

The Supermodular Stochastic Ordering

Margaret Meyer

Bruno Strulovici*

First draft: Feb. 8, 2010; this version: April 18, 2013

Abstract

In many economic applications involving comparisons of multivariate distributions, supermodularity of an objective function is a natural property for capturing a preference for greater interdependence. One multivariate distribution dominates another according to the *supermodular stochastic ordering* if it yields a higher expectation than the other for all supermodular objective functions. We prove that this ordering is equivalent to one distribution being derivable from another by a sequence of elementary, bivariate, interdependence-increasing transformations, and develop methods for determining whether such a sequence exists. For random vectors resulting from common and idiosyncratic shocks, we provide non-parametric sufficient conditions for supermodular dominance. Moreover, we characterize the orderings corresponding to supermodular objective functions that are also increasing or symmetric. We use the symmetric supermodular ordering to compare distributions generated by heterogeneous lotteries. Applications to welfare economics, committee decision-making, insurance, finance, and parameter estimation are discussed.

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

1 Introduction

In many economic contexts, it is of interest to know whether one set of random variables displays a greater degree of interdependence than another. The stochastic dominance approach expresses

*We are grateful for valuable comments from David Cox, Ian Jewitt, Paul Klemperer, and seminar participants at Montréal, Oslo, Oxford Statistics and Oxford Economics, Wisconsin, Yale, the Risk, Uncertainty, and Decision Conference, the Third Transatlantic Theory Workshop, and the European Summer Symposium in Economic Theory. Strulovici gratefully acknowledges financial support from the NSF (Grant No. 1151410). Meyer: Nuffield College and Dept. of Economics, Oxford University, Oxford OX1 1NF, UK, margaret.meyer@nuffield.ox.ac.uk; Strulovici: Dept. of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA, b-strulovici@northwestern.edu.

attitudes towards interdependence through properties of objective functions whose expectations are used to evaluate distributions. Since the expected values of additively separable objective functions depend only on marginal distributions, attitudes towards interdependence must be represented through non-separability properties. We argue that supermodularity (Topkis, 1968, 1978) of an objective function is a natural property with which to capture a preference for greater interdependence. Supermodularity of a function captures the idea that its arguments are complements, not substitutes: When an increasing function of two or more variables is supermodular and the values of any two variables are increased together, the resulting increase in the function is larger than the sum of the increases that would result from increasing each of the values separately. Our main objective in this paper is to characterize the partial ordering on distributions of n -dimensional random vectors which is equivalent to one distribution's yielding a higher expectation than another for all supermodular objective functions. Following the statistics literature, we refer to this partial ordering as the “supermodular stochastic ordering” (Shaked and Shanthikumar, 1997).

There are many branches of economics where the supermodular stochastic ordering is a valuable tool for comparing distributions with respect to their degree of interdependence. We describe applications of our methods and results to the assessment of i) ex post inequality under uncertainty; ii) multidimensional deprivation; iii) the equilibrium duration of search by committees, iv) the dependence among claims in a portfolio of insurance policies or among assets in a financial institution's portfolio; v) systemic risk in financial systems; and vi) the richness of datasets for parameter estimation. Our approach also permits a non-parametric comparison of copulas.

For the special case of two-dimensional random vectors, the economics and statistics literatures have provided a complete characterization of the supermodular ordering. Specifically, Levy and Paroush (1974), Epstein and Tanny (1980), and Tchen (1980) have shown that one bivariate distribution dominates another according to the supermodular ordering if and only if the first distribution dominates the second in the sense of both upper-orthant and lower-orthant dominance.¹ This equivalence breaks down for three or more dimensions (Joe, 1990, and Müller and Scarsini, 2000). In general, the supermodular ordering is strictly stronger than the combination of upper-orthant and lower-orthant dominance.

Focusing on random vectors with supports on a finite lattice, we characterize the supermodular ordering for an arbitrary number of dimensions. Section 3 proves (Theorem 1) that one distribution is preferred to the other by every supermodular objective function if and only if the first distribution can be derived from the other by a sequence of nonnegative “elementary transformations.” Our elementary transformations play a role similar to the mean-preserving spreads defined by Rothschild and Stiglitz (1970) for univariate distributions to capture the notion of increased riskiness.

¹A multivariate distribution G dominates another distribution F according to upper-orthant (respectively, lower-orthant) dominance if for any vector z , a random vector distributed according to G has a higher probability of being above (respectively, below) z in each component than a vector distributed according to F . The ordering corresponding to the combination of upper-orthant and lower-orthant dominance is known as the concordance ordering (Joe, 1990).

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

Our restrictive definition of elementary transformations allows a very simple proof of the known characterization of the supermodular ordering for bivariate distributions. Our simple proof is based on the observation that, for any pair of bivariate distributions with identical marginals, if we allow elementary transformations to have weights of arbitrary sign, then there is a unique weighted sequence of such transformations that converts one distribution into the other.

For three or more dimensions, even with our restrictive definition of elementary transformations, there are many weighted sequences of such transformations that convert one distribution into the other. How, then, can we determine whether g dominates f according to the supermodular ordering? We introduce two different methods. One is to formulate a linear program such that the optimum value of the program is zero if and only if there exist non-negative weights on elementary transformations that will convert f to g . An alternative method, based on Minkowski’s and Weyl’s representation theorems for polyhedral cones, allows us to compute once and for all, for any given support, a minimal set of inequalities that characterize the supermodular ordering. This method can be used for optimization problems such as mechanism design, where each mechanism or policy generates a multivariate distribution, and the set of mechanisms to be compared is large.

In some applications, it is natural to assume that objective functions are not only supermodular but also increasing in their arguments. Theorem 2 demonstrates that comparison of two distributions according to the increasing supermodular ordering can be decomposed into a two-step comparison, 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. In section 3.5, we prove that the supermodular ordering on a continuous support can be characterized in terms of the supermodular ordering on all discretizations of the support, provided that the multivariate distributions have continuous densities.

One important class of interdependent random vectors are those generated by both common and

idiosyncratic shocks. Section 4 studies precisely this class. 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 i.i.d. variables be assessed? Our Theorem 5 answers both questions. We use the supermodular ordering to compare interdependence, and we present easily checkable sufficient conditions (which can be shown to be tight) on the structure of mixture distributions for two such distributions to be rankable according to the supermodular ordering. Our sufficient conditions thus provide a useful non-parametric ordering of the relative size of common vs. idiosyncratic shocks.

In some applications, it is natural to focus on objective functions that are symmetric. In Meyer and Strulovici (2012), we study the *symmetric supermodular ordering*, which corresponds to one distribution’s generating a higher expected value than another for all symmetric supermodular objective functions. In particular, we show that two distributions are ranked according to the symmetric supermodular ordering if and only if the “symmetrized” versions of the distributions are ranked according to the supermodular ordering. For the class of n -dimensional random vectors representing n independent lotteries, we identify in Theorem 6 sufficient conditions for symmetric supermodular dominance and show that these conditions have a natural interpretation in terms of lower dispersion among one set of lotteries than another, holding fixed the average of the lotteries. At a mathematical level, moreover, these sufficient conditions (which can be shown to be tight) are very closely related to the sufficient conditions identified in Theorem 5 for supermodular dominance of mixture distributions, and the proofs of Theorems 5 and 6 are likewise very similar.

Section 7 discusses a wide range of applications of the supermodular ordering. Section 7.1 focuses on two applications in welfare economics: it shows how the ordering and Theorem 6 can be applied to make comparisons of inequality in the presence of uncertainty and to compare multidimensional distributions of economic status. Section 7.2 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 7.3 applies the symmetric supermodular ordering to examine how the degree of systemic risk in banking networks depends on the structure of the interconnections among banks. Finally,

section 7.4 considers an application of the supermodular ordering to prediction and parameter estimation, showing how the ordering may be used to compare the “richness” of datasets.

Section 8 presents a brief conclusion. All proofs not in the text are in the Appendix.

2 General Setting

We consider multivariate distributions with n variables and identical, finite support. The i^{th} variable takes values in \mathcal{L}_i which is a totally ordered set with m_i elements. The Cartesian product $\times_i \mathcal{L}_i$ is denoted \mathcal{L} and is endowed with the usual partial order: $x \leq y$ if and only if $x_i \leq y_i$ for all $i \in \mathcal{N} = \{1, \dots, n\}$.

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 $(1, 0) + e_2 = (0, 0) + e_1 + e_2 = (1, 1)$.

Lattice vs. Vector Structures. The lattice structure of \mathcal{L} and its partial order are used to compare distributions. In particular, supermodularity of objective functions is defined with respect to that partial order. One may label the $d = \prod_{i=1}^n m_i$ elements (or “nodes”) of \mathcal{L} and view real functions on \mathcal{L} as vectors of \mathbb{R}^d , where each coordinate of the vector corresponds to the value of the function at a specific node of \mathcal{L} . This representation will prove particularly important for our dual characterizations of interdependence relations. A multivariate distribution whose support is \mathcal{L} (or a subset of \mathcal{L}) can be represented as an element of the unit simplex Δ_d of \mathbb{R}^d .

Orderings of Multivariate Distributions. For any function $w : \mathcal{L} \rightarrow \mathbb{R}$ 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:

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

3 The Supermodular Stochastic Ordering

Supermodular Functions and Elementary Transformations For any $x, y \in \mathcal{L}$, $x \wedge y$ denotes the component-wise minimum (or “meet”) of x and y , i.e., the element of \mathcal{L} such that $(x \wedge y)_i = \min\{x_i, y_i\} \in \mathcal{L}_i$ for all $i \in \mathcal{N}$. Let $x \vee y$ similarly denote the component-wise maximum (or “join”) of x, y . A function w is *supermodular* (on \mathcal{L}) if $w(x \wedge y) + w(x \vee y) \geq w(x) + w(y)$ for all $x, y \in \mathcal{L}$. If, for all $x, y \in \mathcal{L}$, the reverse inequality holds, the function w is *submodular*.

The set of all supermodular functions is denoted \mathcal{S} . The **supermodular stochastic ordering** is a partial order, denoted \prec_{SPM} , on the set of distributions over \mathcal{L} , and is defined as follows:

$$f \prec_{SPM} g \iff \forall w \in \mathcal{S}, \quad E[w|f] \leq E[w|g] \quad (2)$$

If $f \prec_{SPM} g$, we will say that distribution g is *more interdependent* than distribution f . For random vectors X and Y with distributions f and g and cumulative distributions F and G , respectively, we will use the expressions $X \prec_{SPM} Y$, $f \prec_{SPM} g$, and $F \prec_{SPM} G$ interchangeably.

To characterize this ordering, we introduce a class of elementary transformations which capture the notion of “increasing interdependence”, analogously to the way that Rothschild and Stiglitz’s (1970) mean-preserving spreads capture the notion of “increasing riskiness”.

For any $x \in \mathcal{L}$ such that $x + e_i + e_j \in \mathcal{L}$, let $t_{i,j}^x$ denote the function defined on \mathcal{L} by

$$t_{i,j}^x(x) = t_{i,j}^x(x + e_i + e_j) = 1, \quad t_{i,j}^x(x + e_i) = t_{i,j}^x(x + e_j) = -1, \quad (3)$$

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

If two 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 α . The α -weighted elementary transformation raises the probability of nodes x and $x + e_i + e_j$ by the common amount α , 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.

If, for example, $\mathcal{L} = \{0, 1, 2\}^2$, there are four elementary transformations, corresponding to the four values of x , $(0, 0)$, $(1, 0)$, $(0, 1)$, and $(1, 1)$, such that $x + e_i + e_j$ belongs to \mathcal{L} . For $\mathcal{L} = \{0, 1\}^3$, there are six elementary transformations, one corresponding to each face of the unit cube. Observe that our definition of elementary transformations confines attention to transformations that i) affect only *two* of the n dimensions (as illustrated by the example of $\mathcal{L} = \{0, 1\}^3$) and ii) affect values only at four *adjacent* points in the lattice, x , $x + e_i$, $x + e_j$, and $x + e_i + e_j$ (as illustrated by $\mathcal{L} = \{0, 1, 2\}^2$).

THEOREM 1 (DUAL CHARACTERIZATION) *$f \prec_{SPM} g$ 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 ,*

$$g = f + \sum_{t \in \mathcal{T}} \alpha_t t. \quad (4)$$

Proof. Supermodular functions are characterized by the property (Topkis, 1968, 1978) that

$$w \in \mathcal{S} \iff w(x + e_i + e_j) + w(x) \geq w(x + e_i) + w(x + e_j) \quad (5)$$

for all $i \neq j$ and $x \in \mathcal{L}$ such that $x + e_i + e_j \in \mathcal{L}$. Equivalently, supermodularity is characterized by the set of equations

$$w \in \mathcal{S} \iff w \cdot t \geq 0 \quad \forall t \in \mathcal{T}. \quad (6)$$

Equation (4) 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 (6), \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, $f \prec_{SPM} g$ if and only if $g - f \in \mathcal{C}(\mathcal{T})$. ■

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

Another implication of Theorem 1 is given by the following:

COROLLARY 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.*

Proof. Suppose that $f \neq g$. Then Theorem 1 implies that at least one α_t in (4) must be strictly positive. Let t_{ij}^z denote a t such that $\alpha_t > 0$. For the supermodular function $w(x) = x_i x_j$, the inequality in (5) is strict for all x , so $w \cdot t_{ij}^z > 0$ 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. ■

3.1 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. For example, the class of objective functions that are both supermodular and increasing incorporates a preference for greater interdependence as well as for higher values of each argument. We now characterize the increasing supermodular ordering.

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 τ_i^x denote the function on \mathcal{L} such that $\tau_i^x(x) = -1$, $\tau_i^x(x + e_i) = 1$, and τ_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

$$g \succ_{FOSD} f \iff w \cdot g \geq w \cdot f \quad \forall w \in \mathcal{I}. \quad (7)$$

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

$$g = f + \sum_{\tau \in \mathcal{U}} \beta_\tau \tau. \quad (8)$$

The *increasing supermodular ordering* (denoted \succ_{ISPM}) is defined as follows:

$$g \succ_{ISPM} f \iff w \cdot g \geq w \cdot f \quad \forall w \in \mathcal{S} \cap \mathcal{I}.$$

Since the functions w are now required to be increasing, $g \succ_{ISPM} f$ (in contrast to $g \succ_{SPM} f$) does not imply that g and f have identical marginals. Rather, $g \succ_{ISPM} f$ implies that each marginal distribution of g dominates the corresponding marginal distribution of f according to first-order stochastic dominance: this can be seen by taking, for each $i \in \{1, \dots, n\}$ and each $k_i \in \mathcal{L}_i$, $w(z) = I_{\{z_i \geq k_i\}}$, which is both increasing and supermodular.

Theorem 2 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 $\mathcal{L}_i = \{0, 1, \dots, m_i - 1\}$. Given two distributions f and g with $\delta \equiv g - f$, define the function γ 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 $i \in \{1, \dots, n\}$ and $k \in \{1, 2, \dots, m_i - 1\}$, let

$$\gamma(ke_i) = Pr(Y_i = k) - Pr(X_i = k) = \sum_{z: z_i = k} \delta(z). \quad (9)$$

Finally, let $\gamma(0, 0, \dots, 0)$ 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 (9) that for all i and k , including $k = 0$,

$$\sum_{z: z_i = k} \gamma(z) = \sum_{z: z_i = k} \delta(z). \quad (10)$$

Equation (10) ensures that $f + \gamma$ has the same marginal distributions as g , so $f + \gamma$ and g can be compared according to \prec_{SPM} .² At the same time, γ contains all the information necessary to determine whether the marginals of g first-order stochastically dominate the marginals of f .

²Strictly 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 $[0, 1]$.

THEOREM 2 (INCREASING SUPERMODULAR ORDERING) *The following statements are equivalent:*

- 1) $g \succ_{\mathcal{I}SPM} f$.
- 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 g dominates the i^{th} marginal distribution of f according to first-order stochastic dominance, and for all supermodular w , $w \cdot g \geq w \cdot (f + \gamma)$.*

3.2 Coarsening and Relabeling

For many applications, the choice of a particular support seems somewhat arbitrary. For example, when comparing multivariate empirical distributions of attributes such as income, health, and education (see Section 7.1), the distributions depend on the way the data for each attribute has been aggregated into discrete categories. We now use Theorem 1 to show that the supermodular ordering is robust to coarsening of the support (aggregation), as well as to monotonic relabeling of coordinates. This is important, since some widely used orderings of interdependence, such as the (bivariate) linear correlation coefficient, fail to satisfy this robustness criterion.

A *coarsening* $\tilde{\mathcal{L}}$ of \mathcal{L} is defined by a partitioning $\tilde{\mathcal{L}}_i$ of \mathcal{L}_i for each i .³ To any coarsening $\tilde{\mathcal{L}}$ of \mathcal{L} corresponds a surjective map $\phi : \mathcal{L} \rightarrow \tilde{\mathcal{L}}$ such that $\phi(x) = \phi(z)$ if and only if x_i and z_i belong to the same element \tilde{x}_i of $\tilde{\mathcal{L}}_i$ for all i . Each element of $\tilde{\mathcal{L}}$ represents a hyperrectangle resulting from slicing \mathcal{L} along each dimension. For any probability distribution f on \mathcal{L} and any coarsening $\tilde{\mathcal{L}}$ of \mathcal{L} , let \tilde{f} denote the “coarsened version” of f , defined by

$$\tilde{f}(\tilde{x}) = \sum_{x \in \mathcal{L}: \phi(x) = \tilde{x}} f(x).$$

THEOREM 3 (COARSENING INVARIANCE) *Suppose that $f \prec_{SPM} g$ and that $\tilde{\mathcal{L}}$ is a coarsening of \mathcal{L} . Then $\tilde{f} \prec_{SPM} \tilde{g}$.*

We can apply Theorem 3 as follows. Suppose that the functions $\phi_i : \mathcal{L}_i \mapsto \mathbb{R}$ are nondecreasing, and let $\phi = (\phi_1, \dots, \phi_n)$. Define $\tilde{\mathcal{L}}_i = \{\phi_i(x_i) : x_i \in \mathcal{L}_i\}$ and $\tilde{\mathcal{L}} = \{\phi(x) : x \in \mathcal{L}\}$. Then it is easy to show that $\tilde{\mathcal{L}} = \times_i \tilde{\mathcal{L}}_i$, so $\tilde{\mathcal{L}}$ is therefore also a lattice.⁴

³For example, if $\mathcal{L} = \{0, 1, 2, 3\} \times \{0, 1, 2\}$, one possible coarsening of \mathcal{L} is $\tilde{\mathcal{L}} = \{\{0, 1\}, \{2, 3\}\} \times \{\{0\}, \{1, 2\}\}$.

⁴The inclusion $\tilde{\mathcal{L}} \subset \times_i \tilde{\mathcal{L}}_i$ is straightforward. To show that the reverse inclusion holds, take any $\tilde{x} \in \times_i \tilde{\mathcal{L}}_i$. For each i , there exists x_i such that $\tilde{x}_i = \phi_i(x_i)$. Letting $x = (x_1, \dots, x_n)$, we have $\tilde{x} = (\phi_1(x_1), \dots, \phi_n(x_n))$, which shows that $\tilde{x} \in \tilde{\mathcal{L}}$.

PROPOSITION 1 Let $\phi = (\phi_1, \dots, \phi_n)$, with each $\phi_i : \mathcal{L}_i \mapsto \tilde{\mathcal{L}}_i$ nondecreasing. If $X \prec_{SPM} Y$, then $\phi(X) \prec_{SPM} \phi(Y)$. Moreover, if each ϕ_i is strictly increasing, then $X \prec_{SPM} Y$ if and only if $\phi(X) \prec_{SPM} \phi(Y)$.

Proof. The first part is an implication of Theorem 3, since each ϕ_i defines a coarsening of \mathcal{L}_i : x_i and z_i belong to the same element \tilde{x}_i of the partition of \mathcal{L}_i if and only if $\phi_i(x_i) = \phi_i(z_i)$. For the second part, suppose that $\phi(X) \prec_{SPM} \phi(Y)$ and that, for each i , ϕ_i is strictly increasing. Then each ϕ_i is invertible, and $X = \phi^{-1}(\phi(X))$, where $\phi^{-1}(\tilde{X}) = (\phi_1^{-1}(\tilde{X}_1), \dots, \phi_n^{-1}(\tilde{X}_n))$ and ϕ_i^{-1} is the inverse of ϕ_i . Applying the first part of the proposition to the function ϕ^{-1} and the relation $\phi(X) \prec_{SPM} \phi(Y)$ then shows that $X \prec_{SPM} Y$. ■

3.3 The Supermodular Ordering and Copulas

A useful approach to modeling the interdependence of random variables, which is widespread in finance and in actuarial science, is based on the concept of a copula.⁵ Given any distribution function F of n variables, with marginal distributions F_1, \dots, F_n , Sklar's theorem (1959) guarantees the existence of a function $C : [0, 1]^n \rightarrow [0, 1]$ such that

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)). \quad (11)$$

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

As noted above, Theorem 1 implies that, for two multivariate distributions F and G to be comparable according to the supermodular ordering, they must have identical marginals: $G_i = F_i$ for all i . This implies that the domain $\tilde{\mathcal{L}}$ is the same for both copulas C_F and C_G . As observed before Proposition 1, $\tilde{\mathcal{L}} = \times_i \tilde{\mathcal{L}}_i$, 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 marginal distributions F_i 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 $\tilde{Y}_i \equiv G_i(Y_i) (= F_i(Y_i))$. Proposition 1 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 (11) that the joint distributions

⁵Copulas have been systematically used, since Li's (2000) influential model, to price credit derivatives. They are used to analyze risk insurance (Denuit et al. (2005)) and, more generally, to think about risk management (Embrechts (2009)). Copulas are also used in statistics and econometrics to model the intertemporal dependence of time series (see Joe (1997, Ch. 8), Ibragimov (2005), and Beare (2010)).

of \tilde{X} and \tilde{Y} on $\tilde{\mathcal{L}}$ coincide, respectively, with the copulas C_F and C_G . We have thus proved the following result.

PROPOSITION 2 $F \prec_{SPM} G$ 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.⁶ In contrast, our methods for characterizing the supermodular ordering, and generating distributions that are ranked according to it, allow nonparametric comparisons. This feature makes our methods useful for comparing multivariate empirical distributions.

3.4 Nonparametric Characterizations of the Supermodular Ordering

Two aspects of our approach greatly facilitate the use of Theorem 1 to determine, given a pair of distributions f and g , whether or not $f \prec_{SPM} g$. The first is our restriction to a *finite* support \mathcal{L} .⁷ The second is our restriction that elementary transformations, defined in (3), 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 transformations on any given \mathcal{L} . The next result guarantees that when looking for a representation of $g - f$ in the form $\sum_{t \in \mathcal{T}} \alpha_t t$, none of the elementary transformations in \mathcal{T} is redundant.

PROPOSITION 3 All elements of \mathcal{T} are extreme rays of $\mathcal{C}(\mathcal{T})$, the convex cone generated by \mathcal{T} .

For two dimensions, a stronger result is easily shown: It is impossible to write any $t \in \mathcal{T}$ as a sum, with weights of *arbitrary* sign, of other elements of \mathcal{T} .⁸ As a consequence, for any bivariate distributions f and g with identical marginals (which is necessary for $f \prec_{SPM} g$), there is a *unique* representation of $g - f$ in the form $\sum_{t \in \mathcal{T}} \alpha_t t$, when we allow the weights α_t to have *arbitrary* signs. To see this, note that if \mathcal{L} has $m_1 \times m_2$ elements, then $g - f$ is fully described by its values at $(m_1 - 1) \times (m_2 - 1)$ points, and there are exactly $(m_1 - 1) \times (m_2 - 1)$ linearly independent elementary transformations.

This uniqueness of the representation $g - f = \sum_{t \in \mathcal{T}} \alpha_t t$ allows a very simple proof of the known characterization of the supermodular ordering in two dimensions. Define I_v and I^v as the indicator

⁶Wei and Hu (2002) obtain positive results for Archimedean copulas and asymmetric extensions thereof, and Burtschell et al (2008) obtain positive results for Gaussian, Student t , Clayton, and Marshall-Olkin families of copulas.

⁷Theorem 4 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.

⁸For three or more dimensions, this stronger result does not hold. To see this, consider $\mathcal{L} = \{0, 1\}^3$, and observe that $t_{13}^{(0,0,0)} = t_{13}^{(0,1,0)} - t_{23}^{(1,0,0)} + t_{23}^{(0,0,0)}$.

functions of the lower-orthant set $\{z|z \leq v\}$ and the upper-orthant set $\{z|z \geq v\}$, respectively. For two dimensions,

$$f \prec_{SPM} g \iff \forall v \in \mathcal{L}, \quad I_v \cdot f \leq I_v \cdot g \quad \text{and} \quad I^v \cdot f \leq I^v \cdot g, \quad (12)$$

that is, supermodular dominance for bivariate distributions is equivalent to the combination of upper-orthant and lower-orthant dominance.^{9,10} Since I_v and I^v are both supermodular, the implication \Rightarrow in (12) is obvious. To prove the reverse implication, let \mathcal{L}^- denote the $(m_1 - 1) \times (m_2 - 1)$ points $x \in \mathcal{L}$ such that $x + e_1 + e_2 \in \mathcal{L}$, and observe that the right-hand side of (12) is equivalent to lower-orthant dominance of g over f for all $v \in \mathcal{L}^-$, coupled with identical marginals for g and f . Indexing the $(m_1 - 1) \times (m_2 - 1)$ transformations in \mathcal{T} by the points in \mathcal{L}^- , we can write the unique representation of $g - f$ as $\sum_{x \in \mathcal{L}^-} \alpha_x t^x$. Hence for each $v \in \mathcal{L}^-$,

$$I_v \cdot (g - f) = I_v \cdot \left(\sum_{x \in \mathcal{L}^-} \alpha_x t^x \right) = \sum_{x \in \mathcal{L}^-} \alpha_x (I_v \cdot t^x) = \alpha_v. \quad (13)$$

The third equality in (13) follows since $I_v \cdot t^v = 1$, whereas for all $x \in \mathcal{L}^-$ such that $x \neq v$, $I_v \cdot t^x = 0$. It follows from (13) that lower-orthant dominance of g over f , coupled with identical marginals, implies that $g - f = \sum_{x \in \mathcal{L}^-} \alpha_x t^x$, with $\alpha_x \geq 0$ for all $x \in \mathcal{L}^-$, and hence that $f \prec_{SPM} g$.

Note that (13) also identifies the weights α_x in the unique decomposition of $g - f$ for bivariate distributions with identical marginals: $\alpha_x = I_x \cdot (g - f) = G(x) - F(x)$. In two dimensions, the indicator functions of lower orthant sets are in fact the extreme rays of the cone \mathcal{S} of supermodular functions, and there is a one-to-one mapping associating with each $t^x \in \mathcal{T}$ the only extreme ray I_v of \mathcal{S} , namely I_x , such that $I_v \cdot t^x = 0$.

For more than two dimensions, however, many decompositions of $g - f$ into weighted sums of elementary transformations exist, and as a consequence such a one-to-one mapping between elementary transformations and extreme supermodular functions does not exist. In addition, for more than two dimensions, the supermodular ordering is in general strictly stronger than the combination of upper-orthant and lower-orthant dominance (Joe, 1990, and Müller and Scarsini, 2000).¹¹

⁹We can use Theorem 2 to provide an analogous proof that for two dimensions, $f \prec_{ISPM} g$ if and only if g dominates f according to upper-orthant dominance.

¹⁰See Levy and Paroush (1974), Epstein and Tanny (1980), and Tchen (1980). The latter two papers proved the implication \Leftarrow in (12) constructively, by defining a notion of a simple ‘‘correlation increasing’’ transformation. (Levy and Paroush’s proof assumed continuous distributions and used integration by parts.) These proofs were laborious, because a) they did not restrict their transformations to affect values at only *adjacent* points in the support and b) they sought a weighted sequence of transformations that, when added to f , yielded g and that produced, after each step, a probability distribution. Our Theorem 1 makes clear that, in searching for a decomposition of $g - f$ into a weighted sum $\sum_{t \in \mathcal{T}} \alpha_t t$, it is irrelevant whether or not partial sums of the form $f + \sum_{t \in \mathcal{T}' \subset \mathcal{T}} \alpha_t t$ are actual probability distributions.

¹¹In Meyer and Strulovici (2012), we contrast five orderings of interdependence, including the supermodular ordering and the combination of upper-orthant and lower-orthant dominance. We show that, for two dimensions, all five orderings are equivalent, but that, for an arbitrary number of dimensions, the five orderings are strictly ranked.

¹² These features make it considerably more difficult to determine, given distributions f and g , whether or not $f \prec_{SPM} g$ when (X_1, \dots, X_n) and (Y_1, \dots, Y_n) have three or more dimensions.

For three or more dimensions, we provide two methods for determining whether a pair of distributions can be ranked according to \prec_{SPM} . Both methods exploit Theorem 1’s dual characterization of the ordering as well as Proposition 3’s result that all elementary transformations as defined in (3) are extreme. We now briefly summarize these methods; details are provided in Section D of the Appendix.

Theorem 1 shows that $f \prec_{SPM} g$ if and only if there exists a representation of $g - f$ in the form $\sum_{t \in \mathcal{T}} \alpha_t t$ with all $\alpha_t \geq 0$. This existence problem can be reformulated as establishing the non-emptiness of the domain of a linear program. This, in turn, leads to the formulation of an auxiliary linear program, based on the set of elementary transformations of \mathcal{L} , such that the optimum value of the program is equal to zero if and only if there exist non-negative coefficients $\{\alpha_t\}_{t \in \mathcal{T}}$ such that $g - f = \sum_{t \in \mathcal{T}} \alpha_t t$. This method has the advantage, in the case where f and g are rankable according to the supermodular ordering, of constructing an explicit sequence of transformations that, added to f , result in g . However, it also has the drawback that one has to solve a different linear program for each pair of distributions to be compared.

A second method, based on Minkowski’s and Weyl’s representation theorems for polyhedral cones, allows one to compute, for any given support \mathcal{L} , a minimal set of inequalities which completely characterize the supermodular ordering. That is, $f \prec_{SPM} g$ if and only if the vector $g - f$ satisfies all of these inequalities. This method can be used to compare many empirical distributions, or to compare interdependence of many mechanism designs. Specifically, we develop an algorithm, based on the “double description method” conceptualized by Motzkin et al. (1953) and developed by Avis and Fukuda (1992), to generate, for any given support, the set of extreme rays of the cone of supermodular functions. Each extreme ray defines an inequality of the minimal set characterizing \prec_{SPM} . Using a package to implement the double description method freely available for Matlab (see Torrisi and Baotic (2005)), we have written and used programs to characterize both the supermodular ordering and the symmetric supermodular ordering studied in Section 5.¹³

3.5 Continuous Support

The analysis so far has focused on discrete supports. We now prove that 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} .

¹²Hu, Xie, and Ruan (2005) have shown that in the special case of three-dimensional Bernoulli random vectors, the equivalence in (12) remains valid.

¹³The code is available online at <http://faculty.wcas.northwestern.edu/bhs675/>.

Recall that 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 coarsening of F on $\tilde{\mathcal{L}}$ is the distribution \tilde{F} such that for all $\tilde{x} \in \tilde{\mathcal{L}}$, $\tilde{F}(\tilde{x})$ is the probability that F puts 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,

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

In light of Theorem 3, 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 4 *Suppose that distributions F and G have continuous densities. $F \prec_{\mathcal{L}SPM} G$ if and only if $\tilde{F} \prec_{SPM} \tilde{G}$ on all finite coarsenings $\tilde{\mathcal{L}}$ of \mathcal{L} .*

We note here that for random vectors X and Y with multivariate normal distributions, necessary and sufficient conditions for $X \prec_{SPM} Y$ and for $X \prec_{ISPM} Y$ are easily stated. Müller and Scarsini (2000) have shown that $X \prec_{SPM} Y$ if and only if X and Y have the same marginal distributions and $Cov(X_i, X_j) \leq Cov(Y_i, Y_j)$ for all $i \neq j$. Arlotto and Scarsini (2009) have shown that $X \prec_{ISPM} Y$ if and only if $EX_i \leq EY_i$ and $Var(X_i) = Var(Y_i)$ for all i and $Cov(X_i, X_j) \leq Cov(Y_i, Y_j)$ for all $i \neq j$. To highlight the relationship between our characterization of the increasing supermodular ordering in Theorem 2 and the latter result, we can rewrite the latter result as $X \prec_{ISPM} Y$ if and only if $EX_i \leq EY_i$ for all i and $X' \prec_{SPM} Y$, where $X'_i \equiv X_i + (EY_i - EX_i)$.

4 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 the class of 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.

Two familiar parametric examples will help to motivate our approach.

Example 1 Let the random vector X be such that $X_i = \theta + \varepsilon_i$, where θ and $\{\varepsilon_i\}_{i \in \{1, \dots, n\}}$ are all independent and normally distributed with mean 0 and where $Var(\theta) = \tau$ and $Var(\varepsilon_i) = (1 - \tau)$.¹⁴ Intuitively, an increase in τ raises the contribution to each X_i of the common shock θ relative to that of the idiosyncratic shock ε_i , while leaving the marginal distribution of each X_i unchanged.

¹⁴This additive-normal structure has been used, for example, to examine how interdependence in agents' information affects behavior in "beauty-contest" coordination games (Myatt and Wallace, 2012) and in voting games (Myatt, 2007).

More formally, an increase in τ raises $Cov(X_i, X_j)$ for each $i \neq j$. It therefore follows from Müller and Scarsini’s (2000) result quoted above for normal distributions that an increase in τ yields a distribution that dominates the original one according to the supermodular ordering.

Example 2 The conclusion in Example 1 need not hold if we relax the assumption of normal distributions. Let $X_i = \theta + \varepsilon_i$, where now θ equals 1 or -1 with probability p and $1 - p$, respectively, and ε_i equals 2 or -2 with probability $1 - p$ and p , respectively. Similarly, let $Y_i = \theta' + \varepsilon'_i$, where θ' equals 2 or -2 with probability $1 - p$ and p , respectively, and ε'_i equals 1 or -1 with probability p and $1 - p$, respectively. Y and X have identical marginals, and the common shock would seem to be more important relative to the idiosyncratic shock in the distribution of Y than in X . Nevertheless, for any $p \neq \frac{1}{2}$, the distributions of Y and X cannot be ranked according to the supermodular ordering.¹⁵

In this section, we develop a general, non-parametric criterion for comparing two joint distributions according to the relative importance of aggregate vs. idiosyncratic shocks, and we prove that if two distributions can be ranked according to this criterion, then the one for which the aggregate shock is relatively more important dominates the other according to the supermodular ordering.

We will consider the following class of “mixture distributions” (mixtures of conditionally independent random variables).¹⁶ To each variable X_r , $r \in \{1, \dots, n\}$, is associated a $q \times m$ 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 values. The vector (X_1, \dots, X_n) is constructed as follows. First, a row index $i \in \{1, \dots, q\}$ is drawn randomly, according to a uniform distribution.¹⁷ This step represents the realization of the aggregate shock. 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 $A(r)$. Because, as we observed earlier (Proposition 1), the supermodular ordering is invariant with respect to monotonic coordinate changes, we take, without loss of generality, the support of each variable to be $\{1, \dots, m\}$.

For mixture distributions of the form just described, greater importance of the aggregate shock relative to the idiosyncratic shocks should correspond to the rows of the matrices $A(r)$ being more different from one another, holding the unconditional distribution of each X_r unchanged by holding the average of the rows of each $A(r)$ fixed.

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 matrix A , the (upper) *cumulative-sum matrix* \bar{A} of A is defined by $\bar{A}_{i,j} = \sum_{k=j}^m A_{i,k}$. Thus, $\bar{A}_{i,j}$ is decreasing in j . If A

¹⁵To see this, note that all upper-orthant and lower-orthant indicator functions are supermodular, and observe that for $p > (<) \frac{1}{2}$, $P(Y_1 \geq 3, Y_2 \geq 3) > (<) P(X_1 \geq 3, X_2 \geq 3)$ and $P(Y_1 \leq -3, Y_2 \leq -3) < (>) P(X_1 \leq -3, X_2 \leq -3)$.

¹⁶In the statistics literature, distributions generated as described below are often referred to as unidimensional latent variable models (Holland and Rosenbaum, 1986).

¹⁷The analysis can easily be extended to accommodate non-uniform distributions of the index i , by appropriate replication of the rows of the matrix.

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 $i \in \{2, \dots, q\}$, the i^{th} row of A dominates the $(i-1)^{\text{th}}$ 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.

Before introducing our ordering of matrices, we recall 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 vectors have identical sums, and ii) for all k , the sum of the k largest entries of a is weakly greater than the sum of the k largest entries of b .*

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 $A \succ_{CCM} B$ (or equivalently $\bar{A} \succ_{CCM} \bar{B}$), 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 all $l \leq q$ and $k \leq m$, $\sum_{i=l}^q \bar{A}_{i,k}^{so} \geq \sum_{i=l}^q \bar{B}_{i,k}^{so}$, 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 $\{1, \dots, m\}$, 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 . In turn, the fact that the q possible conditional distributions are everywhere more diverse for matrix A than for matrix B , while the unconditional distribution is the same, implies that the aggregate shock is more important in the mixture distribution generated by A than in the mixture distribution generated by B .

The main result of this section is the following theorem.

THEOREM 5 *Let $(A(1), \dots, A(n))$ and $(B(1), \dots, B(n))$ be two sets of row-stochastic matrices generating the random vectors (X_1, \dots, X_n) and (Y_1, \dots, Y_n) , respectively. Suppose that, i) for each $r \in \{1, \dots, n\}$, $A(r)$ is stochastically ordered, and ii) for each $r \in \{1, \dots, n\}$, $A(r) \succ_{CCM} B(r)$. Then $(X_1, \dots, X_n) \succ_{SPM} (Y_1, \dots, Y_n)$.*

We have examples showing that Theorem 5 does not hold if we drop either condition i) or condition ii).¹⁸

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 $\{X_r\}_{r=1}^n$ will be than about what the realizations of $\{Y_r\}_{r=1}^n$ will be. In the special case where both $A(r)$ and $B(r)$ are stochastically ordered, $A(r) \succ_{CCM} B(r)$ reduces to

$$\sum_{i=l}^q \sum_{j=k}^m A_{i,j}(r) = \sum_{i=l}^q \bar{A}_{i,k}(r) \geq \sum_{i=l}^q \bar{B}_{i,k}(r) = \sum_{i=l}^q \sum_{j=k}^m B_{i,j}(r) \quad \forall l \geq 2, k \geq 2, \quad (15)$$

coupled with the condition that $A(r)$ and $B(r)$ have matching column sums. Inequality (15) can be read as saying that the matrix $A(r)$ dominates $B(r)$ in the sense of “upper-orthant dominance”.¹⁹

To provide some insight into the proof of Theorem 5, we focus on the special case where the random vectors X and Y have *symmetric* mixture distributions: this is the case where the sets of matrices $\{A(r)\}_r$ and $\{B(r)\}_r$ are each independent of 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 ,

$$Ew(X_1, \dots, X_n) = \frac{1}{q} \sum_{i=1}^q E[w(X_1, \dots, X_n) | \bar{A}_{i,\bullet}] \geq \frac{1}{q} \sum_{i=1}^q E[w(Y_1, \dots, Y_n) | \bar{B}_{i,\bullet}] = Ew(Y_1, \dots, Y_n), \quad (16)$$

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

Let $\bar{p} \equiv (\bar{p}_1, \dots, \bar{p}_m)$ denote an arbitrary upper-cumulative vector corresponding to a discrete distribution on support $\{1, \dots, m\}$. We have $\bar{p}_1 = 1$ and $\bar{p}_{k-1} \geq \bar{p}_k$ for all k . For any supermodular

¹⁸Jogdeo (1978) showed that for any stochastically ordered row-stochastic matrices $\{A(r)\}$, the distribution of (X_1, \dots, 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, \dots, 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 5 where for each r , the matrix $B(r)$ consists of q identical rows.

¹⁹Athey 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.

objective function w on \mathbb{R}^n , define $\bar{w}(\bar{p})$ by

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

Using this definition, (16) can be rewritten as

$$Ew(X_1, \dots, X_n) = \frac{1}{q} \sum_{i=1}^q \bar{w}(\bar{A}_{i,\bullet}) \geq \frac{1}{q} \sum_{i=1}^q \bar{w}(\bar{B}_{i,\bullet}) = Ew(Y_1, \dots, Y_n). \quad (17)$$

The function \bar{w} is defined on a convex lattice of \mathbb{R}^m and, importantly, inherits several properties from the supermodularity of w , as shown in the following lemma.²⁰ A function $h(x_1, \dots, x_r, \dots, x_m)$ is *componentwise convex* if, when considered as a function of just x_r , it is convex, for each $r \in \{1, \dots, m\}$, for all values of the other $m - 1$ arguments.²¹

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

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 (17) holds. With $q = 2$, condition *i*) in Lemma 2 implies that A is stochastically ordered, and conditions *ii*) and *iii*) are equivalent to $A \succ_{CCM} B$. Recall that for all row-stochastic matrices, the first column of the corresponding cumulative-sum matrix has all entries equal to 1.

LEMMA 2 *Suppose that $q = 2$ and that there exists a nonnegative vector ε such that for all $k \in \{2, \dots, m\}$, *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 $(X_1, \dots, X_n) \succ_{SPM} (Y_1, \dots, Y_n)$.*

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

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

Equivalently, we need to show that

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

²⁰The domain of \bar{w} is a simplex and is clearly convex. Moreover, the inequalities $\bar{p}_1 \geq \bar{p}_2 \geq \dots \geq \bar{p}_m$ 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)).

²¹Functions 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”.

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

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

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

Starting from the stochastically ordered matrix \bar{A} , the matrix \bar{B} described in Lemma 2 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 5 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 2, affecting only two of the q rows. From (16), 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 (17). 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 .

Example 3 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 2×3 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 5 implies that for any n , (X_1, \dots, X_n) dominates (Y_1, \dots, Y_n) , (Z_1, \dots, Z_n) , and (V_1, \dots, V_n) according to \succ_{SPM} .

For symmetric mixture distributions generated from matrices with only two rows, and for any n , we can show that the pair of conditions in Theorem 5 are necessary as well as sufficient for the random vectors to be supermodularly ordered. To illustrate this result, observe first that matrices B and C above cannot be ranked according to \succ_{CCM} . It follows from the necessity of the \succ_{CCM} condition that Y and Z cannot be ranked according to \succ_{SPM} , whatever the value of $n \geq 2$. In fact, because the third column of \bar{C} majorizes (strictly) the third column of \bar{B} , we can deduce

that for $w(x) = I_{\{x_1 \geq 3, x_2 \geq 3\}}$, we have $Ew(Z) > Ew(Y)$, and because the second column of \bar{B} majorizes (strictly) the second column of \bar{C} , we can deduce that for $w(x) = I_{\{x_1 \geq 2, x_2 \geq 2\}}$, we have $Ew(Y) > Ew(Z)$. Second, observe that 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 $w(x) = I_{\{x_1 \geq 3, x_2 \geq 2\}}$.

Example 2 (revisited) Let $X_i = \theta + \varepsilon_i$, where θ equals 1 or -1 with probability p and $1-p$, respectively, and ε_i equals 2 or -2 with probability $1-p$ and p , respectively. Similarly, let $Y_i = \theta' + \varepsilon'_i$, where θ' equals 2 or -2 with probability $1-p$ and p , respectively, and ε'_i equals 1 or -1 with probability p and $1-p$, respectively. Set $p = \frac{2}{3}$. The random vectors X and Y have symmetric mixture distributions on $\mathcal{L} = \{-3, -1, 1, 3\}^n$, generated by the 3×4 matrices P and Q , respectively:

$$P = \begin{pmatrix} \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \end{pmatrix} \quad Q = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

Both P and Q are stochastically ordered. The duplication of the bottom two rows in P and the top two rows in Q reflects the fact that the aggregate shock θ is twice as likely to be high in P as low and twice as likely to be low in Q as high. The rows of P and Q have a common arithmetic average, confirming that X and Y have identical marginal distributions. However, neither $P \succ_{CCM} Q$ nor $Q \succ_{CCM} P$ holds: the third column of \bar{P} is majorized by the third column of \bar{Q} , while the reverse is true for the second columns of \bar{P} and \bar{Q} . As we noted above, the random vectors X and Y cannot be ranked according to \succ_{SPM} , because for $w(x) = I_{\{x_1 \geq 3, x_2 \geq 3\}}$, $Ew(X) < Ew(Y)$, while for $w(x) = I_{\{x_1 \leq -3, x_2 \leq -3\}}$, $Ew(X) > Ew(Y)$.

5 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 a form of ex post anonymity across individuals: permutations of the vector of realized incomes leave welfare unchanged. In finance and insurance contexts, losses may be evaluated according to a convex function of the total loss across all assets or all insurance policies. Any convex function of the sum of outcomes across random variables is both symmetric and supermodular.

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 for all permutations $\sigma(x)$ of x .

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

$$f^{symm}(x) = \frac{1}{n!} \sum_{\sigma \in \Sigma(n)} f(\sigma(x)), \quad (18)$$

for any x , where $\Sigma(n)$ is the set of all permutations of $\{1, \dots, n\}$. If w is a supermodular function, then w^{symm} is supermodular. The following equivalence result was proved in Meyer and Strulovici (2012, Section 2.3.1):

PROPOSITION 4 *Given distributions f, g defined on a symmetric lattice, $f \prec_{SSPM} g$ if and only if $f^{symm} \prec_{SPM} g^{symm}$.*

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

Meyer and Strulovici (2012) showed that the symmetric supermodular ordering has a very simple form for random vectors for which each component has a binary support $\{0, 1\}$, so the lattice is $\mathcal{L} = \{0, 1\}^n$. For such a random vector $X = (X_1, \dots, X_n)$, define $c(X) \equiv \sum_{i=1}^n I_{\{X_i=1\}}$. The ‘‘count function’’ $c(X)$ gives the number of components of X for which the realization takes the value 1. For random variables Z and V with support $S \subseteq \mathbb{R}$, we say Z dominates V according to the (univariate) convex ordering, written $V \prec_{\mathcal{X}} Z$, if $Ew(V) \leq Ew(Z)$ for all convex functions $w : S \rightarrow \mathbb{R}$.²² The convex ordering is equivalent to the ordering of greater riskiness studied by Rothschild and Stiglitz (1970).

PROPOSITION 5 *For random vectors Y and X distributed on $\mathcal{L} = \{0, 1\}^n$, $X \prec_{SSPM} Y$ if and only if $c(X) \prec_{\mathcal{X}} c(Y)$.*

This proposition is easily proved, by noting, first, that any symmetric function w defined on $\mathcal{L} = \{0, 1\}^n$ can be written as $w(X_1, \dots, X_n) = \phi(c(X_1, \dots, X_n))$, for some $\phi : \{0, 1, \dots, n\} \rightarrow \mathbb{R}$, and second, that a function w of this form is supermodular if and only if $\phi(\cdot)$ is convex.

The next section applies these results on the symmetric supermodular ordering to develop a generalization of well-known results in the statistics literature concerning the variability of distributions of the number of successes in independent trials, when success probabilities differ across trials.²³

6 Comparing Distributions Generated from Heterogeneous Lotteries

Let $(X_1, \dots, X_n) \in \{0, 1\}^n$ (resp., $(Y_1, \dots, Y_n) \in \{0, 1\}^n$) 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 $\sum_{i=1}^n a_i =$

²²A function defined on $S \subset \mathbb{R}$ is convex if it can be extended to a convex function on \mathbb{R} .

²³This topic has been explored by Hoeffding (1956), Karlin and Novikoff (1963), and Gleser (1975), among others.

$\sum_{i=1}^n b_i$, 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 $c(Y)$?²⁴ Karlin and Novikoff (1963) showed that if (a_1, \dots, a_n) majorizes (b_1, \dots, b_n) , then $c(X) \prec_X c(Y)$.

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 4 and 5, combined with Karlin and Novikoff's result, imply that if (a_1, \dots, a_n) majorizes (b_1, \dots, b_n) , then i) $(X_1, \dots, X_n) \prec_{SSPM} (Y_1, \dots, Y_n)$ 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 $X' = (X'_1, \dots, X'_n)$ denote the random vector whose distribution matches 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 (as defined in (3)) 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.

What are the lessons of this example for n independent Bernoulli trials, when the expected number of successes is held fixed but the vector of individual success probabilities is varied? The example shows 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.

In this section, we focus on multivariate distributions representing the outcome of n independent lotteries, each with an *arbitrary* finite support. Our objective is to explore in greater generality 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

²⁴As in Section 5, $c(\cdot)$ denotes the count function.

sets of n independent lotteries, we present in Theorem 6 sufficient conditions for their outcome distributions to be rankable according to the symmetric supermodular ordering, or equivalently, for the degree of negative interdependence of the symmetrized versions of their outcome distributions to be rankable according to the supermodular ordering. Theorem 6 can be used to compare different production designs in the presence of complementarity among tasks, and it can also be used to compare ex post inequality of reward schemes under uncertainty (see Section 7.1).

We will compare random vectors (X_1, \dots, X_n) and (Y_1, \dots, Y_n) that are 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 $\{1, \dots, m\}$, and the $\{X_i\}$ (resp. $\{Y_i\}$) are independent.²⁵ Just as above we compared sets of n independent Bernoulli trials with the same average success probability, here we want to compare sets of n independent lotteries with the same average distribution over the m prizes. This constraint translates into the requirement on the matrices A and B that for each j , the j^{th} column of A has the same sum as the j^{th} column of B .

As above, denote by (X'_1, \dots, X'_n) and (Y'_1, \dots, Y'_n) the random vectors whose distributions match the symmetrized distributions of (X_1, \dots, X_n) and (Y_1, \dots, Y_n) , respectively. The common marginal distribution of the $\{X'_i\}$ is the average of the rows of matrix A . Hence, our requirement that the matrices being compared have matching column sums implies that the marginal distribution of the $\{X'_i\}$ is identical to the marginal distribution of the $\{Y'_i\}$.

In the Bernoulli example above, dispersion of the n -vector of success probabilities was captured by majorization. For lotteries with m -point supports represented by the n rows of a matrix, we want a generalization of majorization to formalize the idea of the rows of a matrix being more different from one another, holding the average of the rows fixed. Our cumulative column majorization ordering defined in Section 4 again turns out to be key.

THEOREM 6 *Let A and B be $n \times m$ row-stochastic matrices generating the independent random vectors (X_1, \dots, X_n) and (Y_1, \dots, Y_n) , respectively. Let (X'_1, \dots, X'_n) and (Y'_1, \dots, Y'_n) have distributions matching the symmetrized distributions of (X_1, \dots, X_n) and (Y_1, \dots, Y_n) , respectively. Suppose that i) A is stochastically ordered, and ii) $A \succ_{CCM} B$. Then $(X_1, \dots, X_n) \prec_{SSPM} (Y_1, \dots, Y_n)$ and $(X'_1, \dots, X'_n) \prec_{SPM} (Y'_1, \dots, Y'_n)$.*

As with Theorem 5, we have examples showing that Theorem 6 does not hold if we drop either condition i) or condition ii).²⁶

²⁵The symmetric supermodular ordering is invariant to monotonic coordinate changes that preserve the symmetry of the lattice, so it is without loss of generality to take the support of each marginal distribution to be $\{1, \dots, m\}$.

²⁶Hu 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 6 where the rows of the matrix B are all identical.

The proof of Theorem 6 closely parallels that of Theorem 5. As in Section 4, it is convenient to work with the cumulative-sum matrices \bar{A} and \bar{B} corresponding to A and B , respectively. The following lemma plays a role analogous to that of Lemma 2 in the proof of Theorem 5:

LEMMA 3 *Suppose that $n = 2$ and that there exists a nonnegative vector ε such that for all $k \in \{2, \dots, m\}$, 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 4 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 . As discussed in Section 3.4, for bivariate distributions with identical marginals, supermodular dominance is equivalent to upper-orthant dominance. For any $k, l \in \{2, \dots, m\}$,

$$\begin{aligned} 2[P(Y'_1 \geq k, Y'_2 \geq l) - P(X'_1 \geq k, X'_2 \geq l)] &= P(Y_1 \geq k, Y_2 \geq l) + P(Y_1 \geq l, Y_2 \geq k) - P(X_1 \geq k, X_2 \geq l) - P(X_1 \geq l, X_2 \geq k) \\ &= \bar{B}_{1k}\bar{B}_{2l} + \bar{B}_{2k}\bar{B}_{1l} - \bar{A}_{1k}\bar{A}_{2l} - \bar{A}_{2k}\bar{A}_{1l}. \end{aligned}$$

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

$$2[P(Y'_1 \geq k, Y'_2 \geq l) - P(X'_1 \geq k, X'_2 \geq l)] = \varepsilon_k[\bar{A}_{2l} - (\bar{A}_{1l} + \varepsilon_l)] + \varepsilon_l[\bar{A}_{2k} - (\bar{A}_{1k} + \varepsilon_k)]. \quad (19)$$

Condition i) ensures that both of the terms in square brackets in (19) 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 3 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 6 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 3, affecting only two of the n rows. The following lemma, combined with Lemma 3, then ensures that each of these transformations raises the expected value of any symmetric and supermodular objective function.²⁷

²⁷Although Lemma 4 is stated for 2-dimensional random vectors X and Y , it would hold for any (common) dimensionality of X and Y greater than or equal to 2.

LEMMA 4 *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 $E[w(X, Z)] = E[E[w(X, Z)|Z]] = E[r(Z)] \leq E[s(Z)] = E[E[w(Y, Z)|Z]] = E[w(Y, Z)]$. ■

As an application of Theorem 6, 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. Suppose that the production function is symmetric and supermodular in the performance levels on the different tasks, reflecting interchangeability and complementarity among tasks. Suppose that an organization designer can choose how to allocate resources across the different tasks, thereby shifting the distributions of performance, subject to a constraint on the average distribution over all tasks. Theorem 6 identifies conditions under which expected production is higher in one setting than the other for all symmetric supermodular production functions.

Bond and Gomes's (2009) multi-task principal-agent model embeds the special case of this problem where $m = 2$. An agent chooses a level $e_i \in [\underline{e}, \bar{e}]$ of effort for each task, incurring a total effort cost $\sum_{i=1}^n e_i$. Performance on each task is binary, with e_i the probability of success. The interesting case is when the principal's benefit is a convex function of the number of successes. 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 \bar{e} on the remainder. In this case, given the total effort exerted, the agent's effort allocation actually minimizes expected social surplus.

Theorem 6 implies these conclusions about the effort allocations that maximize and minimize expected social surplus. With binary task outcomes, a convex function of the sum of successes is a symmetric supermodular function of the vector of outcomes (see Proposition 5 and the argument following it). 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 6 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.²⁸ 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 \bar{e}) majorizes all vectors with the same total.

Using Theorem 6, we can examine, for the case of 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, for the class of distributions considered here, we have the following positive results.

PROPOSITION 6 *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 , $(X_1, \dots, X_n) \prec_{SSPM} (Y_1, \dots, Y_n)$;*

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 , $(X_1, \dots, X_n) \prec_{SSPM} (Y_1, \dots, Y_n)$.

The “optimal” matrix B identified by part ii) of Proposition 6 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.²⁹

²⁸Bond and Gomes’s conclusions also follow from Karlin and Novikoff’s (1963) result for Bernoulli trials, discussed above.

²⁹Note 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 2×3 matrix both of whose rows match p , and let A be the 2×3 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 $w(z_1, z_2) = I_{\{z_1 \geq 3, z_2 \geq 3\}}$, $Ew(X_1, X_2) = 0 < \frac{1}{16} = Ew(Y_1, Y_2)$.

7 Applications

7.1 Welfare and Inequality

Comparisons of Inequality in the Presence of Uncertainty

A first application of the supermodular ordering concerns welfare economics. In many group settings where individual outcomes (e.g. rewards) are uncertain, members of the group may be concerned, *ex ante*, about how unequal their *ex post* rewards will be.³⁰ As argued by Meyer and Mookherjee (1987), an aversion to *ex post* inequality can be formalized by adopting an *ex post* welfare function that is symmetric and supermodular in the realized utilities of the individuals. Given two mechanisms for allocating rewards (formally, two joint distributions of utilities), when can we be sure that one mechanism generates higher expected welfare than the other, for all symmetric and supermodular *ex post* welfare functions? Our characterization results allow us to answer this question.

Consider a specific illustration. Intuitively, when groups dislike *ex post* inequality, tournament reward schemes, which distribute a fixed set of rewards among individuals, one to each person, should be particularly unappealing, since they generate a form of negative interdependence among rewards: if one person receives a higher reward, this must be accompanied by another person's receiving a lower reward. This reasoning suggests the conjecture that tournaments should be dominated, in the sense of the 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 for an arbitrary number of individuals (dimensions), but only for the special case of a symmetric tournament (one in which each individual has an equal chance of winning each of the rewards), and their method of proof was laborious. Theorem 6 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. A symmetric tournament is the special case in which each of the $n!$ allocations is equally likely. 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 across individuals the tournament is, the more disparate are the rows of the corresponding matrix B . Given any tournament, consider the associated reward scheme (which is not a tournament, except for extreme cases) which gives each individual the same marginal distribution over rewards as he

³⁰See Meyer and Mookherjee, 1987; Meyer, 1990; Ben-Porath et al, 1997; Gajdos and Maurin, 2004; Kroll and Davidovitz, 2003; Adler and Sanchirico, 2006; Chew and Sagi, 2012. This concern is distinct from concerns about the mean level of rewards and about their riskiness.

receives in the tournament but which determines rewards independently. For any tournament, however asymmetric, we now show 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.³¹

PROPOSITION 7 *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.*

Proof. 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 6 therefore yields the result.

Multidimensional Poverty and Multidimensional Deprivation

A second application in welfare economics concerns comparisons of poverty and deprivation when individual-level data are available on different dimensions of economic status, for example, on attributes such as income, health, and education.³² In such a context, as noted by Atkinson (2003), there are two alternative approaches to comparisons of multidimensional poverty and deprivation

³¹For 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 comparing the “average” degree of negative interdependence across the whole set of individuals. Equivalently, as Proposition 4 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.

³²See Atkinson and Bourguignon, 1982, Bourguignon and Chakravarty, 2002, the Symposium in Honor of Amartya

between two datasets (e.g. two countries, two time periods). One approach is to compare deprivation in the two datasets dimension by dimension. Another is to first aggregate across dimensions to generate a deprivation measure for each individual in each dataset and then to sum these measures to generate an aggregate deprivation measure for the whole dataset. Importantly, under the second approach, deprivation comparisons will be sensitive to the degree of interdependence displayed by the joint distributions of attributes in the two datasets.

Pursuing the second approach, one might classify an individual as multidimensionally poor if and only if his “achievement level” x_i on each dimension i falls below a threshold level z_i . This method of identifying those who are deprived in a multidimensional context has been termed the “intersection approach” (Alkire and Foster, 2011, and Atkinson, 2003). Such an individual deprivation measure would have the form $d(x_1, \dots, x_n) = I_{\{x_i \leq z_i \forall i\}}$. Since this is a lower-orthant indicator function, it is a supermodular function of (x_1, \dots, x_n) . Therefore, if one population joint distribution of achievement levels in the different dimensions is more interdependent than another, in the sense of the supermodular ordering, the aggregate level of deprivation obtained by summing this deprivation measure over individuals will be higher in the former case than in the latter. Alternatively, the “union approach” classifies an individual as poor if and only if there is at least one dimension i in which $x_i \leq z_i$. In this approach, the individual deprivation measure would have the form $d(x_1, \dots, x_n) = 1 - I_{\{x_i \geq z_i \forall i\}}$, which is a submodular function of (x_1, \dots, x_n) , since the supermodular upper-orthant indicator function appears with a negative sign in front of it. With individual deprivation identified in this manner, lower interdependence in the multidimensional distribution of achievement levels, in the sense of the supermodular ordering, will imply lower aggregate deprivation.

In the intersection approach, there is a complementarity among the different dimensions in the determination of individual deprivation. A natural generalization of this approach 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 :

$$d(x_1, \dots, x_n) = \phi\left(\sum_{i=1}^n I_{\{x_i \leq z_i\}}\right), \quad (20)$$

where ϕ 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 (20) where ϕ is increasing and concave. 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 ϕ convex (concave), the deprivation function in (20) is a symmetric supermodular (symmetric submodular) function of (x'_1, \dots, x'_n) . Therefore, for a given vector of thresholds (z_1, \dots, z_n) , aggregate deprivation will be lower in one population than another, for all deprivation measures in the class in (20) with ϕ convex (concave),

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.

if and only if the distribution of (x'_1, \dots, x'_n) in one population is more (less) interdependent, in the sense of the symmetric supermodular ordering, than in the other. Proposition 5 then shows that in this context, symmetric supermodular dominance is equivalent to univariate convex dominance for distributions of $\sum_{i=1}^n x'_i = \sum_{i=1}^n I_{\{x_i \leq z_i\}}$.

7.2 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 (Boland and Proschan, 1988; Baldiga and Green, 2011) or information (Prat, 2002) of members of decision-making groups. In a strategic model of consensus-building within a committee, Caillaud and Tirole (2007) study how the degree of interdependence of members' ex ante uncertain payoffs from a proposal affects the proposer's optimal persuasion strategy. In a strategic model of search and voting, Moldovanu and Shi (2012) examine how the degree of alignment in committee members' preferences affects their equilibrium search strategy and welfare. In both of these latter papers, however, in order to carry out comparative statics analysis with respect to the degree of alignment in individuals' preferences, restrictive assumptions are made. First, both papers analyze only the case where the voting rule requires unanimous approval by committee members. In addition, Caillaud and Tirole (2007) restrict attention to payoff distributions where each member's payoff can assume only two different values, and Moldovanu and Shi (2012) focus on a parametric family of payoff functions in which the degree of alignment of preferences is represented by a single parameter. 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 $X_t = (X_{1t}, \dots, X_{nt})$, 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 $(Y_1, \dots, Y_n) \sim g$, $(X_1, \dots, X_n) \sim f$, and the distribution of (Y_1, \dots, Y_n) dominates the distribution of (X_1, \dots, X_n) according to the supermodular ordering, 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, \dots, z_n) satisfy the n simultaneous equations

$$c_i = E \left[(X_i - z_i) I_{\{X_j \geq z_j \forall j\}} \right], \quad i = 1, \dots, n. \quad (21)$$

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

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, i.e. the expression in square brackets on the right-hand side of (21), is a supermodular function of (X_1, \dots, X_n) , for all (z_1, \dots, z_n) . To confirm this, observe that we can 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, if committee members' interests become more aligned, each member's expected gain from one more search increases, for any vector of reservation levels. Since the right-hand side of the equilibrium condition is also decreasing in z_i , it follows that when alignment of interests increases, member i 's optimal z_i increases, for all z_{-i} . Consequently, we can show that, for any number n of committee members, if the committee is symmetric ($c_i = c$ for all i and the distributions of attributes are symmetric across members), then when interests become more aligned, the common equilibrium reservation value increases (that is, the members become choosier).

To examine how the voting rule affects the comparative-static analysis of changes in the alignment of members' interests, suppose now that a candidate is hired if and only if at least m of the n members vote to stop searching. For a given (z_1, \dots, z_n) , define $K(z_1, \dots, z_n) \equiv \{k | X_k \geq z_k\}$. Then the equilibrium reservation levels satisfy

$$c_i = E [(X_i - z_i) I_{\{|K| \geq m\}}], \quad i = 1, \dots, n. \quad (22)$$

When unanimity is required to reject a candidate ($m = 1$), the expression in square brackets on the right-hand side of (22) 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, \dots, X_n) , for all (z_1, \dots, z_n) .³³ Consequently, for this alternative voting rule, the comparative statics result derived above continues to 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. 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. This failure of supermodularity can have a bite: we have examples for three-person committees for which, for the intermediate voting rule requiring two or more yes votes for a candidate to be accepted, an increase in the alignment of members' interests results in lower, rather than higher, equilibrium reservation values (that is, committee members become less choosy).

³³It 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.

7.3 Systematic and Systemic Risk

Macroeconomists need to be able to gauge and compare levels of “systematic risk”. At the level of a single country, this involves assessing the degree of covariation among levels of output in different sectors, while at the level of the world economy, it involves assessing the degree of interdependence among output levels in different countries. Hennessy and Lapan (2003) have proposed using the supermodular ordering to make such comparisons. In the actuarial literature, the supermodular ordering has recently received considerable attention as a means of comparing the degrees of dependence among claims in a portfolio of insurance policies (see Müller and Stoyan, 2002, and Denuit, Dhaene, Goovaerts, and Kaas, 2005). In finance, this ordering has been proposed as a method for assessing both the dependence among asset returns in a portfolio (Epstein and Tanny, 1980) and the interdependence between a single institution’s portfolio and the market as a whole (Patton, 2009). Our Theorem 5 provides a flexible method for generating or modeling distributions that are comparable according to the supermodular ordering, by changing the relative importance of common and idiosyncratic shocks.

Moreover, the recent financial crisis has stimulated the development of measures of interdependence for the components of the financial system as a whole (measures of “systemic risk”) and not just for individual assets. 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, while Diebold and Yilmaz (2011) develop measures of connectedness for the system. Beale et al (2011) and Allen, Babus, and Carletti (2012) focus on understanding the interplay between the effect of diversification at the level of the financial institution, which lowers individual risk, and increasing similarity of institutions’ portfolios, which raises systemic risk.

Allen, Babus, and Carletti (2012) model a particular diversification strategy of banks, namely asset-swapping, and they examine how the pattern of asset swaps, in conjunction with the maturity of bank debt, affect market outcomes and welfare. Here we generalize a stylized version of their model, focusing on how different patterns of asset swaps (represented by different networks) generate multivariate distributions of bank failures with different degrees of interdependence. We use the symmetric supermodular ordering as a measure of systemic risk.

Consider, for example, the six banks in the model of Allen et al (2012) and the two networks of asset swaps they analyzed, the “clustered” and the “unclustered”. Each bank $i \in \{1, \dots, 6\}$ funds a project, whose return, $\theta_i \in \{L, H\}$. The returns θ_i are independent and identically distributed, 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, $Y'_i = \frac{1}{3} \sum_{j=4}^6 \theta_j$ for $i \geq 4$. 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. In that case, therefore, $X'_i = \frac{1}{3}(\theta_{i-1 \bmod 6} + \theta_i + \theta_{i+1 \bmod 6})$ for all i . The marginal distribution of each bank’s return is the same in the two networks, but the degree of interdependence of bank returns

differs. Thus, this is a setting where the degree of diversification is held fixed, and only systemic risk varies. Suppose a bank fails (default status=1) if its return is less than or equal to some level $d \in [L, H)$, otherwise it is solvent (default status=0). Let banks' default statuses in the clustered network be described by $(Y_1, \dots, Y_6) \in \{0, 1\}^6$, so $Y_i = I_{\{Y'_i \leq d\}}$, and in the unclustered network by $(X_1, \dots, X_6) \in \{0, 1\}^6$, so $X_i = I_{\{X'_i \leq d\}}$.

To compare the interdependence in the distribution of bank failures across the two different networks, we adopt a systemic cost function which is a supermodular and symmetric function of banks' default statuses: supermodularity reflects the judgment that the additional cost to the system from two bank failures is higher than the sum of the marginal costs from each individual failure, and symmetry reflects the fact that the banks in this setting are of equal size. Since the vectors of default statuses are binary random vectors, Proposition 5 implies that the distribution of (Y_1, \dots, Y_6) dominates the distribution of (X_1, \dots, X_6) according to the symmetric supermodular ordering if and only if the total number of bank defaults in the clustered network dominates the total number in the unclustered network according to the univariate convex ordering. It is straightforward to use this equivalence to show that, for any probability of project success p and for any common failure threshold d for banks, $(X_1, \dots, X_6) \prec_{SSPM} (Y_1, \dots, Y_6)$. Hence for any supermodular and symmetric systemic cost function, expected systemic cost is higher under the clustered than under the unclustered network.^{34 35}

7.4 Prediction and Parameter Estimation

The supermodular ordering may be used to compare the “richness” of data samples to estimate parameters. This section makes a first step in this direction. The question is approached from the viewpoint of an expert who gets rewarded based on the accuracy of his prediction, $\tilde{\theta}$, about the value of an unknown parameter θ . The parameter θ is revealed after the expert has made his prediction. For example, one may compare predicted and realized earnings of a firm. The expert gets an “accuracy payment” $\pi(\tilde{\theta} - \theta)$, where π is concave and maximized at 0, and receives utility $u(\pi(\tilde{\theta} - \theta))$, where u is increasing and concave. The expert's prediction is based on some data (X_1, \dots, X_n) , where the distribution $F_i(\cdot|\theta)$ of X_i , conditional on θ , lies in some fixed support \mathcal{L}_i . We focus, for this illustration, on the case in which the estimator $\tilde{\theta}$ is an affine function of the

³⁴It is not the case that $(X_1, \dots, X_6) \prec_{SPM} (Y_1, \dots, Y_6)$. To see why, observe that in the clustered network, banks 1 and 6 are in different clusters, so Y_1 and Y_6 are independent; but in the unclustered network, banks 1 and 6 directly swap assets, so X_1 and X_6 are positively dependent, and hence, since the support of (X_1, X_6) and (Y_1, Y_6) is $\{0, 1\}^2$, $(X_1, X_6) \succ_{SPM} (Y_1, Y_6)$. 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. Compare the remark in the footnote immediately preceding Proposition 7.

³⁵Allen et al (2012) restrict attention to the case where $p = \frac{1}{2}$ and where a bank defaults if and only if the return on all three of the projects in its portfolio is L (i.e. $d = L$). Their model involves additional features, such as different maturities of debt, through which interdependence of banks' returns indirectly influences welfare.

observed variables:³⁶ $\tilde{\theta} = \sum \kappa_i X_i$ for some nonnegative weights $\{\kappa_i\}_{i=1,\dots,n}$.

Our objective is to find an ordering for the “richness” of the dataset used by the expert, holding fixed the marginal distributions $F_i(\cdot|\theta)$ for each individual observation. Intuitively, the dataset is “rich” if it comes from multiple independent sources about the value of θ , instead of synthesizing closely related opinions or sources of information. The next proposition shows that if one dataset displays more interdependence than another in the sense of the supermodular ordering, for all true θ 's, then the more interdependent dataset is less valuable to the expert.

PROPOSITION 8 *Let the datasets (X_1, \dots, X_n) and (Y_1, \dots, Y_n) be generated by joint distributions $F(\cdot|\theta)$ and $G(\cdot|\theta)$, respectively. Suppose that $F(\cdot|\theta) \prec_{SPM} G(\cdot|\theta)$ for all θ . Then, $E[u(\pi(\tilde{\theta}-\theta))|\theta] \geq E[u(\pi(\tilde{\theta}-\theta))|\theta]$ for all $\{\kappa_i\}$, $u(\cdot)$, $\pi(\cdot)$, and θ .*

Proof. The function $-u \circ \pi$ is convex, as is easily shown. Therefore, $(x_1, \dots, x_n) \mapsto -u(\pi(\sum_i \kappa_i x_i - \theta))$ is supermodular in (x_1, \dots, x_n) , for all θ and nonnegative vectors $\{\kappa_i\}$. The conclusion follows from the definition of the supermodular ordering. ■

Proposition 8 has a purely statistical interpretation, as a special case. One possible objective for the expert is to minimize $(\tilde{\theta} - \theta)^2$ (letting u be the identity and $\pi(z) = -z^2$), or more generally, $(\tilde{\theta} - \theta)^{2p}$ for any integer p . Proposition 8 then implies that the variance, as well as any even higher-order moment of the error, is lower under F than it is under G .

8 Conclusion

The supermodular ordering is relevant for both theoretical and empirical work. In empirical work, the constructive methods developed in this paper allow comparisons of portfolios according to the interdependence among their assets' returns, of systemic risk in financial sectors according to the interdependence among the returns of banks, and of multidimensional inequality across countries or time periods by taking into account the interdependence among the different dimensions affecting economic wellbeing. The ordering is also relevant for econometrics: We have shown that it bears a close relation to the study of copulas and that it can be used to compare the richness of datasets for parameter estimation.

In economic theory, we have illustrated how our characterization results for the supermodular ordering can be used to rank reward schemes in the presence of uncertainty according to the amount of ex post inequality they generate. The supermodular ordering is also 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. While applications to two-sided matching problems

³⁶While special, affine estimators are pervasive in statistics and econometrics. If, for example, (X_1, \dots, X_n) are exchangeable and have mean θ , then $\tilde{\theta}$ will be the sample average of those variables.

have received some attention,³⁷ multi-dimensional applications remain largely unexplored.³⁸

This paper has focused primarily on comparing multivariate distributions according to the value that their degree of interdependence provides to agents or to a social planner. Another important question concerns the relationship between increased interdependence and the comparative statics of decisions.³⁹ We provided an example of this type of investigation in a multidimensional context in Section 7.2, when we analyzed the impact of greater alignment in preferences, in the sense of the supermodular ordering, on the search and voting behavior of committees of arbitrary size.⁴⁰ A more systematic exploration of the role of the supermodular ordering in comparative statics analysis of decisions should be a fruitful area for future research.

In their influential work on monotone comparative statics, Milgrom and Shannon (1994) emphasize in their conclusion the need for a better theory of what “more correlated” means, and they suggest that supermodularity of objective functions may be a promising approach for constructing such theory. By characterizing the supermodular ordering in terms of elementary transformations, extending it to the symmetric and increasing supermodular orderings, and providing, in Sections 4 and 6, general methods for constructing distributions comparable according to these orderings, we have taken a new step in this direction.

References

- ACHARYA, V.V., PEDERSEN, L.H., PHILIPPON, T., AND RICHARDSON, M. (2010) “Measuring Systemic Risk,” N.Y.U. W.P.
- ADLER, M. D., AND SANCHIRICO, C. W. (2006) “Inequality and Uncertainty: Theory and Legal Applications,” *U. of Pennsylvania Law Review*, 155, 279-377.
- 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.

³⁷Fernandez and Gali (1999) used 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.

³⁸One exception is Prat (2002), but he compares only a perfectly positively dependent joint distribution with an independent one.

³⁹The literature on affiliation (introduced into economics by Milgrom and Weber, 1982) is relevant here, but this literature is limited by the facts that affiliation is an extremely stringent condition (see de Castro, 2009) and that there is no widely accepted notion of “greater affiliation”.

⁴⁰Gollier (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.

- ARLOTTO, A., AND SCARSINI, M. (2009) “Hessian Orders and Multinormal Distributions,” *J. of Multivariate Analysis*, 100, 2324-30.
- 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?” *11th Annual ACM Symposium on Computational Geometry*, Vancouver, B.C., 20-8.
- 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.
- BOURGUIGNON, F., AND CHAKRAVARTY, S.R. (2002) “Multi-Dimensional Poverty Orderings,” DELTA W.P., 2002-22.
- BURTSCHHELL, X., GREGORY, J., AND J.-P. LAURENT (2008) “A Comparative Analysis of CDO Pricing Models,” Working paper, BNP Paribas.
- CAILLAUD, B., TIROLE, J. (2007) “Consensus Building: How to Persuade a Group,” *American Economic Review*, 97, 1877–900.
- DE CASTRO, L. (2009) “Affiliation and Dependence in Economic Models,” D.P. No. 1479, Ctr. for Math. Stud. in Econ. and Mgt. Sci., Northwestern Univ.
- 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.
- Economica*, 79, 721-746.
- DENUIT, M., DHAENE, J., GOOVAERTS, M.J., AND KAAS, R. (2005) *Actuarial Theory for Dependent Risks: Measures, Orders, and Models*, Wiley.

- DIEBOLD, F.X., AND YILMAZ, K. (2011) “On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms,” NBER W.P. 17490.
- 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.
- 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.
- 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., XIE, C., AND RUAN, L. (2005) “Dependence Structures of Multivariate Bernoulli Random Vectors,” *J. of Multivariate Analysis*, 94, 172–95.
- HU, T., AND YANG, J. (2004) “Further Developments on Sufficient Conditions for Negative Dependence of Random Variables,” *Statistics and Probability Letters*, 66, 369–81.
- IBRAGIMOV, R. (2005) “Copula-Based Dependence Characterizations and Modeling for Time Series,” *Harvard Institute of Economic Research*, D.P. No. 2094.

- 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.
- KROLL, Y., AND DAVIDOVITZ, L. (2003) “Inequality Aversion versus Risk Aversion,” *Economica*, 70, 19–29.
- 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 SHANNON, C. (1994) “Monotone Comparative Statics,” *Econometrica*, 62, 157–80.
- MILGROM, P., AND WEBER, R. (1982) “A Theory of Auctions and Competitive Bidding,” *Econometrica*, 50, 1089-122.
- MOLDOVANU, B., AND SHI, X. (2012) “Specialization and Partisanship in Committee Search,” *Theoretical Economics*, forthcoming.
- MOTZKIN, T., RAIFFA, H., THOMPSON, G., AND THRALL, R. (1953) “The Double Description Method”. In H. W. Kuhn and A. W. Tucker Ed., *Contributions to the Theory of Games II*, 51–73.
- 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. (2007) “On the Theory of Strategic Voting,” *Review of Economic Studies*, 74, 255-81.

- 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.
- ROTHSCHILD, M., AND STIGLITZ, J. E. (1970) “Increasing Risk: I. A Definition,” *J. of Economic Theory*, 2, 225–43.
- 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.

Appendices

A Proof of Theorem 2

The equivalence of conditions 2) and 3) follows from Theorem 1, the definition of γ , and the decomposition result in (8). 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 $\delta = g - f$, $g \succ_{\mathcal{I}SPM} f$ 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 (10) 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 $i \in \{1, \dots, n\}$ and $k \in \{1, \dots, m_i - 1\}$, $w(z) = I_{\{z_i \geq k\}}$, $g \succ_{\mathcal{I}SPM} f$ 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). \blacksquare

B Proof of Theorem 3

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

$$g = f + \sum_{t \in \mathcal{T}(\mathcal{L})} \alpha_t t, \quad (23)$$

where $\mathcal{T}(\mathcal{L})$ is the set of elementary transformations on \mathcal{L} . Let Φ denote the operator which to any function w on \mathcal{L} associates the function on $\tilde{\mathcal{L}}$ defined by $\Phi(w)(\tilde{x}) = \sum_{x \in \mathcal{L}: \phi(x)=\tilde{x}} w(x)$. Φ is a linear operator, and by construction, $\tilde{f} = \Phi(f)$. Applying Φ to (23) yields

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

For any $t = t_{i,j}^x \in \mathcal{T}(\mathcal{L})$, either $\Phi(t)$ belongs to $\mathcal{T}(\tilde{\mathcal{L}})$, which is the case if and only if $\phi(x)$, $\phi(x + e_i)$, $\phi(x + e_j)$, and $\phi(x + e_i + e_j)$ are all distinct, or $\Phi(t)$ is everywhere equal to zero. Therefore,

$$\tilde{g} = \tilde{f} + \sum_{\tilde{t} \in \mathcal{T}(\tilde{\mathcal{L}})} \tilde{\alpha}_{\tilde{t}} \tilde{t},$$

for some nonnegative coefficients $\{\tilde{\alpha}_{\tilde{t}}\}_{\tilde{t} \in \mathcal{T}(\tilde{\mathcal{L}})}$. Applying Theorem 1 again, but in the reverse direction, proves that $\tilde{f} \prec_{SPM} \tilde{g}$. \blacksquare

C Proof of Proposition 3 (For Online Appendix)

Without loss of generality, we prove the claim for the case where $\mathcal{L}_i = \{0, 1, \dots, m_i - 1\}$ (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 $t^* \equiv t_{i,j}^{x-e_i-e_j}$ is well-defined. Suppose that, contrary to the claim, there exist nonnegative coefficients α_s such that

$$t^* = \sum_{s \in \mathcal{T} \setminus \{t^*\}} \alpha_s s. \quad (24)$$

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 $t_{k,l}^{x-e_k-e_l}$ 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 (24) 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 $t_{k,l}^{x-e_k-e_l}$ 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 $t_{k,l}^{x-e_k-e_l}$ 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. \blacksquare

D Constructive Methods for Comparing Distribution Interdependence using the Supermodular Ordering

D.1 The Linear Programming Approach: Comparing Two Specific Distributions

From Theorem 1, $f \prec_{SPM} g$ if and only there exist nonnegative coefficients $\{\alpha_t\}_{t \in \mathcal{T}}$ such that $g - f = \sum_{t \in \mathcal{T}} \alpha_t t$. Given a specific pair of distributions f and g , we can formulate the problem of determining whether such a set of coefficients exists as a linear programming problem. Let $T = |\mathcal{T}|$ 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 $f \prec_{SPM} g$ if and only if there exists $\alpha \in \mathbb{R}^T$ such that i) $\alpha \geq 0$ and ii) $E\alpha = g - f$. Now define the d -dimensional vector δ^+ such that $\delta_i^+ = |(g - f)_i|$, and let E^+ denote the matrix whose i^{th} row, denoted E_i^+ , satisfies $E_i^+ = (-1)^{\varepsilon_i} E_i$, where $\varepsilon_i = 1$ if $(g - f)_i < 0$ and 0 otherwise. The condition $E\alpha = g - f$ can be re-expressed as $E^+\alpha = \delta^+$. Now consider the following⁴¹ linear program (A):

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

subject to

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

THEOREM 7 (PAIRWISE COMPARISON) *The linear program (A) always has an optimal solution. $f \prec_{SPM} g$ if and only if the optimum value is zero, and in that case $g = f + \sum_{t \in \mathcal{T}} \alpha_t^* t$, where (α^*, β^*) is any minimizer of (A) and $\beta^* = 0$.*

⁴¹This corresponds to the auxiliary program for the determination of a basic feasible solution described in Bertsimas and Tsitsiklis (1997, Section 3).

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

D.2 The Double Description Method

The linear programming approach just described has the drawback of requiring a new program to be solved each time a new pair of distributions is to be compared.

When many distributions are to be compared, for example as part of a larger optimization problem, it is more convenient to have an explicit representation of the supermodular ordering for the common support of these distributions. We present a method for generating such a representation in the form of a list of inequalities that are satisfied by the vector $g - f$ if and only if $f \prec_{SPM} g$. For any given finite support \mathcal{L} , this method computes these inequalities once and for all, a computation made possible by the support's finiteness.

Recall that $f \prec_{SPM} g$ 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 inequalities by exploiting the geometric properties of \mathcal{S} . \mathcal{S} is a convex cone characterized by the fact that w is supermodular (i.e., belongs to \mathcal{S}) if and only if it makes a nonnegative scalar product with all elementary transformations on \mathcal{L} as defined by (3). In matrix form, $\mathcal{S} = \{w \in \mathbb{R}^d : Aw \geq 0\}$, where $A = E'$ is the matrix whose rows consist of all elementary transformations (i.e., the transpose of the matrix E introduced in the previous subsection). A is called the *representation matrix* of the polyhedral cone \mathcal{S} . Minkowski's theorem states that to any representation matrix corresponds a *generating matrix* R such that

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

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

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

Minkowski's theorem thus proves the existence, for any finite support \mathcal{L} , of a finite list of inequalities 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), builds on Minkowski's and Weyl's representation theorems for polyhedral cones. A polyhedral cone can be represented either by a set of inequalities (i.e., by the intersection of a number half-spaces) or by extreme rays. The double description method provides an algorithm to determine one description from the other. The set of elementary transformations defined by (3) is trivially computable, and can be automatically generated for any given support \mathcal{L} . From this input, the double description method can compute the set of extreme supermodular functions. Using Fukuda's algorithm for the double description method, we have computed the inequalities characterizing the supermodular order for a range of problems that are intractable by hand.

D.3 Complexity of the Double Description Method (For Online Appendix)

Although the double description method is very useful in theory, its computational complexity is unsurprisingly exponential in the size of \mathcal{L} . We now provide an exact computation of the algorithm’s complexity.

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

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

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 Theorem 3 showed, 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.

E Continuous Support: Proof of Theorem 4

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

$$w_N(\tilde{x}) + w_N(\tilde{x} + \tilde{e}_i + \tilde{e}_j) \geq w_N(\tilde{x} + \tilde{e}_i) + w_N(\tilde{x} + \tilde{e}_j). \quad (25)$$

Given the equal spacing of the chosen partition, the denominator arising in (14) is the same for all \tilde{x} ’s. Therefore, (25) reduces to showing that⁴²

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

⁴²Because 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.

where $d_i = |\mathcal{L}_i|/N$ is the length of each hyperrectangle along dimension i (with a similar definition for α_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 . There remains to show that $E[w_N|F_N]$ converges to $E[w|F]$ as $N \rightarrow +\infty$. 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} ,

$$\left| \int_{x \in \tilde{x}} (w_N(\tilde{x}) - w(x))f(x) \right| = \int_{x \in \tilde{x}} |(w_N(\tilde{x}) - w(x))(f(x) - f(\chi(\tilde{x})))| dx. \quad (26)$$

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 (26), 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

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

F Proofs for Section 4

Proof of Lemma 1: 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 $\bar{p}_r^i = P(X_i \geq r)$, $r \in \{1, \dots, m\}$. Define

$$\hat{w}(\bar{p}^1, \dots, \bar{p}^n) = E[w(X_1, \dots, X_n) | \bar{p}^1, \dots, \bar{p}^n]. \quad (27)$$

We will use the following lemma both now and in Section F.1.3, when we consider asymmetric distributions.

LEMMA 5 *For any supermodular w , $\hat{w}(\bar{p}^1, \dots, \bar{p}^n)$ has the following properties:*

$$\begin{aligned} \frac{\partial^2 \hat{w}}{\partial \bar{p}_r^i \partial \bar{p}_s^i} &= 0 \text{ for all } i \in \{1, \dots, n\} \text{ and } r, s \in \{1, \dots, m\}, \\ \frac{\partial^2 \hat{w}}{\partial \bar{p}_r^i \partial \bar{p}_s^j} &\geq 0 \text{ for all } i \neq j \in \{1, \dots, n\} \text{ and } r, s \in \{1, \dots, m\}. \end{aligned}$$

Proof. The first part of the lemma is standard, and comes from the linearity of the objective with respect to the probability distribution, which holds also in terms of the cumulative distribution vector. The second

part comes from supermodularity of w . Indeed, by the discrete equivalent of an integration by parts,⁴³ we have

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

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

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

which is nonnegative, by supermodularity of w . ■

To conclude the proof of Lemma 1, observe that $\bar{w}(\bar{p}) = \hat{w}(\bar{p}, \dots, \bar{p})$. Second-order derivatives of \bar{w} involve only second-order derivatives of \hat{w} . Lemma 5 then yields the result.

F.1 Proof of Theorem 5

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

Let A and B denote $q \times m$ row-stochastic matrices and \bar{A} and \bar{B} their cumulative-sum equivalents, so $\bar{A}_{i,k}$ and $\bar{B}_{i,k}$ lie in $[0, 1]$ and are decreasing in k , and for each i , $\bar{A}_{i,1} = \bar{B}_{i,1} = 1$. Let \bar{A} be stochastically ordered, so $\bar{A}_{i,k}$ is increasing in i . Finally, let $A \succ_{CCM} B$, so for each k , the column vector $\bar{A}_{\bullet,k}$ majorizes the column vector $\bar{B}_{\bullet,k}$.

F.1.1 Step 1: Proof when \bar{B} is stochastically ordered

When \bar{B} is stochastically ordered, its entries are nondecreasing with the row index and nonincreasing in the column index. Throughout the proof, we *exclude* the first column of ones from cumulative probability matrices, which play no role in the analysis. Having removed that column, we will first consider the case in which \bar{B} has strictly monotonic entries across row and column indices, so that

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

Analysis when \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 which are dominated by \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 6 *There exists a stochastically ordered cumulative-probability matrix C such that i) $C_{\bullet,\bar{k}} = \bar{B}_{\bullet,\bar{k}}$ for all $\bar{k} \leq k$; 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} .*

⁴³The 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. Let C solve the optimization problem

$$\inf_E \sum_{i \geq 2} \left(\sum_{j \geq i} E_{j,k} \right) \quad (28)$$

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 (28) 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 that⁴⁴

$$C_{i,k} \leq \bar{B}_{i,k} \quad \text{and} \quad C_{i+1,k} > \bar{B}_{i+1,k}. \quad (29)$$

We will show that it is possible to increase $C_{i,k}$ by a small amount ε , 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 (28). Such change only affects the $i + 1$ partial sum of (28), and decreases it by an amount ε , which will yield the desired contradiction.

Let \tilde{k} denote the largest column index such that $C_{i+1, \tilde{k}} = C_{i+1,k}$ for all $\tilde{k} \in [k, \tilde{k}]$,⁴⁵ and let D denote the matrix that is identical to C for all rows other than i and $i + 1$ and for all columns outside of $[k, \tilde{k}]$, and such that

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

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

We first check D is row-monotonic for ε 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 ε 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 $C_{i,k} \leq \bar{B}_{i,k} \leq$

⁴⁴The 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 $i = \max\{j < \bar{i} : C_{j,k} \leq \bar{B}_{j,k}\}$.

⁴⁵Possibly, \tilde{k} is equal to the number of columns of C .

$\bar{B}_{i,k-1} - \chi = C_{i,k-1} - \chi$. 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} .⁴⁶

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 $\varepsilon \leq \chi/2$,⁴⁷ 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

$$\sum_{j \geq i+1} D_{j,\tilde{k}} \geq \sum_{j \geq i+1} \bar{B}_{j,\tilde{k}} \quad (30)$$

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

$$\sum_{j \geq i+2} D_{j,\tilde{k}} = \sum_{j \geq i+2} C_{j,\tilde{k}} \geq \sum_{j \geq i+2} \bar{B}_{j,\tilde{k}} \quad (31)$$

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 (29)),

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

Combining this with (31) implies (30).

Finally, because the rows i and $i + 1$ of the matrices C and D satisfy the assumptions of Lemma 2, it follows that the mixture distribution corresponding to C SPM-dominates that corresponding to D .⁴⁸ 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 ε , thus providing the desired contradiction. \blacksquare

To conclude the proof of Step 1 of Theorem 5, it suffices to apply Lemma 6 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} .

Proof when \bar{B} is not strictly monotonic

When \bar{B} is not strictly monotonic, we approximate \bar{A} and \bar{B} by a sequence of cumulative 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,

⁴⁶If \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.

⁴⁷Indeed, we have $C_{i,k} \leq C_{i+1,k} - \chi$ from both inequalities of (29) and strict monotonicity of \bar{B} .

⁴⁸Lemma 2 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 2 applies to arbitrarily many rows, as long as only two rows are changed.

respectively, to \bar{A} and \bar{B} as $N \rightarrow \infty$. 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} ,⁴⁹ 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.

F.1.2 Step 2: Proof when \bar{B} is not stochastically ordered

Let \bar{B}^{so} denote the stochastically ordered version of \bar{B} , whose k^{th} column consists of the entries of the k^{th} column of \bar{B} , ordered from the smallest to the largest. \bar{B}^{so} is also row monotonic. Indeed, $\bar{B}_{i,k}^{so}$ 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 $\bar{B}_{i,k+1}^{so}$. 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 2. 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 $1, 2, \dots, q$, by their ordinal values $1, 2, \dots, q$, and we use the same cardinal-to-ordinal transformation to label the entries in each column of \bar{B} .⁵⁰ 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 2, 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 no longer stochastically ordered, but both rows still (stochastically) dominate all rows with indices less than

⁴⁹Recall that we have excluded the first column of ones that may appear in cumulative matrices.

⁵⁰For 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 $\bar{B}_{1,2}^{so} = .1$, $\bar{B}_{2,2}^{so} = .3$, and $\bar{B}_{3,2}^{so} = .4$, then entries are converted to $\bar{B}_{1,2} = 2$, $\bar{B}_{2,2} = 3$, and $\bar{B}_{3,2} = 1$, so that $\bar{B}_{1,2}^{so} = 1$, $\bar{B}_{2,2}^{so} = 2$, and $\bar{B}_{3,2}^{so} = 3$. If there are ties, the way ties are broken does not matter, as is clear from the algorithm.

$q - 2$. 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 $q - 1$, $q - 2$, \dots , 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 2 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, $\bar{B}_{i,k}^{so}$, of $\bar{B}_{\bullet,k}^{so}$ if $D_{q,k}$ is smaller than $\bar{B}_{i,k}^{so}$, 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, $D(i)_{q,k} = \bar{B}_{i,k}^{so}$, then we use the fact that $D(i)_{q,k-1} \geq \bar{B}_{i,k-1}^{so} \geq \bar{B}_{i,k}^{so}$. This demonstrates row monotonicity of $D(i)$, for all $i \in \{1, \dots, q - 1\}$ and, hence, the applicability of Lemma 2 for each transformation described in the algorithm above.

F.1.3 Step 3: Asymmetric Distributions

Using the definition of the function \hat{w} in (27), we can write

$$\begin{aligned} Ew(X_1, \dots, X_n) &= \frac{1}{q} \sum_{i=1}^q \hat{w}(\bar{A}(1)_{i,\bullet}, \dots, \bar{\alpha}(n)_{i,\bullet}) \\ Ew(Y_1, \dots, Y_n) &= \frac{1}{q} \sum_{i=1}^q \hat{w}(\bar{B}(1)_{i,\bullet}, \dots, \bar{B}(n)_{i,\bullet}), \end{aligned} \quad (32)$$

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 $\bar{B}(r)_{i,\bullet}^{so}$ 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 2 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 the proof of Lemma 2, we must check that

$$\hat{w}(\bar{p}, \bar{p}(2), \dots, \bar{p}(n)) + \hat{w}(\bar{q}, \bar{q}(2), \bar{q}(n)) \geq \hat{w}(\bar{p} + \varepsilon, \bar{p}(2), \dots, \bar{p}(n)) + \hat{w}(\bar{q} - \varepsilon, \bar{q}(2), \dots, \bar{q}(n)), \quad (33)$$

where, for $r \geq 2$, $\bar{p}(r) = \bar{A}(r)_{i,\bullet}$ and $\bar{q}(r) = \bar{A}(r)_{j,\bullet}$. Generalizing the argument used to prove Lemma 2, we have

$$\hat{w}(\bar{p} + \varepsilon, \bar{p}(2), \dots, \bar{p}(n)) - \hat{w}(\bar{p}, \bar{p}(2), \dots, \bar{p}(n)) = \int_0^1 \sum_{k=2}^m \frac{\partial \hat{w}}{\partial \bar{p}_k^1}(\bar{p} + \alpha \varepsilon, \bar{p}(2), \dots, \bar{p}(n)) \varepsilon_k d\alpha,$$

and similarly,

$$\hat{w}(\bar{q}, \bar{q}(2), \dots, \bar{q}(n)) - \hat{w}(\bar{q} - \varepsilon, \bar{q}(2), \dots, \bar{q}(n)) = \int_0^1 \sum_{k=2}^m \frac{\partial \hat{w}}{\partial \bar{p}_k^1}(\bar{q} - \varepsilon + \alpha \varepsilon, \bar{q}(2), \dots, \bar{q}(n)) \varepsilon_k d\alpha.$$

Letting $\delta(1) = \bar{q} - \varepsilon - \bar{p}$ and $\delta(r) = \bar{q}(r) - \bar{p}(r)$ for $r \geq 2$, we have, for each $k \in \{2, \dots, m\}$,

$$\begin{aligned} \frac{\partial \hat{w}}{\partial \bar{p}_k^1}(\bar{q} - \varepsilon + \alpha \varepsilon, \bar{q}(2), \dots, \bar{q}(n)) &- \frac{\partial \hat{w}}{\partial \bar{p}_k^1}(\bar{p} + \alpha \varepsilon, \bar{p}(2), \dots, \bar{p}(n)) \\ &= \int_0^1 \sum_{r=1}^n \sum_{k=2}^m \frac{\partial^2 \hat{w}}{\partial \bar{p}_k^1 \partial \bar{p}_k^r}(\bar{p} + \alpha \varepsilon + \beta \delta(1), \beta \delta(2), \dots, \beta \delta(n)) \delta_{\bar{k}}(r) d\beta \\ &\geq 0, \end{aligned} \quad (34)$$

where the inequality follows from the fact, as established in Lemma 5, that all cross-partial derivatives of \hat{w} are nonnegative. Summing (34) over k and integrating over α then shows that (33) holds.

Inequality (33) 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 2 can be applied. Therefore,

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

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

$$\sum_{i=1}^q \hat{w}(\bar{A}(1)_{i,\bullet}, \bar{A}(2)_{i,\bullet}, \dots, \bar{A}(n)_{i,\bullet}) \geq \sum_{i=1}^q \hat{w}(\bar{B}(1)_{i,\bullet}^{so}, \bar{A}(2)_{i,\bullet}, \dots, \bar{A}(n)_{i,\bullet}) \quad (35)$$

$$\geq \sum_{i=1}^q \hat{w}(\bar{B}(1)_{i,\bullet}^{so}, \bar{B}(2)_{i,\bullet}^{so}, \dots, \bar{A}(n)_{i,\bullet}) \quad (36)$$

$$\geq \dots \quad (37)$$

$$\geq \sum_{i=1}^q \hat{w}(\bar{B}(1)_{i,\bullet}^{so}, \bar{B}(2)_{i,\bullet}^{so}, \dots, \bar{B}(n)_{i,\bullet}^{so}). \quad (38)$$

Finally, we use the algorithm described in Section F.1.2 to convert $\bar{B}^{so}(r)$ into $\bar{B}(r)$ for all r simultaneously. Supermodularity and componentwise-convexity of \hat{w} ensure that

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

Combining this with (35) and (32) then yields $Ew(X_1, \dots, X_n) \geq Ew(Y_1, \dots, Y_n)$ for all supermodular w .

G Proof of Theorem 6 and Proposition 6

The proof of Theorem 6 is closely related to the proof of Theorem 5, using, almost without change, the arguments of Sections F.1.1 and F.1.2. There are two key differences. The first is to replace Lemma 2 by Lemma 3. Following on from this, the second is that wherever, in the statement or proof of Theorem 5, the relation \succ_{SPM} appears, at the corresponding point in the proof of Theorem 6, the relation \prec_{SSPM} must appear instead.

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 F.1.1. 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 3, instead of Lemma 2. Because each row now represents the distribution of a different random variable, and random variables are independently distributed, Lemma 4 guarantees that the result of Lemma 3 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 F.1.2, again replacing Lemma 2 by Lemma 3. Because each step preserves row monotonicity, as shown in Section F.1.2, all rows correspond to actual probability distributions. This ensures that once again, Lemma 4 can be applied at every step to extend the result of Lemma 3 to the multivariate distributions over all n variables.

G.1 Proof of Proposition 6 (For Online Appendix)

Proof of *i*): Assume that $p_j > 0$ for all $j \in \{1, \dots, m\}$. (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, \dots, 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 $(0, \dots, 0, \lambda_k, 1, \dots, 1)'$, 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 λ_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 λ_k while the entry in column $k + 1$ is 0, for $i = i_{k+1}$, the entry in column k is 1 while that in column $k + 1$ is λ_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 λ_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 stochas-

tically 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 6.

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 all of whose components are equal. 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 6.