined by the Cartesian product of the $\tilde{x}_i$'s: $\tilde{F}(\tilde{x}) = F(\times_i \tilde{x}_i)$. For any function $w$ on $\mathcal{L}$, the coarsened version $\tilde{w}$ of $w$ on $\tilde{\mathcal{L}}$ is the average of $w$ over the hyperrectangle defined by each $\times_i \tilde{x}_i$. Formally,

$$\tilde{w}(\tilde{x}) = \frac{\int_{\times_i \tilde{x}_i} w(x)dx}{\int_{\times_i \tilde{x}_i} dx}. \tag{14}$$

In light of the robustness to coarsening of the ordering $\prec_{SPM}$ noted in Section 2, it is not surprising that the supermodular ordering on $\mathcal{L}$ is stronger than the supermodular ordering on every finite coarsening of $\mathcal{L}$. With continuous densities, the following equivalence result holds.

**Theorem 6** *Suppose that distributions $F$ and $G$ have continuous densities. $F \prec_{CSPM} G$ if and only if $\tilde{F} \prec_{SPM} \tilde{G}$ on all finite coarsenings $\tilde{\mathcal{L}}$ of $\mathcal{L}$.*

## 6.2 Comparing Copulas

A useful approach to examining the interdependence of random variables, which is widespread in econometrics and finance and gaining prominence in the study of intergenerational mobility and income dynamics, is based on the concept of a copula.[44] Given any distribution function $F$ of

[44]Copulas are used in statistics and econometrics to model the intertemporal dependence of time series (see for example Joe (1997, Ch. 8), and Beare (2010)). For applications in risk management and derivative pricing, see

$n$ variables, with marginal distributions $F_1, \ldots, F_n$, Sklar's theorem (1959) guarantees the existence of a function $C : [0,1]^n \to [0,1]$ such that

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)). \tag{15}$$

$C$ is called the *copula* of $F$. With discrete support, the values of the copula are pinned down on the domain $\tilde{\mathcal{L}} = \{(F_1(x_1), \ldots, F_n(x_n)) : (x_1, \ldots, x_n) \in \mathcal{L}\}$. The copula of a discrete distribution is therefore essentially unique.

Since $X_i \sim F_i$ implies that $F_i(X_i) \sim U[0,1]$, the copula is a distribution function each of whose marginal distributions is uniform on $[0,1]$. By normalizing marginal distributions to be uniform, copulas allow an exclusive focus on interdependence. Nevertheless, there remains the need to find an appropriate way of comparing copulas. The supermodular ordering, since it is based on complementarities in objective functions, provides an economically meaningful way to compare interdependence in copulas.

**Proposition 7** *$F \prec_{SPM} G$ on $\mathcal{L}$ if and only if $F$ and $G$ have identical marginals and their copulas satisfy $C_F \prec_{SPM} C_G$ on $\tilde{\mathcal{L}}$.*

Several works have examined whether copulas within specific parametric families with continuous supports can be ranked according to the supermodular ordering.[45] In contrast, the methods we have developed in this paper for characterizing and applying the supermodular ordering allow non-parametric comparisons of copulas. This feature makes our methods useful for comparing empirical copulas, as well as theoretical ones.[46]

Embrechts (2009) and Li (2000). Bonhomme and Robin (2006) use copulas to model individual earnings trajectories, and Chetty et al (2014) use them to examine intergenerational mobility.

[45]Positive results have been obtained for Archimedean copulas and asymmetric extensions thereof by Wei and Hu (2002) and for Gaussian, Student $t$, Clayton, and Marshall-Olkin families of copulas by Burtschell et al (2009).

[46]Chetty et al (2014) decompose the joint distribution of parent and child income into the copula and the marginal distributions; this approach allows them to disentangle changes over time in relative mobility from changes in inequality. They coarsen the empirical copula by aggregating parent and child incomes into quintiles.

# Appendix

## A Proofs for Section 2

**Proof of Theorem 1** Supermodular functions are characterized by the property (Topkis, 1978) that

$$w \in \mathcal{S} \iff w(x+e_i+e_j)+w(x) \geq w(x+e_i)+w(x+e_j) \tag{16}$$

for all $i \neq j$ and $x \in \mathcal{L}$ such that $x+e_i+e_j \in \mathcal{L}$. Equivalently,

$$w \in \mathcal{S} \iff w \cdot t \geq 0 \quad \forall t \in \mathcal{T}. \tag{17}$$

Equation (3) holds if and only if $g-f$ belongs to the convex cone $\mathcal{C}(\mathcal{T})$ generated by $\mathcal{T}$, defined by $\mathcal{C}(\mathcal{T}) = \{\sum_{t \in \mathcal{T}} \alpha_t t : \alpha_t \geq 0 \quad \forall t \in \mathcal{T}\}$. From (17), $\mathcal{S}$ is the dual cone of $\mathcal{C}(\mathcal{T})$. Since $\mathcal{C}(\mathcal{T})$ is closed and convex, this implies (Luenberger, 1969, p. 215) that $\mathcal{C}(\mathcal{T})$ is the dual cone of $\mathcal{S}$:

$$\delta \in \mathcal{C}(\mathcal{T}) \iff w \cdot \delta \geq 0 \quad \forall w \in \mathcal{S}.$$

Therefore, $f \prec_{\mathcal{S}PM} g$ if and only if $g - f \in \mathcal{C}(\mathcal{T})$. ■

**Proof of Proposition 1** Suppose that the hypotheses hold but that $f \neq g$. Then Theorem 1 implies that at least one $\alpha_t$ in (3) must be strictly positive. Let $t^z_{ij}$ denote a $t \in \mathcal{T}$ such that $\alpha_t > 0$. For the supermodular function $w(x) = x_i x_j$, the inequality in (16) is strict for all $x$, so $w \cdot t^z_{ij} > 0$ and thus $w \cdot g > w \cdot f$. Therefore $E(Y_i Y_j) > E(X_i X_j)$, and since any $t \in \mathcal{T}$ leaves marginal distributions unchanged, it follows that $Cov(Y_i, Y_j) > Cov(X_i, X_j)$, yielding a contradiction. ■

**Proof of Proposition 2** Without loss of generality, we prove the claim for the case where $\mathcal{L}_i = \{0, 1, \ldots, m_i - 1\}$ (other cases are treated with an obvious modification of the function $w$ below). Consider a point $x \in \mathcal{L}$ and a pair of dimensions $i, j$ such that the elementary transformation $t^* \equiv t^{x-e_i-e_j}_{i,j}$ is well-defined. Suppose that, contrary to the claim, there exist nonnegative coefficients $\alpha_s$ such that

$$t^* = \sum_{s \in \mathcal{T} \setminus \{t^*\}} \alpha_s s. \tag{18}$$

Define the function $w$ on $\mathcal{L}$ 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 strictly positive scalar product with all $t \in \mathcal{T}$ except for those of the form $t^{x-e_k-e_l}_{k,l}$ for some dimensions $k, l$. Since $t^*$ is one of the elementary transformations of this form, taking the scalar product of $w$ with both sides of (18) yields

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

This equation in turn implies that $\alpha_s = 0$ for all transformations $s \in \mathcal{T} \setminus \{t\}$ except possibly those of the form $t^{x-e_k-e_l}_{k,l}$ for some $k, l$. However, $t^*$ cannot be a positive linear combination of only transformations of this form. To see this, observe that any $s \neq t^*$ of the form $t^{x-e_k-e_l}_{k,l}$ for some $k, l$ must take value 0 at $x - e_i - e_j$, whereas $t^*$ evaluated at $x - e_i - e_j$ equals 1. ■

**Proof of Theorem 2** There always exists a feasible vector $(\alpha, \beta)$, namely $(\alpha, \beta) = (0, \delta^+)$. Moreover, the optimum value of program (4) is nonnegative since the feasibility constraints require that $\beta$ have nonnegative components. If $f \prec_{\mathcal{SPM}} g$, there exists $\alpha^* \geq 0$ such that $E^+\alpha^* = \delta^+$, so the optimum value of the program must be zero, since that value is achieved by $(\alpha, \beta) = (\alpha^*, 0)$. Reciprocally, if there exists $(\alpha^*, \beta^*)$ such that the value of the program is zero, then necessarily $\beta^* = 0$ and $E^+\alpha^* = \delta^+$. ■

**Proof of Theorem 3** The equivalence of conditions 2) and 3) follows from Theorem 1, the definition of $\gamma$, and the decomposition result in (6). It is obvious that 2) implies 1). We now show that 1) implies 3). For any supermodular $w$, let

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

where $z_i e_i$ is the vector with $i^{th}$ component equal to $z_i$ and all other components equal to 0. Clearly, $w^0(z_i e_i) = 0$ for all $i$ and $z_i$, and therefore, since $\gamma(z) = 0$ for all $z \neq z_i e_i$ for some $i$ and some $z_i$, $w^0 \cdot \gamma = 0$. Moreover, $w^0$ is supermodular, since it is the sum of supermodular functions, and $w^0$ is increasing, since for any $z \in \mathcal{L}$ and $i$ such that $z + e_i \in \mathcal{L}$, supermodularity of $w^0$ yields

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

Letting $\delta = g - f$, $g \succ_{\mathcal{ISPM}} f$ implies, therefore, that $w^0 \cdot \delta \geq 0$ and hence, since $w^0 \cdot \gamma = 0$, we have $w^0 \cdot (\delta - \gamma) \geq 0$. Furthermore,

$$\begin{aligned}
(w - w^0) \cdot (\delta - \gamma) &= \sum_{z \in \mathcal{L}} \left[ (\delta(z) - \gamma(z)) \left( \sum_{i=1}^{n} w(z_i e_i) - (n-1)w(0) \right) \right] \\
&= \sum_{z \in \mathcal{L}} \left[ (\delta(z) - \gamma(z)) \left( \sum_{i=1}^{n} w(z_i e_i) \right) \right] \\
&= \sum_{i=1}^{n} \sum_{k=0}^{m_i - 1} \left( \sum_{z : z_i = k} (\delta(z) - \gamma(z)) \right) w(k e_i) \\
&= 0,
\end{aligned}$$

where the second line follows since $\sum_{z \in \mathcal{L}} (\delta(z) - \gamma(z)) = 0$ and the final equality follows since (8) holds for all $i$ and all $k$. Thus, since $w^0 \cdot (\delta - \gamma) \geq 0$, it follows that $w \cdot (\delta - \gamma) \geq 0$, proving the first part of condition 3). Finally, taking, for each $i \in \mathcal{N}$ and $k \in \{1, \ldots, m_i - 1\}$, $w(z) = I_{\{z_i \geq k\}}$, $g \succ_{\mathcal{ISPM}} f$ implies that $\sum_{z : z_i \geq k} g(z) \geq \sum_{z : z_i \geq k} f(z)$, proving the second part of 3). ■

**Proof of Proposition 4** Recall that for $\mathcal{L} = \{0, 1, \ldots, m-1\}^n$, $k \in \{1, \ldots, m-1\}$, and $x \in \mathcal{L}$, $\bar{c}^k(x) \equiv \sum_{i=1}^{n} I_{\{x_i \geq k\}}$ and $\bar{c}(x) \equiv (\bar{c}^1(x), \ldots, \bar{c}^{m-1}(x))$. Since all permutations of $x$ correspond to the same $\bar{c}(x)$, a function $w$ is symmetric if and only if it can be written as $w(x) = \phi(\bar{c}(x))$, for some $\phi$ defined on $\tilde{L}^{m-1}$, the range of $\bar{c}(x)$. We now show that a function $w$ of this form is supermodular if and only if $\phi(\cdot)$ is supermodular and componentwise-convex.

As stated in (17), a function $w$ is supermodular if and only if $w \cdot t \geq 0$ for every elementary transformation $t^x_{i,j} \in \mathcal{T}$, as defined in (2). For $\mathcal{L} = \{0, 1, \ldots, m-1\}^n$, there are two distinct types of elementary transformation $t^x_{i,j}$, those in which for some $k \in \{1, \ldots, m-1\}$, $x_i = x_j = k-1$, and those in which $x_i \neq x_j$. For $w$ symmetric, $w \cdot t \geq 0$ for all $t = t^x_{i,j}$ such that $x_i = x_j = k-1$ if and only if the corresponding function $\phi$ defined above is componentwise-convex with respect to its $k^{th}$ argument, $\bar{c}^k(x)$. Moreover, for $w$ symmetric, $w \cdot t \geq 0$ for all $t = t^x_{i,j}$ such that $x_i = k-1$ and $x_j = l-1$, $k \neq l$, if and only if the corresponding $\phi$ is supermodular with respect to $\bar{c}^k(x)$ and $\bar{c}^l(x)$. Applying these two results for all $i, j \in \mathcal{N}$

and all $k, l \in \{1, \ldots, m-1\}$ shows that a symmetric $w$ defined on $\mathcal{L}$ is supermodular if and only if the correspoding $\phi$ defined on $\tilde{L}^{m-1}$ is componentwise-convex and supermodular in all its arguments. ■

## A.1 Incomparability of Mixture Distributions: Example

Let $X_i = \theta + \varepsilon_i$, where $\theta$ and $\{\epsilon_i\}_{i\in\mathcal{N}}$ are all independent, $\theta$ equals 2 or -2 with probability $p$ and $1-p$, respectively, and each $\varepsilon_i$ equals 1 or -1 with probability $1-p$ and $p$, respectively. Similarly, let $Y_i = \theta' + \varepsilon'_i$, where $\theta'$ and $\varepsilon'_{i\in\mathcal{N}}$ are all independent, $\theta'$ equals 1 or -1 with probability $1-p$ and $p$, respectively, and each $\varepsilon'_i$ equals 2 or -2 with probability $p$ and $1-p$, respectively. $X$ and $Y$ have identical marginals, and the common shock would seem to be more important relative to the idiosyncratic shock in the distribution of $X$ than in $Y$. Nevertheless, for any $p \neq \frac{1}{2}$, the distributions of $Y$ and $X$ cannot be ranked according to the supermodular ordering. To see this, note that all upper-orthant and lower-orthant indicator functions are supermodular, and observe that for $p > (<)\, \frac{1}{2}$, $P(X_1 \geq 3, X_2 \geq 3) < (>)\, P(Y_1 \geq 3, Y_2 \geq 3)$ and $P(X_1 \leq -3, X_2 \leq -3) > (<)\, P(Y_1 \leq -3, Y_2 \leq -3)$.

# Online Appendix

## B Complexity of the Double Description Method

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 $\mathcal{L} = \times_{i=1}^{n} \mathcal{L}_i$ of $\mathbb{R}^n$ with $|\mathcal{L}_i| = m_i$, the dimension of the vector space generated by associating a dimension to each node of $\mathcal{L}$ is $d = \prod_{i=1}^{n} m_i$. To compute the number $p$ of inequalities, first recall Proposition 2, 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 $\mathcal{L}$, which it is straightforward to calculate:

$$p = \sum_{1 \leq i < j \leq n} (m_i - 1)(m_j - 1)\Pi_{k \notin \{i,j\}} m_k.$$

Suppose, for example, that $m_i$ is exactly $m$ for each of the $n$ dimensions. Then

$$p = \frac{n(n-1)}{2}(m-1)^2 m^{n-2} \sim \frac{n(n-1)}{2} m^n \quad \text{and} \quad d = m^n.$$

Therefore, the double description method has complexity $O(\exp(m^n(n \log m + 2 \log n)))$. In practice, therefore, the inequalities characterizing the supermodular ordering can be computed via this method only for "small-size" problems. However, the "size" of a problem can be reduced by aggregating data into coarser categories, and as discussed in Section 2, aggregation of data preserves the supermodular ordering. Thus, with an appropriate degree of coarsening of categories, the double description method can be used to achieve a tractable comparison of distributions according to the supermodular ordering.

## C Proof of the Mixture Distribution Theorem (Theorem 4)

We begin with the special case of Theorem 4 where the random vectors $X$ and $Y$ have *symmetric* mixture distributions: this is the case where $A(r)$ and $B(r)$ have dimension $q \times m$ and do not vary with $r$. Denote by $\bar{A}$ (resp. $\bar{B}$) the common cumulative-sum matrix generating the $X_r$'s (resp. the $Y_r$'s). Then we seek to show that for all supermodular $w$,

$$Ew(X_1, \ldots, X_n) = \frac{1}{q} \sum_{i=1}^{q} E[w(X_1, \ldots, X_n)|\bar{A}_{i,\bullet}) \geq \frac{1}{q} \sum_{i=1}^{q} E[w(Y_1, \ldots, Y_n)|\bar{B}_{i,\bullet}) = Ew(Y_1, \ldots, Y_n), \tag{19}$$

where $\bar{A}_{i,\bullet}$ (resp. $\bar{B}_{i,\bullet}$) denotes the $i^{th}$ row of $\bar{A}$ (resp. $\bar{B}$).

Let $\bar{p} \equiv (\bar{p}_1, \ldots, \bar{p}_m)$ denote an arbitrary upper-cumulative vector corresponding to a discrete distribution on support $\{1, \ldots, m\}$. We have $\bar{p}_1 = 1$ and $\bar{p}_{k-1} \geq \bar{p}_k$ for all $k$. For any supermodular objective function $w$ on $\mathbb{R}^n$, define $\bar{w}(\bar{p})$ by

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

Using this definition, (19) can be rewritten as

$$Ew(X_1, \ldots, X_n) = \frac{1}{q} \sum_{i=1}^{q} \bar{w}(\bar{A}_{i,\bullet}) \geq \frac{1}{q} \sum_{i=1}^{q} \bar{w}(\bar{B}_{i,\bullet}) = Ew(Y_1, \ldots, Y_n). \tag{20}$$

The function $\bar{w}$ is defined on a convex lattice of $\mathbb{R}^m$ and inherits several properties from the supermodularity of $w$, as shown in the following lemma.[47] A function $h(x_1, \ldots, x_j, \ldots, x_m)$ is *componentwise convex* if, when considered as a function of just $x_j$, it is convex for each $j$, for all values of the other $m-1$ arguments.[48]

**Lemma 1** *If $w$ is supermodular, $\bar{w}$ is supermodular and componentwise convex.*

*Proof.* Changing any component $\bar{p}_k$ of $\bar{p}$ affects all of the $X_i$'s and hence has a complicated effect on $\bar{w}$. It is therefore useful to consider, as an intermediate step, a setting where each of the independent variables $X_i$ has its own upper-cumulative distribution vector $\bar{p}^i$, so $\bar{p}^i_r = P(X_i \geq r)$, $r \in \{1, \ldots, m\}$. Define

$$\hat{w}(\bar{p}^1, \ldots, \bar{p}^n) = E[w(X_1, \ldots, X_n)|\bar{p}^1, \ldots, \bar{p}^n]. \tag{21}$$

We will use the following lemma both now and in Section C.3, when we consider asymmetric distributions.

**Lemma 2** *For any supermodular $w$, $\hat{w}(\bar{p}^1, \ldots, \bar{p}^n)$ has the following properties:*

$$\frac{\partial^2 \hat{w}}{\partial \bar{p}^i_r \partial \bar{p}^i_s} = 0 \text{ for all } i \in \mathcal{N} \text{ and } r, s \in \{1, \ldots, m\},$$

$$\frac{\partial^2 \hat{w}}{\partial \bar{p}^i_r \partial \bar{p}^j_s} \geq 0 \text{ for all } i \neq j \in \mathcal{N} \text{ and } r, s \in \{1, \ldots, m\}.$$

*Proof.* The first part of the lemma is standard, and comes from the 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 $w$. By the discrete equivalent of an integration by parts,[49] we have

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

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

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

which is nonnegative, by supermodularity of $w$. ■

To conclude the proof of Lemma 1, observe that $\bar{w}(\bar{p}) = \hat{w}(\bar{p}, \ldots, \bar{p})$. Second-order derivatives of $\bar{w}$ involve only second-order derivatives of $\hat{w}$. Lemma 2 then yields the result. ■

Now suppose that the aggregate shock takes only two possible values, so both the matrices $\bar{A}$ and $\bar{B}$ have only two rows ($q = 2$). The following lemma shows how Lemma 1, in conjunction with stochastic ordering of $A$ and $A \succ_{CCM} B$, ensures that (20) holds. With $q = 2$, condition *i)* in Lemma 3 implies that $A$ is stochastically ordered, and conditions *ii)* and *iii)* are equivalent to $A \succ_{CCM} B$. (For all row-stochastic matrices, the first column of the corresponding cumulative-sum matrix has all entries equal to 1.)

**Lemma 3** *Suppose that $q = 2$ and that there exists a nonnegative vector $\varepsilon$ such that for all $k \in \{2, \ldots, m\}$, i) $\bar{A}_{2,k} \geq \bar{A}_{1,k} + \varepsilon_k$; ii) $\bar{B}_{1,k} = \bar{A}_{1,k} + \varepsilon_k$; and iii) $\bar{B}_{2,k} = \bar{A}_{2,k} - \varepsilon_k$. Then $(X_1, \ldots, X_n) \succ_{SPM} (Y_1, \ldots, Y_n)$.*

[47] The domain of $\bar{w}$ is a simplex and is clearly convex. Moreover, the inequalities $\bar{p}_1 \geq \bar{p}_2 \geq \cdots \bar{p}_m$ reduce to pairwise inequalities of the form $\bar{p}_i \geq \bar{p}_j$, and define a lattice, as is well known (Topkis, 1968, 1978).

[48] Functions that are both supermodular and componentwise convex have been studied by Marinacci and Montrucchio (2005) and by Müller and Scarsini (2012), where they are termed "ultramodular".

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

*Proof.* The function $\bar{w}$ is polynomial in $\bar{p}$ and hence twice differentiable. Moreover, by Lemma 1, it is supermodular and componentwise convex, which implies that all of its second-order derivatives are everywhere nonnegative on its domain. Letting $\bar{p}$ (resp. $\bar{q}$) denote the first (resp. second) row of $\bar{A}$, we need to show that for any $m$-vectors $\bar{p}, \bar{q}$, and $\varepsilon \geq 0$ such that $\bar{p} + \varepsilon \leq \bar{q}$ and $\varepsilon_1 = 0$, the following inequality holds:

$$\bar{w}(\bar{p}) + \bar{w}(\bar{q}) \geq \bar{w}(\bar{p} + \varepsilon) + \bar{w}(\bar{q} - \varepsilon). \tag{22}$$

Equivalently, we need to show that

$$\bar{w}(\bar{p} + \varepsilon) - \bar{w}(\bar{p}) = \int_0^1 \sum_{k=2}^{m} \bar{w}_k(\bar{p} + \alpha\varepsilon)\varepsilon_k d\alpha \leq \int_0^1 \sum_{k=2}^{m} \bar{w}_k(\bar{q} - \varepsilon + \alpha\varepsilon)\varepsilon_k d\alpha = \bar{w}(\bar{q}) - \bar{w}(\bar{q} - \varepsilon),$$

where $\bar{w}_k$ denotes the $k^{th}$ partial derivative of $\bar{w}$. Let $\delta = \bar{q} - \varepsilon - \bar{p} \geq 0$. For each $k \in \{2, \ldots, m\}$,

$$\bar{w}_k(\bar{q} - \varepsilon + \alpha\varepsilon) - \bar{w}_k(\bar{p} + \alpha\varepsilon) = \int_0^1 \sum_{\tilde{k}=2}^{m} \bar{w}_{k\tilde{k}}(\bar{p} + \alpha\varepsilon + \beta\delta)\delta_{\tilde{k}} d\beta \geq 0, \tag{23}$$

where the inequality holds since, by Lemma 1, all second-order derivatives of $\bar{w}$ are nonnegative. Summing these inequalities over $k$ from 2 to $m$ and integrating with respect to $\alpha$ from 0 to 1 then yields the result. ∎

Starting from the stochastically ordered matrix $\bar{A}$, the matrix $\bar{B}$ described in Lemma 3 is obtained by a simple transformation that shifts a small amount of weight from the stochastically dominant row (row 2) to the dominated row (row 1), in (possibly) every column except the first. Such a transformation clearly makes the rows of the cumulative-sum matrix more similar, while keeping the column sums fixed, thus reducing the importance of the aggregate shock while leaving the unconditional distribution of each variable unchanged.

The proof of Theorem 4 for the case of symmetric mixture distributions is completed by showing that, given any $A$ and $B$ such that $A$ is stochastically ordered and $A \succ_{CCM} B$, $\bar{A}$ can be converted into $\bar{B}$ through a sequence of simple transformations of the form in Lemma 3, affecting only two of the $q$ rows. We proceed in two steps, first proving the claim for the case where $B$ is stochastically ordered (Step 1) and then extending the argument to the case where $B$ is not stochastically ordered (Step 2). From (19), the unconditional expectation of any objective function $w$ is the average of the $q$ possible expected values of $w$, conditional on the realization of the aggregate shock, i.e., the average of the $q$ possible values of $\bar{w}$, as in (20). Therefore, given Lemma 1, for any supermodular $w$, each simple transformation in the sequence reduces the average value of $\bar{w}$ and hence reduces the expected value of $w$.

Since $\bar{A}$ and $\bar{B}$ are the cumulative-sum equivalents of the $q \times m$ row-stochastic matrices $A$ and $B$, $\bar{A}_{i,k}$ and $\bar{B}_{i,k}$ lie in $[0, 1]$ and are weakly decreasing in $k$. Moreover, for any cumulative-sum matrix, the first column has all entries equal to 1, so we will henceforth ignore the first column of all such matrices. $A$ stochastically ordered means that $\bar{A}_{i,k}$ is weakly increasing in $i$. $A \succ_{CCM} B$ means that for each $k$, the column vector $\bar{A}_{\bullet,k}$ majorizes the column vector $\bar{B}_{\bullet,k}$. Below, we sometimes abuse notation slightly and use the expression $\bar{A} \succ_{CCM} \bar{B}$ to mean the same thing as $A \succ_{CCM} B$.

## C.1 Step 1: Analysis when $B$ is stochastically ordered

When $B$ is stochastically ordered, $\bar{B}_{i,k}$ is weakly increasing in $i$. We first consider the case in which $\bar{B}$ has *strictly* monotonic entries across row and column indices (ignoring, as noted, above, the first column), so

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

**The case where $\bar{B}$ has strictly monotonic entries**

The proof consists in building, by induction on $k$, a sequence of matrices whose first $k$ columns are identical to those of $\bar{B}$ and such that the mixture distributions generated from them are dominated by that generated from $\bar{A}$ according to the supermodular ordering. Let $k$ denote the smallest column index such that the $k^{th}$ columns $\bar{A}_{\bullet,k}$ and $\bar{B}_{\bullet,k}$ of $\bar{A}$ and $\bar{B}$ are distinct.

**Lemma 4** *There exists a stochastically ordered cumulative-probability matrix $C$ such that i) $C_{\bullet,\tilde{k}} = \bar{B}_{\bullet,\tilde{k}}$ for all $\tilde{k} \leq k$; ii) for all $k$, $C_{\bullet,k}$ majorizes $\bar{B}_{\bullet,k}$; and iii) the mixture distribution corresponding to $C$ is SPM-dominated by that corresponding to $\bar{A}$.*

*Proof.* Let $C$ solve the optimization problem

$$\inf_{E} \sum_{i \geq 2} \left( \sum_{j \geq i} E_{j,k} \right) \tag{24}$$

subject to the following constraints:

1. $E_{i,k} \in [0,1]$ for all $i, k$;
2. $E$ satisfies row monotonicity (the entries in each row of $E$ are decreasing in the column index);
3. $E$ is stochastically ordered (the 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 $E$ majorizes the corresponding column of $\bar{B}$);
5. the mixture distribution corresponding to $E$ is SPM-dominated by that corresponding to $\bar{A}$;
6. $E_{\bullet,\tilde{k}} = \bar{B}_{\bullet,\tilde{k}}$ for all $\tilde{k} < k$.

The set of $E$'s satisfying these constraints is compact (as a closed, bounded subset of a finite dimensional space) and nonempty (since $\bar{A}$ belongs to it), and the objective (24) is continuous. Therefore, its minimum is reached by some $C$.

We will show that $C_{\bullet,k}$ is equal to $\bar{B}_{\bullet,k}$, which will prove the lemma. Suppose, by contradiction, that $C_{\bullet,k} \neq \bar{B}_{\bullet,k}$. Since $C_{\bullet,k}$ majorizes $\bar{B}_{\bullet,k}$ and $C_{\bullet,k} \neq \bar{B}_{\bullet,k}$, there must exist a row $i$ such that[50]

$$C_{i,k} \leq \bar{B}_{i,k} \quad \text{and} \quad C_{i+1,k} > \bar{B}_{i+1,k}. \tag{25}$$

We will show that it is possible to increase $C_{i,k}$ by a small amount $\varepsilon$, and decrease $C_{i+1,k}$ by the same amount and modify some other entries, in such a way that the resulting matrix $D$ satisfies all the constraints of the minimization problem (24). Such change only affects the $i+1$ partial sum of (24), and decreases it by an amount $\varepsilon$, which will yield the desired contradiction.

Let $\bar{k}$ denote the largest column index such that $C_{i+1,\tilde{k}} = C_{i+1,k}$ for all $\tilde{k} \in [k, \bar{k}]$,[51] and let $D$ denote the matrix identical to $C$ for all rows other than $i$ and $i+1$ and for all columns outside of $[k, \bar{k}]$, and such that

1. $D_{i,\tilde{k}} = C_{i,\tilde{k}} + \varepsilon$

[50]The set $\mathcal{I}(k) = \{i : \sum_{j \geq i} C_{j,k} > \sum_{j \geq i} \bar{B}_{j,k}\}$ is nonempty. Let $\bar{i} = \max \mathcal{I}(k)$. It suffices to take $i = \max\{j < \bar{i} : C_{j,k} \leq \bar{B}_{j,k}\}$.

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

2. $D_{i+1,\tilde{k}} = C_{i+1,\tilde{k}} - \varepsilon = C_{i+1,k} - \varepsilon$

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

We first check $D$ is row-monotonic for $\varepsilon$ small enough. First, $D$ inherits this property from $C$ for all rows other than $i$ and $i+1$. For row $i$, we need to check that adding $\varepsilon$ to $C_{i,k}$ does not raise it above $C_{i,k-1}$ (if $k = 1$, there is nothing to check). This comes from the fact that $C_{i,k} \leq C_{i,k-1} - \chi$, since $C_{i,k} \leq \bar{B}_{i,k} \leq \bar{B}_{i,k-1} - \chi = C_{i,k-1} - \chi$. For $i+1$, we must check that reducing $C_{i,\bar{k}}$ by some small amount does not take it below $C_{i,\bar{k}+1}$. This comes from the definition of $\bar{k}$.[52]

Second, we check that $D$ is stochastically ordered. This is clearly true for all columns outside of $[k, \bar{k}]$, where $D$ inherits this property from $C$. For columns $\tilde{k} \in [k, \bar{k}]$, we use that $C_{i,k} + \varepsilon \leq C_{i+1,k} - \varepsilon$ for all $\varepsilon \leq \chi/2$,[53] which yields the inequalities

$$D_{i,\tilde{k}} \leq D_{i,k} = C_{i,k} + \varepsilon \leq C_{i+1,k} - \varepsilon = D_{i+1,\tilde{k}}.$$

We now show that the columns of $D$ majorize those of $\bar{B}$. It suffices to check that

$$\sum_{j \geq i+1} D_{j,\tilde{k}} \geq \sum_{j \geq i+1} \bar{B}_{j,\tilde{k}} \tag{26}$$

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

$$\sum_{j \geq i+2} D_{j,\tilde{k}} = \sum_{j \geq i+2} C_{j,\tilde{k}} \geq \sum_{j \geq i+2} \bar{B}_{j,\tilde{k}} \tag{27}$$

For $\tilde{k} > k$, we have

$$D_{i+1,\tilde{k}} = C_{i+1,k} - \varepsilon \geq \bar{B}_{i+1,k} - \varepsilon \geq \bar{B}_{i+1,\tilde{k}}$$

where the last inequality holds for $\varepsilon \leq \chi$. For $\tilde{k} = k$, we have, for $\varepsilon < C_{i+1,k} - \bar{B}_{i+1,k}$ (which is strictly positive, by our choice of $i$, see (25)),

$$D_{i+1,k} = C_{i+1,k} - \varepsilon \geq \bar{B}_{i+1,k}$$

Combining this with (27) implies (26).

Finally, because rows $i$ and $i+1$ of the matrices $C$ and $D$ satisfy the assumptions of Lemma 3, the mixture distribution corresponding to $C$ SPM-dominates that corresponding to $D$.[54] By transitivity, this implies that the mixture distribution corresponding to $\bar{A}$ SPM-dominates that corresponding to $D$.

Therefore, $D$ satisfies all of the constraints of the minimization problem above and, compared to $C$, improves the objective by $\varepsilon$, thus providing the desired contradiction. ■

To conclude the proof of Step 1 of Theorem 4, it suffices to apply Lemma 4 iteratively, transforming the first column of $\bar{A}$ into that of $\bar{B}$, then the second, until $\bar{A}$ is entirely converted into $\bar{B}$.

**The case where $\bar{B}$ is not strictly monotonic**

[52] 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 $\varepsilon < \chi$ without creating negative entries.

[53] Indeed, we have $C_{i,k} \leq C_{i+1,k} - \chi$ from both inequalities of (25) and strict monotonicity of $\bar{B}$.

[54] Lemma 3 concerns matrices with only two rows. However, by construction of the mixture distribution, the objective is linearly separable in the rows of the cumulative matrix generating the distribution, and gives equal weight to each row. Therefore, Lemma 3 applies to arbitrarily many rows, as long as only two rows are changed.

When $\bar{B}$ is not strictly monotonic, we approximate $\bar{A}$ and $\bar{B}$ by a sequence of cumulative-sum matrices $\bar{A}(N), \bar{B}(N)$ with the following properties: i) $\bar{A}(N), \bar{B}(N)$ are strictly monotonic (and, in particular, stochastically ordered), with minimal increase $\chi_N = 1/N$, ii) $\bar{A}(N)$ majorizes $\bar{B}(N)$, and iii) $\bar{A}(N)$ and $\bar{B}(N)$ converge, respectively, to $\bar{A}$ and $\bar{B}$ as $N \to \infty$. The previous analysis shows that the mixture distribution corresponding to $\bar{A}(N)$ SPM-dominates that corresponding to $\bar{B}(N)$ 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 $1-(q+(m-1))/N$ where $q \times (m-1)$ are the matrix dimensions of $\bar{A}$ and $\bar{B}$,[55] and add the matrix $E(N)$ such that $E(N)_{i,j} = \frac{1}{N}(i+(m-j))$ to the scaled down matrices to obtain $\bar{A}(N)$ and $\bar{B}(N)$. By construction, and given the hypotheses on $\bar{A}$ and $\bar{B}$, these matrices are strictly increasing with minimal increase $1/N$ and have entries less than 1. Moreover, one may easily check, for each $N$, each column of $\bar{A}(N)$ still majorizes the corresponding column of $\bar{B}(N)$, since the scaling and addition operations do not affect the ranking of those partial sums.

## C.2 Step 2: Analysis when $B$ 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. $\bar{B}^{so}$ is also row monotonic. Indeed, $\bar{B}^{so}_{i,k}$ is the $i^{th}$ smallest entry in the column $\bar{B}_{\bullet,k}$. Since $\bar{B}$ is row monotonic, that entry must be larger than the $i^{th}$ smallest entry in the column $\bar{B}_{\bullet,k+1}$, which is equal to $\bar{B}^{so}_{i,k+1}$. Moreover, majorization comparisons are the same between columns of $\bar{A}$ and $\bar{B}^{so}$ as they were with $\bar{A}$ and $\bar{B}$. Therefore, $\bar{A}$ dominates $\bar{B}^{so}$ according to the cumulative column criterion and, applying the previous analysis to $\bar{A}$ and $\bar{B}^{so}$, we conclude that the mixture distribution corresponding to $\bar{A}$ SPM-dominates that corresponding to $\bar{B}^{so}$. It then suffices to show that the mixture distribution corresponding to $\bar{B}^{so}$ SPM-dominates that corresponding to $\bar{B}$.

We convert $\bar{B}^{so}$ to $\bar{B}$ by a sequence of pairwise row transformations, of the form defined in Lemma 3. To clarify the exposition of the algorithm, for each column of $\bar{B}^{so}$, we refer the cardinal values of the ordered entries, in rows $1, 2, \ldots, q$, by their ordinal values $1, 2, \ldots, q$, and we use the same cardinal-to-ordinal transformation to label the entries in each column of $\bar{B}$.[56] Starting from the last row, $q$, of $\bar{B}^{so}$, whose entries are equal to $q$ after the cardinal-to-ordinal transformation, we will move these '$q$'-labeled entries upwards, gradually, so as to position them as in $\bar{B}$. We do this by a sequence of entry permutations between rows $q$ and $i$, for $i$ starting from $q-1$ until $i$ reaches 1. This will be done so that, after the step involving rows $q$ and $i$, the rows with indices strictly below $q$ remain stochastically ordered, and the $q^{th}$ row continues to be row monotonic and to stochastically dominate each of the rows with indices strictly below $i$. This guarantees that the application of Lemma 3, at each step, is valid. Each transformation results in a matrix corresponding to a mixture distribution that is SPM-dominated by the mixture distribution corresponding to the previous matrix. By transitivity, therefore, the mixture distribution corresponding to $\bar{B}$ is SPM-dominated by that corresponding to $\bar{B}^{so}$.

Starting with rows $q$ and $q-1$, we flip entries of $\bar{B}^{so}$ for each column $j$ in which $\bar{B}_{q,j} \neq q$. The result is that some entries in the last row of the matrix are now equal to $q-1$, with the corresponding entries in row $q-1$ equal to $q$, for exactly those columns where $\bar{B}_{q,j} \neq q$. As a result, the $q$ and $q-1$ rows of $\bar{B}^{so}$ are

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

[56] For example, if the second column of $\bar{B}$ has entries $\bar{B}_{1,2} = .3$, $\bar{B}_{2,2} = .4$, and $\bar{B}_{3,2} = .1$, so that $\bar{B}^{so}_{1,2} = .1$, $\bar{B}^{so}_{2,2} = .3$, and $\bar{B}^{so}_{3,2} = .4$, then entries are converted to $\bar{B}_{1,2} = 2$, $\bar{B}_{2,2} = 3$, and $\bar{B}_{3,2} = 1$, so that $\bar{B}^{so}_{1,2} = 1$, $\bar{B}^{so}_{2,2} = 2$, and $\bar{B}^{so}_{3,2} = 3$. If there are ties, the way ties are broken does not matter, as is clear from the algorithm.

no longer stochastically ordered, but both rows still (stochastically) dominate all rows with indices less than $q-2$. The next step is to flip entries between rows $q$ and $q-2$ of the new resulting matrix, for columns in which the $q^{th}$-row entry does not match $q^{th}$-row entry of $\bar{B}$. As a result, the $q^{th}$ row now (possibly) contains entries labeled '$q-2$' while row $q-2$ row may contain some '$q-1$' entries. Notice that, i) rows $q$, $q-1$, and $q-2$ still dominate all rows with indices less than $q-3$, and ii) row $q-1$ dominates row $q-2$. Point ii) holds because row $q-2$ inherited a '$q-1$' only if row $q-1$ inherited a '$q$' entry. Proceeding systematically by decreasing, at each step, the index $i$ of the row whose entries are swapped with those of row $q$, the result after these $q-1$ steps is that the $q^{th}$ row now has the same entries as the $q^{th}$ row of $\bar{B}$, and that the first $q-1$ rows of the resulting matrix are still stochastically ordered.

The next stage of the algorithm leaves the new $q^{th}$ row untouched. In $q-2$ steps analogous to the $q-1$ steps in the first stage, it transforms row $q-1$ into row $q-1$ of $\bar{B}$; it does so while preserving at each step the stochastic ordering of the first $q-2$ rows and guaranteeing that row $q-1$ dominates rows with which it has not yet been flipped. Applying this larger algorithmic loop to each row $q-1$, $q-2, \ldots 2$, in decreasing index order, we eventually transform $\bar{B}^{so}$ into $\bar{B}$ through a sequence of steps, each of which generates a matrix corresponding to a mixture distribution that is SPM-dominated by the previous one.

Finally, we must check that each step preserves row monotonicity, that is, the property that entries in each row are weakly decreasing in the column index. This is necessary because Lemma 3 applies only to pairs of rows that satisfy this condition. Consider the first stage of the conversion from $\bar{B}^{so}$ to $\bar{B}$, which consists of a series of pairwise transformations between the $q^{th}$ row of $\bar{B}^{so}$ and its $i^{th}$ row, for $i$ decreasing from $q-1$ to 1. Let $D(i)$ denote the matrix that results after the step involving rows $q$ and $i$, and let $D = D(1)$ denote the resulting matrix at the end of this entire first stage. The submatrix of $D$ where the last row has been removed is the stochastically ordered version of the submatrix of $\bar{B}$ where the last row has been removed. In particular, the former submatrix satisfies row monotonicity. Moreover, row $j$ of $D(i)$ is identical to row $j$ of $D$ for $j \geq i$ and $j \neq q$, and is equal to the $j^{th}$ row of $\bar{B}^{so}$ for $j < i$. All rows $j$ of $D(i)$ with $j < q$ thus satisfy row monotonicity. It remains to show that row $q$ of $D(i)$ also satisfies row monotonicity. Observe that $D(i)_{q,k}$ is equal to the $i^{th}$ largest entry, $\bar{B}^{so}_{i,k}$, of $\bar{B}^{so}_{\bullet,k}$ if $D_{q,k}$ is smaller than $\bar{B}^{so}_{i,k}$, and to $D_{q,k}$ otherwise. Now consider any two consecutive columns $k-1$ and $k$. We must show that $D(i)_{q,k-1} \geq D(i)_{q,k}$. If $D(i)_{q,k} = D_{q,k}$, then we use the fact that $D_{i,k-1} \geq D_{q,k-1} \geq D_{q,k}$. If, instead, $D(i)_{q,k} = \bar{B}^{so}_{i,k}$, then we use the fact that $D(i)_{q,k-1} \geq \bar{B}^{so}_{i,k-1} \geq \bar{B}^{so}_{i,k}$. This demonstrates row monotonicity of $D(i)$, for all $i \in \{1, \ldots, q-1\}$ and, hence, the applicability of Lemma 3 for each transformation described in the algorithm above.

## C.3 Extension of the proof to asymmetric distributions

In the general case where the random vectors $X$ and $Y$ have *asymmetric* mixture distributions, the matrices $A(r)$ and $B(r)$, of dimension $q \times m_r$, vary with $r$. Now, using the function $\hat{w}(\bar{p}^1, \ldots, \bar{p}^n) = E[w(X_1, \ldots, X_n)|\bar{p}^1, \ldots, \bar{p}^n]$ defined in (21) in the proof of Lemma 1, we can write

$$\begin{aligned} Ew(X_1, \ldots, X_n) &= \frac{1}{q} \sum_{i=1}^{q} \hat{w}(\bar{A}(1)_{i,\bullet}, \ldots, \bar{A}(n)_{i,\bullet}) \\ Ew(Y_1, \ldots, Y_n) &= \frac{1}{q} \sum_{i=1}^{q} \hat{w}(\bar{B}(1)_{i,\bullet}, \ldots, \bar{B}(n)_{i,\bullet}), \end{aligned} \tag{28}$$

where $\bar{A}(r)_{i,\bullet}$ denotes the $i^{th}$ row of $\bar{A}(r)$ and $\bar{B}(r)_{i,\bullet}$ the $i^{th}$ row of $\bar{B}(r)$.

For each $r$, let $\bar{B}(r)^{so}_{i,\bullet}$ denote the $i^{th}$ row of $\bar{B}^{so}(r)$, the stochastically ordered version of $\bar{B}(r)$. The argument proceeds by first transforming $\bar{A}(1)$ into $\bar{B}(1)^{so}$, in a manner analogous to what we did for the symmetric case in Step 1. We need to check that Lemma 3 can be applied as in Step 1. To do so, pick two realizations, $i$ and $j$, of the aggregate shock, and consider the $i^{th}$ and $j^{th}$ rows of the matrices $\{\bar{A}(r)\}_{1\leq r\leq n}$. Using notation analogous to that used in inequality (22) in the proof of Lemma 3, and for $r \geq 2$ defining $\bar{p}(r) = \bar{A}(r)_{i,\bullet}$ and $\bar{q}(r) = \bar{A}(r)_{j,\bullet}$, we must check that the following inequality holds:

$$\hat{w}(\bar{p}, \bar{p}(2), \ldots, \bar{p}(n)) + \hat{w}(\bar{q}, \bar{q}(2), \bar{q}(n)) \geq \hat{w}(\bar{p}+\varepsilon, \bar{p}(2), \ldots, \bar{p}(n)) + \hat{w}(\bar{q}-\varepsilon, \bar{q}(2), \ldots, \bar{q}(n)). \tag{29}$$

To generalize the argument used to prove Lemma 3, we begin by writing

$$\begin{aligned}
\hat{w}(\bar{p}+\varepsilon, \bar{p}(2), \ldots, \bar{p}(n)) - \hat{w}(\bar{p}, \bar{p}(2), \ldots, \bar{p}(n)) &= \int_0^1 \sum_{k=2}^{m_1} \frac{\partial \hat{w}}{\partial \bar{p}^1_k}(\bar{p}+\alpha\varepsilon, \bar{p}(2), \ldots, \bar{p}(n))\varepsilon_k d\alpha \\
\hat{w}(\bar{q}, \bar{q}(2), \ldots, \bar{q}(n)) - \hat{w}(\bar{q}-\varepsilon, \bar{q}(2), \ldots, \bar{q}(n)) &= \int_0^1 \sum_{k=2}^{m_1} \frac{\partial \hat{w}}{\partial \bar{p}^1_k}(\bar{q}-\varepsilon+\alpha\varepsilon, \bar{q}(2), \ldots, \bar{q}(n))\varepsilon_k d\alpha.
\end{aligned}$$

Now let $\delta(1) = \bar{q} - \varepsilon - \bar{p}$ and $\delta(r) = \bar{q}(r) - \bar{p}(r)$ for $r \geq 2$. Analogously to (23) in the proof of Lemma 3 we have, for each $k \in \{2, \ldots, m_1\}$,

$$\begin{aligned}
\frac{\partial \hat{w}}{\partial \bar{p}^1_k}(\bar{q}-\varepsilon+\alpha\varepsilon, \bar{q}(2), \ldots, \bar{q}(n)) &- \frac{\partial \hat{w}}{\partial \bar{p}^1_k}(\bar{p}+\alpha\varepsilon, \bar{p}(2), \ldots, \bar{p}(n)) \qquad (30)\\
&= \int_0^1 \sum_{r=1}^{n} \sum_{\tilde{k}=2}^{m_r} \frac{\partial^2 \hat{w}}{\partial \bar{p}^1_k \partial \bar{p}^r_{\tilde{k}}}(\bar{p}+\alpha\varepsilon+\beta\delta(1), \beta\delta(2), \ldots, \beta\delta(n))\delta_{\tilde{k}}(r) d\beta \\
&\geq 0,
\end{aligned}$$

where the inequality follows from the fact, as established in Lemma 2, that all cross-partial derivatives of $\hat{w}$ are nonnegative. Summing (30) over $k$ from 2 to $m_1$ and integrating over $\alpha$ then shows that (29) holds.

Inequality (29) in turn ensures that, when we convert $\bar{A}(1)$ into $\bar{B}(1)^{so}$, in a manner analogous to Step 1 above, for every transformation in the sequence Lemma 3 can be applied. Therefore,

$$\sum_{i=1}^{q} \hat{w}(\bar{A}(1)_{i,\bullet}, \bar{A}(2)_{i,\bullet}, \ldots, \bar{A}(n)_{i,\bullet}) \geq \sum_{i=1}^{q} \hat{w}(\bar{B}(1)^{so}_{i,\bullet}, \bar{A}(2)_{i,\bullet}, \ldots, \bar{A}(n)_{i,\bullet}).$$

Iterating this conversion for $r = 2, \ldots, n$, we get the chain of inequalities

$$\begin{aligned}
\sum_{i=1}^{q} \hat{w}(\bar{A}(1)_{i,\bullet}, \bar{A}(2)_{i,\bullet}, \ldots, \bar{A}(n)_{i,\bullet}) &\geq \sum_{i=1}^{q} \hat{w}(\bar{B}(1)^{so}_{i,\bullet}, \bar{A}(2)_{i,\bullet}, \ldots, \bar{A}(n)_{i,\bullet}) \\
&\geq \sum_{i=1}^{q} \hat{w}(\bar{B}(1)^{so}_{i,\bullet}, \bar{B}(2)^{so}_{i,\bullet}, \ldots, \bar{A}(n)_{i,\bullet}) \\
&\geq \cdots \\
&\geq \sum_{i=1}^{q} \hat{w}(\bar{B}(1)^{so}_{i,\bullet}, \bar{B}(2)^{so}_{i,\bullet}, \ldots, \bar{B}(n)^{so}_{i,\bullet}). \qquad (31)
\end{aligned}$$

Finally, we use the algorithm described in Section C.2 to convert $\bar{B}^{so}(r)$ into $\bar{B}(r)$ for all $r$ simultaneously. Supermodularity and componentwise-convexity of $\hat{w}$ ensure that

$$\sum_{i=1}^{q} \hat{w}(\bar{B}(1)^{so}_{i,\bullet}, \bar{B}(2)^{so}_{i,\bullet}, \ldots, \bar{B}(n)^{so}_{i,\bullet}) \geq \sum_{i=1}^{q} \hat{w}(\bar{B}(1)_{i,\bullet}, \bar{B}(2)_{i,\bullet}, \ldots, \bar{B}(n)_{i,\bullet}).$$

Combining this with (31) and (28) then yields $Ew(X_1, \ldots, X_n) \geq Ew(Y_1, \ldots, Y_n)$ for all supermodular $w$.

# D Proofs for Section 4

**Proof of Theorem 5** The proof closely parallels that of Theorem 4. The following lemma plays a role analogous to that of Lemma 3 in the proof of Theorem 4.

**Lemma 5** *Suppose that $n = 2$ and that there exists a nonnegative vector $\varepsilon$ such that for all $k \in \{2, \ldots, m\}$, i) $\bar{A}_{2,k} \geq \bar{A}_{1,k} + \varepsilon_k$; ii) $\bar{B}_{1,k} = \bar{A}_{1,k} + \varepsilon_k$; and iii) $\bar{B}_{2,k} = \bar{A}_{2,k} - \varepsilon_k$. Then $(X_1, X_2) \prec_{SSPM} (Y_1, Y_2)$ and $(X_1', X_2') \prec_{SPM} (Y_1', Y_2')$.*

*Proof.* Proposition 3 implies that $(X_1, X_2) \prec_{SSPM} (Y_1, Y_2)$ if and only if $(X_1', X_2') \prec_{SPM} (Y_1', Y_2')$. We will prove that $(X_1', X_2') \prec_{SPM} (Y_1', Y_2')$. Conditions *ii)* and *iii)* in the statement of the lemma imply that the column sums of $\bar{B}$ match those of $\bar{A}$, from which it follows that the common marginal distribution of $X_1'$ and $X_2'$ matches the common marginal distribution of $Y_1'$ and $Y_2'$. Epstein and Tanny (1980) have shown that for bivariate distributions with identical marginals, supermodular dominance is equivalent to upper-orthant dominance. For any $k, l \in \{2, \ldots, m\}$,

$$\begin{aligned} 2[P(Y_1' \geq k, Y_2' \geq l) \quad &- \quad P(X_1' \geq k, X_2' \geq l)] \\ &= \quad P(Y_1 \geq k, Y_2 \geq l) + P(Y_1 \geq l, Y_2 \geq k) - P(X_1 \geq k, X_2 \geq l) - P(X_1 \geq l, X_2 \geq k) \\ &= \quad \bar{B}_{1k}\bar{B}_{2l} + \bar{B}_{2k}\bar{B}_{1l} - \bar{A}_{1k}\bar{A}_{2l} - \bar{A}_{2k}\bar{A}_{1l}. \end{aligned}$$

Substituting for $\bar{B}_{1k}$, $\bar{B}_{1l}$, $\bar{B}_{2k}$, and $\bar{B}_{2l}$ using conditions *ii)* and *iii)*, and then simplifying, yields

$$2[P(Y_1' \geq k, Y_2' \geq l) - P(X_1' \geq k, X_2' \geq l)] = \varepsilon_k[\bar{A}_{2l} - (\bar{A}_{1l} + \varepsilon_l)] + \varepsilon_l[\bar{A}_{2k} - (\bar{A}_{1k} + \varepsilon_k)]. \tag{32}$$

Condition *i)* ensures that both of the terms in square brackets in (32) are nonnegative. Hence the distribution of $(Y_1', Y_2')$ dominates that of $(X_1', X_2')$ according to upper-orthant dominance and therefore also according to the supermodular ordering. ■

The transformation in Lemma 5 converting the matrix $\bar{A}$ into $\bar{B}$ shifts a small amount of weight from the stochastically dominant row 2 to the dominated row 1, in (possibly) every column except the first. This transformation clearly makes the independent lotteries represented by the rows of the matrix more similar to one another, while keeping the column sums fixed. The lemma shows that this increasing similarity of the lotteries translates into symmetric supermodular dominance of the distribution of the lottery outcomes, or equivalently, into less negative interdependence of the symmetrized distribution of the lottery outcomes.

The proof of Theorem 5 is completed by showing that given any $n \times m$ matrices $A$ and $B$ such that $A$ is stochastically ordered and $A \succ_{CCM} B$, $\bar{A}$ can be converted into $\bar{B}$ through a sequence of simple transformations of the form in Lemma 5, affecting only two of the $n$ rows. As in the proof of Theorem 4, we proceed in two steps, first proving the claim for the case where $B$ is stochastically ordered (Step 1) and then extending the argument to the case where $B$ is not stochastically ordered (Step 2). The following lemma, combined with Lemma 5, then ensures that each simple transformation in the sequence raises the expected value of any symmetric and supermodular objective function.

**Lemma 6** *Suppose that $X$ and $Y$ are 2-dimensional random vectors such that $X \prec_{SSPM} Y$ and that $Z$ is a $p$-dimensional random vector independent of $X$ and $Y$. Then for any $p$, the $(p+2)$-dimensional random vectors $(X, Z)$ and $(Y, Z)$ satisfy $(X, Z) \prec_{SSPM} (Y, Z)$.*

*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. Therefore, $X \prec_{SSPM} Y$ implies that $r(z) \leq s(z)$ for all $z$. Since also $Z$ is independent of $X$ and $Y$, it follows that $E[w(X,Z)] = E[E[w(X,Z)|Z]] = E[r(Z)] \leq E[s(Z)] = E[E[w(Y,Z)|Z]] = E[w(Y,Z)]$. ■

**Step 1: Proof that the distribution corresponding to $\bar{A}$ is SSPM-dominated by that corresponding to $\bar{B}^{so}$.** We use the proof of Section C.1. The condition that the distribution corresponding to $\bar{A}$ SPM-dominates that corresponding to $E$ is replaced by the condition that the distribution corresponding to $\bar{A}$ is SSPM-dominated by that corresponding to $E$. The proof that the distribution generated by the constructed matrix $D$ SSPM-dominates that generated by $C$ is based on Lemma 5, instead of Lemma 3. Because each row now represents the distribution of a different random variable, and random variables are independently distributed, Lemma 6 guarantees that the result of Lemma 5 pertaining to changes to the distributions of variables $i$ and $i+1$ extends to the multivariate distributions over all $n$ random variables.

**Step 2: Proof that the distribution corresponding to $\bar{B}^{so}$ is SSPM-dominated by that corresponding to $\bar{B}$.** We use the proof of Section C.2, again replacing Lemma 3 by Lemma 5. Because each step preserves row monotonicity, as shown in Section C.2, all rows correspond to actual probability distributions. This ensures that once again, Lemma 6 can be applied at every step to extend the result of Lemma 5 to the multivariate distributions over all $n$ variables. ■

**Proposition 8** *For any row-stochastic matrix $A$ ($B$), let $X$ ($Y$) denote a random vector whose components are independently distributed and generated by the rows of $A$ ($B$). Given any $m$-dimensional probability vector $p$, and any $n$, i) there exists a unique $n \times m$ row-stochastic matrix $A$ whose $j^{th}$ column, for each $j$, sums to $np_j$, such that for all $n \times m$ row-stochastic matrices $B$ with the same column sums as $A$, $(X_1, \ldots, X_n) \prec_{SSPM} (Y_1, \ldots, Y_n)$;*
*ii) for the $n \times m$ matrix $B$ with all rows equal to the probability vector $p$ and for any stochastically ordered row-stochastic matrix $A$ whose $j^{th}$ column sums to $np_j$, $(X_1, \ldots, X_n) \prec_{SSPM} (Y_1, \ldots, Y_n)$.*

The "optimal" matrix $B$ identified by part ii) of Proposition 8 is the one in which all of the lotteries are identical. In the production context described above, for example, this corresponds to allocating resources symmetrically across tasks. The "worst" matrix $A$ identified by part i) is the one in which the stochastically ordered lotteries described by the rows are as disparate as possible, subject to their average equaling the vector $p$. The lottery represented by row $i$ assigns positive probability either to a single outcome (i.e., it is degenerate) or to a set of outcomes with adjacent (column) indices, and there is at most one outcome to which the lotteries in rows $i$ and $i+1$ both assign positive probability. In the production context described above, this matrix allocates resources to the various tasks as differently as is feasible, given the overall resource constraints.[57]

**Proof of Proposition 8**

[57]Note that in part i) of the proposition, $A$ yields a distribution that is dominated according to $\succ_{SSPM}$ by that from any other matrix with matching column sums, while in part ii), $B$ yields a distribution that is guaranteed to dominate only those from stochastically ordered matrices with matching column sums. Let $p = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$, let $B$ equal the $2 \times 3$ matrix both of whose rows match $p$, and let $A$ be the $2 \times 3$ matrix whose first row is $(\frac{1}{2}, 0, \frac{1}{2})$ and whose second row is $(0, 1, 0)$. $A$ and $B$ have matching column sums, but $A$ is not stochastically ordered. The bivariate distributions generated from $A$ and $B$ cannot be ranked according to $\succ_{SSPM}$: For $w(z_1, z_2) = I_{\{z_1 \geq 3, z_2 \geq 2\}} + I_{\{z_1 \geq 2, z_2 \geq 3\}}$, $Ew(X_1, X_2) = \frac{1}{2} > \frac{1}{4} = Ew(Y_1, Y_2)$, while for $w(z_1, z_2) = I_{\{z_1 \geq 3, z_2 \geq 3\}}$, $Ew(X_1, X_2) = 0 < \frac{1}{16} = Ew(Y_1, Y_2)$.

Proof of *i):* Assume that $p_j > 0$ for all $j \in \{1, \ldots, m\}$. (If for some $j$, $p_j = 0$, then all entries in the $j^{th}$ column of $A$ would necessarily equal 0.) Given the one-to-one mapping between row-stochastic matrices and their cumulative-column equivalents, it is sufficient to prove the existence of a unique cumulative-sum matrix $\bar{A}$ satisfying the claim.

Let $\lfloor x \rfloor$ denote the largest integer below $x$, and for a vector $v$, let $v'$ denote its transpose. Given a probability vector $(p_1, \ldots, p_m)$, define $\bar{p}_k = \sum_{j=k}^{m} p_j$. Note that $\bar{p}_1 = 1$ and $\bar{p}_k$ is strictly decreasing in $k$. Consider the cumulative-column matrix $\bar{A}$ whose first column consists of all 1's and whose $k^{th}$ column has the form $(0, \ldots, 0, \lambda_k, 1, \ldots, 1)'$, where $\lambda_k \equiv n\bar{p}_k - \lfloor n\bar{p}_k \rfloor \in [0, 1)$ and where the index of the row in which $\lambda_k$ appears is $i_k \equiv n - \lfloor n\bar{p}_k \rfloor$. Note that since $\lfloor n\bar{p}_k \rfloor$ is weakly decreasing in $k$, $i_k$ is weakly increasing in $k$.

By construction, the $k^{th}$ column of $\bar{A}$ sums to $\lambda_k + 1(\lfloor n\bar{p}_k \rfloor) = n\bar{p}_k$, as required. By construction also, all entries of $\bar{A}$ are in $[0, 1]$. To confirm that $\bar{A}$ is a valid cumulative-column matrix, we need to confirm that for each row, the entries are weakly decreasing in the column index $k$. If $i_k < i_{k+1}$, then this is clearly true, since for $i < i_k$, the entries in columns $k$ and $k+1$ are both 0, for $i = i_k$, the entry in column $k$ is $\lambda_k$ while the entry in column $k$ is 0, for $i = i_{k+1}$, the entry in column $k$ is 1 while that in column $k+1$ is $\lambda_k$, and for $i > i_{k+1}$, the entries in column $k$ and $k+1$ are both 0. If, instead, $i_k = i_{k+1}$, then we need to check that $\lambda_k \geq \lambda_{k+1}$. Now given the definition of $i_k$, $i_k = i_{k+1}$ implies that $\lfloor n\bar{p}_k \rfloor = \lfloor n\bar{p}_{k+1} \rfloor$, and since $\bar{p}_k > \bar{p}_{k+1}$, it then follows from the definition of $\lambda_k$ that $\lambda_k > \lambda_{k+1}$.

By construction, for each column $k$ of $\bar{A}$, the entries are weakly increasing in the row index, so $\bar{A}$ is stochastically ordered. Since for each $k \geq 2$, all but at most one element of column $k$ equals 0 or 1, it is clear that for each $k$, the $k^{th}$ column of $\bar{A}$ majorizes all vectors whose components lie in $[0, 1]$ and sum to $n\bar{p}_k$. Furthermore, among all such vectors, the $k^{th}$ column of $\bar{A}$ is the unique vector with increasing components which majorizes all others. Therefore, for any other cumulative-column matrix $\bar{B}$ whose $k^{th}$ column sums to $n\bar{p}_k$, $A \succ_{CCM} B$, and $\bar{A}$ is the unique matrix for which this statement is true. The claim in *part i)* then follows from Theorem 5.

Proof of *ii):* Since each row of the matrix $B$ described in part *ii)* is identical, every column of $\bar{B}$ consists of a vector with equal components. Thus, the $k^{th}$ column of $\bar{B}$ is majorized by any vector whose components lie in $[0, 1]$ and sum to $n\bar{p}_k$, so for any other cumulative-column matrix $\bar{A}$ whose $k^{th}$ column sums to $n\bar{p}_k$, we have $A \succ_{CCM} B$. With $A$ stochastically ordered, the claim in part *ii)* then follows from Theorem 5. ■

# E Proofs for Sections 5 and 6

**Proof of Proposition 5** Given an arbitrary tournament, let it be summarized by a bistochastic matrix $B$, whose $i^{th}$ row describes individual $i$'s marginal distribution over the $n$ prizes. For any symmetric ex post welfare function, the realized ex post welfare under the tournament is independent of the allocation of prizes, since by assumption, each prize must be allocated to exactly one individual. Therefore, the expected ex post welfare generated by *any* tournament is the same as that generated by the (degenerate) tournament summarized by the $n \times n$ identity matrix $I$—in this tournament, individual $i$ receives the prize of rank $i$ with probability 1. Moreover, this degenerate tournament coincides with the degenerate independent joint distribution where individual $i$ receives the prize of rank $i$ with probability 1. For proving the proposition, it is therefore sufficient to show that the independent joint distribution with marginals represented by the rows of $I$ is dominated according to the symmetric supermodular ordering by the independent joint distribution summarized by any bistochastic matrix $B$. Now the identity matrix $I$ is stochastically ordered and clearly dominates any other bistochastic matrix according to the cumulative column majorization criterion.

Theorem 5 therefore yields the result. ■

**Proof of Proposition 6 and claim following it** Define $N_c = \sum_{i=1}^{6} I_{\{Y_1=1\}}$ and $N_u = \sum_{i=1}^{6} I_{\{X_1=1\}}$: these are the total number of solvent banks in the clustered and unclustered networks, respectively. Proposition 4 implies that $(Y_1, \ldots, Y_6) \succ_{SSPM} (X_1, \ldots, X_6)$ if and only if the distribution of $N_c$ dominates that of $N_u$ according to the univariate convex ordering, which we will write as $N_c \succ_{CX} N_u$. $N_c \succ_{CX} N_u$ if and only if the distribution of $N_c$ is derivable from that of $N_u$ by a sequence of mean-preserving spreads.

Suppose first that the common default threshold for banks, $d$, satisfies $d \in [L, \frac{2L+H}{3})$, so a bank defaults if and only if all three projects in its portfolio fail. We will show that for each $k \in \{0, \ldots, 6\}$, conditional on $k$ of the 6 projects succeeding, we have $N_c \succ_{CX} N_u$. Since these conditional distributions are independent of $p$, the probability that any given project succeeds, it will then follow that for all $p$, $N_c \succ_{CX} N_u$ holds unconditionally. For each $k \in \{0, 1, 4, 5, 6\}$, the conditional distributions of $N_c$ and $N_u$ are degenerate and equal. Conditional on $k = 3$, i) $N_c = 3$ if all three banks whose projects fail are in the same cluster (probability $\frac{1}{10}$) and $N_c = 6$ otherwise; and ii) $N_u = 5$ if the three banks whose projects fail are adjacent to one another in the circle (probability $\frac{3}{10}$) and $N_u = 6$ otherwise. Hence, conditional on $k = 3$, the distribution of $N_c$ is a mean-preserving spread of that of $N_u$. Conditional on $k = 2$, i) $N_c = 3$ if three of the four banks whose projects fail are in the same cluster (probability $\frac{2}{5}$) and $N_c = 6$ otherwise; and ii) $N_u$ takes the values $4, 5, 6$ if the two banks whose projects succeed are adjacent, separated by one bank, and opposite one another in the circle, respectively; these three events occur with respective probabilities $\frac{2}{5}, \frac{2}{5}, \frac{1}{5}$. Thus, conditional on $k = 2$, the distribution of $N_c$ is a mean-preserving spread of that of $N_u$. It follows that, for $d \in [L, \frac{2L+H}{3})$ and any $p$, $N_c \succ_{CX} N_u$ holds unconditionally.

Now suppose $d \in [\frac{L+2H}{3}, H)$, so a bank is solvent if and only if all three projects in its portfolio succeed. This case is the mirror image of the case where $d \in [L, \frac{2L+H}{3})$, with "solvent" replacing "defaulting" and $6 - N_c$ and $6 - N_u$ replacing $N_c$ and $N_u$, respectively. Hence, the arguments above immediately imply that for all $p$, $6 - N_c \succ_{CX} 6 - N_u$, which is equivalent to $N_c \succ_{CX} N_u$.

Finally, suppose $d \in [\frac{2L+H}{3}, \frac{L+2H}{3})$, so a bank defaults if two or three of the projects in its portfolio fail. For each $k \in \{0, 1, 3, 5, 6\}$, the conditional distributions of $N_c$ and $N_u$ are degenerate and equal. Conditional on $k = 4$, the distributions of $N_c$ and $N_u$ for the current default threshold match, respectively, the distributions of $N_c$ and $N_u$, conditional on $k = 2$, for $d = L$. Finally, for the current default threshold (under which project successes and failures are mirror images in terms of their effect on bank default), the distributions of $6 - N_c$ and $6 - N_u$ conditional on $k = 2$ match, respectively, the distributions of $N_c$ and $N_u$ conditional on $k = 4$. Hence, for $d \in [\frac{2L+H}{3}, \frac{L+2H}{3})$ and for any $p$, $N_c \succ_{CX} N_u$ holds unconditionally. This completes the proof of Proposition 6.

■

**Proof of Theorem 6** For the "only if" part, choose any coarsening $\tilde{\mathcal{L}}$ and supermodular function $\tilde{w}$ on $\tilde{\mathcal{L}}$. The function $w$ on $\mathcal{L}$ defined by $w(x) = \tilde{w}(\tilde{x}(x))$, where $\tilde{x}(x)$ is the hyperrectangle containing $x$, is also supermodular. Therefore, $E[w|G] \geq E[w|F]$. Equivalently, $E[\tilde{w}|\tilde{G}] \geq E[\tilde{w}|\tilde{F}]$. Since the inequality holds for any $\tilde{w}$, we conclude that $\tilde{G} \succ_{SPM} \tilde{F}$.

For the "if" part, consider, for any $N > 1$, the coarsening $\mathcal{L}(N)$ of $\mathcal{L}$ in which each $\mathcal{L}_i$ is partitioned into $N$ intervals of equal length. Given any supermodular function $w$ on $\mathcal{L}$, let $w_N, F_N, G_N$ denote the coarsened versions of $w, F, G$ on $\mathcal{L}(N)$. We first show that $w_N$ is supermodular. For any $\tilde{x} \in \mathcal{L}(N)$ and dimensions $i, j$ such that $\tilde{x} + e_i + e_j$ belongs to $\mathcal{L}(N)$, we must show that

$$w_N(\tilde{x}) + w_N(\tilde{x} + \tilde{e}_i + \tilde{e}_j) \geq w_N(\tilde{x} + \tilde{e}_i) + w_N(\tilde{x} + \tilde{e}_j). \tag{33}$$

Given the equal spacing of the chosen partition, the denominator arising in (14) is the same for all $\tilde{x}$'s. Therefore, (33) reduces to showing that[58]

$$\int_{x\in\tilde{x}} (w(x) + w(x + d_i + d_j) - w(x + d_i) - w(x + d_j))dx \geq 0,$$

where $d_i = |\mathcal{L}_i|/N$ is the length of each hyperrectangle along dimension $i$ (and similarly for $d_j$). The inequality holds by supermodularity of $w$, which proves that $w_N$ is supermodular. As a result, $E[w_N|G_N] \geq E[w_N|F_N]$ for all $N$. It remains to show that $E[w_N|F_N]$ converges to $E[w|F]$ as $N \to \infty$. We have

$$E[w_N|F_N] - E[w|F] = \sum_{\tilde{x}\in\mathcal{L}_N} \int_{x\in\tilde{x}} (w_N(\tilde{x}) - w(x))f(x)dx.$$

By construction, $\int_{x\in\tilde{x}} w(x)dx = \int_{x\in\tilde{x}} w_N(\tilde{x})dx$. Therefore, letting $\chi(\tilde{x})$ denote any element of $\tilde{x}$,

$$\left|\int_{x\in\tilde{x}} (w_N(\tilde{x}) - w(x))f(x)\right| = \int_{x\in\tilde{x}} |(w_N(\tilde{x}) - w(x))(f(x) - f(\chi(\tilde{x})))|dx. \tag{34}$$

Fix $\varepsilon > 0$. The density $f$ of $F$ is continuous, and hence uniformly continuous on the compact domain $\mathcal{L}$. Therefore, there exists $\bar{N}$ such that for all $N > \bar{N}$, $|f(x) - f(y)| < \varepsilon$ for all $x, y$ of $\mathcal{L}$ belonging to the same hypercube of $\mathcal{L}(N)$. This, combined with (34), implies that

$$\left|\int_{x\in\tilde{x}} (w_N(\tilde{x}) - w(x))f(x)\right| < \varepsilon \int_{x\in\tilde{x}} (|w_N(\tilde{x})| + |w(x)|)dx.$$

Integrating over $\mathcal{L}(N)$, we get

$$|E[w_N|F_N] - E[w|F]| < \varepsilon(\|w\|_1 + \|w_N\|_1).$$

It remains to show that $\|w_N\|_1$ is bounded above, uniformly in $N$. This is implied by

$$\|w_N\|_1 = \sum_{\tilde{x}\in\mathcal{L}(N)} |w_N(\tilde{x})| \leq \sum_{\tilde{x}\in\mathcal{L}(N)} \int_{x\in\tilde{x}} |w(x)|dx = \|w\|_1 < \infty.$$

■

**Proof of Proposition 7** Theorem 1 implies that, for two distributions $F$ and $G$ to be comparable according to the supermodular ordering, they must have identical marginals: $F_i = G_i$ for all $i$. This in turn implies that the domain $\tilde{\mathcal{L}}$ is the same for both copulas $C_F$ and $C_G$. As is easily shown, $\tilde{\mathcal{L}} = \times_i \tilde{\mathcal{L}}_i$, where $\tilde{\mathcal{L}}_i = \{F_i(x_i) : x_i \in \mathcal{L}_i\}$, so $\tilde{\mathcal{L}}$ is a lattice. By definition, the $F_i$'s are nondecreasing. Moreover, without loss of generality, we can assume that for each $i$, each level $x_i$ is achieved with positive probability (otherwise, we can simply remove that level from the support $\mathcal{L}_i$), hence the $F_i$'s are strictly increasing from $\mathcal{L}_i$ to $\tilde{\mathcal{L}}_i$. Now define $\tilde{X}_i \equiv F_i(X_i)$ and $\tilde{Y}_i \equiv G_i(Y_i)(= F_i(Y_i))$. As observed in section 2, this implies that $X \prec_{SPM} Y$ if and only if $\tilde{X} \prec_{SPM} \tilde{Y}$. Finally, observe from the definition of a copula in (15) that the joint distributions of $\tilde{X}$ and $\tilde{Y}$ on $\tilde{\mathcal{L}}$ coincide, respectively, with the copulas $C_F$ and $C_G$. This proves the result. ■

[58] Because the distributions $F$ and $G$ are absolutely continuous, it is not necessary to specify in which elements of the partition the boundaries of these elements are located.

# References


Acharya, V.V., Pedersen, L.H., Philippon, T., and Richardson, M. (2010) "Measuring Systemic Risk," N.Y.U. W.P.

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

Alkire, S., and Foster, J. (2011) "Counting and Multidimensional Poverty Measurement," *J. of Public Economics*, 95, 476-87.

Allen, F., Babus, A., and Carletti, E. (2012) "Asset Commonality, Debt Maturity and Systemic Risk," *J. of Financial Economics*, 104, 519-34.

Athey, S., and Levin, J. (2001) "The Value of Information in Monotone Decision Problems", draft, Stanford University.

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

Atkinson, A. B. (2003) "Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches," *J. of Economic Inequality*, 1, 51-65.

Avis, D., and Fukuda, K. (1982) "A Pivoting Algorithm for Convex Hulls and Vertex Enumeration of Arrangements and Polyhedra," *Discrete and Computational Geometry*, 8, 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., 20-8.

Baldiga, K., and Green, J.R. (2013) "Assent-Maximizing Social Choice," *Social Choice and Welfare*, 40, 439–60.

Beale, N., Rand, D., Battey, H., Croxson, K., May, R., and Nowak, M. (2011) "Individual versus Systemic Risk and the Regulator's Dilemma," *Proc. of the National Academy of Sciences*, 108, 12647-52.

Beare, B. (2010) "Copulas and Temporal Dependence," *Econometrica*, 78, 395–410.

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

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

Boland, P., and Proschan, F. (1988) "Multivariate Arrangement Increasing Functions with Applications in Probability and Statistics," *J. Multivariate Anal.*, 25, 286–98.

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

Bonhomme, S., and Robin, J.-M. 2006 "Modeling Individual Earnings Trajectories Using Copulas: France, 1990-2002," *Contributions to Economic Analysis*, 275, 441-78.

Burtschell, X., Gregory, J., and J.-P. Laurent (2009) "A Comparative Analysis of CDO Pricing Models under the Factor Copula Framework," *J. of Derivatives*, 16, 9-37.

Caillaud, B., and Tirole, J. (2007) "Consensus Building: How to Persuade a Group," *American Economic Review*, 97, 1877–900.

Chetty, R., Hendren, N., Kline, P., Saez, E., and Turner, N. (2014) "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility," *American Econ. Review*, 104, 141-47.

Chew, S.H., and Sagi, J. (2012) "An Inequality Measure for Stochastic Allocations," *J. of Economic Theory*, 147, 1517-44.

Cousin, A., and Laurent, J.-P. (2008) "Comparison Results for Exchangeable Credit Risk Portfolios," *Insurance: Mathematics and Economics*, 42, 1118–27.

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

Dwielwulski, P., and Quah, J. "Testing for Production with Complementarities," Oxford Dept. of Economics Disc. Paper 722, 2014.

Embrechts, P. (2009) "Copulas: A Personal View," *J. of Risk and Insurance*, 76, 639–50.

Epstein. L., and Tanny, S. (1980) "Increasing Generalized Correlation: A Definition and Some Economic Consequences," *Canadian J. of Economics*, 13, 6–34.

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

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

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

Fukuda, K., and Prodon, A. (1996) "Double Description Method Revisited." In *Combinatorics and Computer Science*, *Lecture Notes in Computer Science*, 1120, 91–111. Springer.

Gajdos, T., and Maurin, E. (2004) "Unequal Uncertainties and Uncertain Inequalities: An Axiomatic Approach," *J. of Economic Theory*, 116, 3–118.

Genest, C., and Verret, F. (2002) "The $TP_2$ Ordering of Kimeldorf and Sampson has the Normal-Agreeing Property," *Statistics and Probability Letters*, 57, 387-391.

Giovagnoli, A., and Wynn, H.P. (2008) "Stochastic Orderings for Discrete Random Variables," *Statistics and Probability Letters*, 78, 28272835.

Gleser, L. (1975) "On the Distribution of the Number of Successes in Independent Trials," *Annals of Probability*, 3, 182–8.

Gollier, C. (2011) "Discounting, Inequalities and Economic Convergence", mimeo, Toulouse School of Economics.

Goldman, A. J., and Tucker, A. W. "Polyhedral Convex Cones". In *Linear inequalities and related systems*, Ann. of Mathematics Studies 38, eds. H. W. Kuhn and A. W. Tucker. Princeton Univ. Press, 19–39.

Gendron-Saulnier, C., and Gordon, S. "Information Choice and Diversity: The Role of Strategic Complementarities", Working Paper, 2014.

Hardy, G.H., Littlewood, J.E., and Polya, G. (1952) *Inequalities*, Cambridge: Cambridge University Press, 2nd edition (1st edition 1934).

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

Hoeffding, W. (1956) "On the Distribution of the Number of Successes in Independent Trials," *Annals of Mathematics and Statistics*, 27, 713–21.

Holland, P. W., and Rosenbaum, P. R. (1986) "Conditional Association and Unidimensionality in Monotone Latent Variable Models," *Annals of Statistics*, 14, 1523–43.

Hu, T., and Yang, J. (2004) "Further Developments on Sufficient Conditions for Negative Dependence of Random Variables," *Statistics and Probability Letters*, 66, 369–81.

Joag-Dev, F., and Proschan, F. (1983) "Negative Association of Random Variables, with Applications," *Annals of Statistics*, 11, 286-95.

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

Joe, H. (1997) *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

Jogdeo, K. (1978) "On a Probability Bound of Marshall and Olkin," *Annals of Statistics*, 6, 232–34.

Karlin, S., and Novikoff, A. (1963) "Generalized Convex Inequalities," *Pacific J. of Mathematics*, 13, 1251–79.

Levy, H., and Paroush, J. (1974) "Toward Multivariate Efficiency Criteria," *J. of Economic Theory*, 7, 129–42.

Li, D. (2000) "On Default Correlation: A Copula Function Approach," *J. of Fixed Income*, 9, 43–54.

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

Marinacci, M., and Montrucchio, L. (2005) "Ultramodular Functions," *Mathematics of Operations Research*, 30, 311–332.

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 D. P. No. 49.

Meyer, M. A., and Mookherjee, D. (1987) "Incentives, Compensation and Social Welfare," *Review of Economic Studies*, 45, 209–26.

Meyer, M.A., and Strulovici, B. (2012) "Increasing Interdependence of Multivariate Distributions," *J. of Economic Theory*, 147, 1460–89.

Meyer, M.A., and Zeng, Q.C. (2013) "Matching with Informational Frictions," draft, Nuffield College, Oxford.

Milgrom, P., and Roberts, J. (1990) "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *American Economic Review*, 80, 511–528

Milgrom, P., and Roberts, J. (1995) "Complementarities and Fit Strategy, Structure, and Organizational Change in Manufacturing," *Journal of Accounting and Economics*, 19, 179–208.

Moldovanu, B., and Shi, X. (2013) "Specialization and Partisanship in Committee Search," *Theoretical Economics*, 8, 751-774.

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

Müller, A. (2013) "Duality Theory and Transfers for Stochastic Order Relations". In H. Li and X. Li eds., Stochastic Orders in Reliability and Risk, Springer, 41-57.

Müller, A., and Scarsini, M. (2000) "Some Remarks on the Supermodular Order," *J. of Multivariate Analysis*, 73, 107-19.

Müller, A., and Scarsini, M. (2012) "Fear of Loss, Inframodularity, and Transfers," *J. of Economic Theory*, 147, 1490-500.

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

Myatt, D., and Wallace, C. (2012) "Endogenous Information Acquisition in Coordination Games," *Review of Economic Studies*, 79, 340-74.

Patton, A. (2009) "Are 'Market Neutral' Hedge Funds Really Market Neutral?," *Review of Financial Studies*, 22, 2495–530.

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

Promislow, S.D., and Young, V.R. (2005) "Supermodular Functions on Finite Lattices," *Order*, 22, 389413.

Rothschild, M., and Stiglitz, J. E. (1970) "Increasing Risk: I. A Definition," *J. of Economic Theory*, 2, 225–43.

Saito, K. (2013) "Social Preferences under Risk: Equality of Opportunity versus Equality of Outcome," *American Economic Review*, 103, 3084-3101.

Shaked, M., and Shanthikumar, J.G. (1997) "Supermodular Stochastic Orders and Positive Dependence of Random Vectors," *J. of Multivariate Analysis*, 61, 86–101.

Sklar, A. (1959) "Fonctions de Répartition à $n$ Dimensions et Leurs Marges," *Publication de l'Institut de Statistique de l'Université de Paris*, 8, 229-31.

Tchen, A.H. (1980) "Inequalities for Distributions with Given Marginals," *Annals of Probability*, 8, 814-27.

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

Topkis, D.M. (1978) "Minimizing a Submodular Function on a Lattice," *Operations Research*, 26, 5-31.

Torrisi, F., and Baotic, M. (2005) "CDDMEX - Matlab Interface for the CDD Solver." Available at http://control.ee.ethz.ch/hybrid/cdd.php.

Wei, G., and Hu, T. (2002) "Supermodular Dependence Orderings on a Class of Multivariate Copulas," *Statistics and Probability Letters*, 57, 375–85.

Ziegler, G. (1997) *Lectures on polytopes*, Springer-Verlag.