Discussion Paper Series

RIEB

Kobe University

DP2014-23

Partial Stochastic Dominance

Takashi KAMIHIGASHI
John STACHURSKI

May 19, 2014

Research Institute for Economics and Business Administration
Kobe University
2-1 Rokkodai, Nada, Kobe 657-8501 JAPAN

Partial Stochastic Dominance[1]

Takashi Kamihigashi[a,b] and John Stachurski[c]

[a]Research Institute for Economics and Business Administration, Kobe University

[b]IPAG Business School, Paris, France

[c]Research School of Economics, Australian National University

May 29, 2014

ABSTRACT. The stochastic dominance ordering over probability distributions is one of the most familiar concepts in economic and financial analysis. One difficulty with stochastic dominance is that many distributions are not ranked at all, even when arbitrarily close to other distributions that are. Because of this, several measures of ″partial″ or ″near″ stochastic dominance have been introduced into the literature—albeit on a somewhat ad hoc basis. This paper argues that there is a single measure of extent of stochastic dominance that can be regarded as the most natural default measure from the perspective of economic analysis.

JEL Classifications: D81, G11

Keywords: Stochastic dominance, stochastic order

1. INTRODUCTION

Stochastic dominance is a fundamental concept in economics, finance and operations research, which permits comparison of outcomes in environments involving

[1]This paper has benefited from helpful discussions with Kieron Meagher and Rabee Tourky, and financial support from the Japan Society for the Promotion of Science and Australian Research Council Discovery Grant DP120100321.

Email addresses: tkamihig@rieb.kobe-u.ac.jp, john.stachurski@anu.edu.au

uncertainty. The basic properties of first order stochastic dominance were initially provided by Lehmann (1955), and introduced to economics by Quirk and Saposnik (1962). As stated by Border (1992), "the concept has been independently discovered too many times for an exhaustive listing." (Here and below, stochastic dominance always refers to first order stochastic dominance unless otherwise specified.)

For probability distributions μ and ν on some set S, stochastic domination of μ by ν means that if X and Y are S-valued random elements with distributions μ and ν respectively, then $\mathbb{E}\, h(X) \leq \mathbb{E}\, h(Y)$ for any increasing bounded function $h\colon S \to \mathbb{R}$. In the one dimensional setting, stochastic domination is equivalent to pointwise ordering of the cumulative distribution functions (cdfs). In particular, cdf F is (weakly) stochastically dominated by cdf G if $G(x) \leq F(x)$ for all $x \in \mathbb{R}$. See, for example, section 6.D of Mas-Colell et al. (1995).

Stochastic dominance is one of the cornerstones of the theory of choice under uncertainty. One immediate consequence of the definition of stochastic dominance is that any von-Neumann–Morgenstern expected utility maximizer with monotone preferences will weakly prefer stochastically dominant distributions. In addition, extensions and alternatives to expected utility theory typically treat the notion of stochastic dominance as fundamental, either as an axiom or as a test of the implications of axiomatic treatments (see, e.g., Machina (1982), Tversky and Kahneman (1992), Halevy and Feltkamp (2005) or Safra and Segal (2008)).

The notion of stochastic dominance also plays a central role in the theory of finance. For example, stochastic dominance is closely related to existence of arbitrage opportunities. In standard settings, existence of arbitrage opportunities implies stochastic dominance and that the converse is true under relatively weak conditions (Jarrow (1986)). In addition, stochastic dominance is fundamental to the study of portfolio diversification, partly because many of the theorems of classical mean-variance portfolio theory apply only to certain classes of utility functions and return distributions (Hadar and Russell (1971)).

In social welfare studies, stochastic dominance is again fundamental, as policies or decision rules that shift outcomes up in terms of stochastic dominance can be regarded as universally preferred under very weak assumptions on individual choices. In fact Blackorby and Donaldson (1980) show that all social welfare functions satisfying three basic axioms are equivalent in the sense that they induce the same order on the set of distributions as first order stochastic dominance. Related analysis can be found in Atkinson (1987) and Foster and Shorrocks (1988).

Stochastic dominance also serves as a standard ordering for considering parametric monotonicity in settings where distributions are compared. These distributions can take the form of distributions over actions, beliefs, valuations, preferences, information or other related quantities (see, e.g., Maskin and Riley (2000), Brock and Durlauf (2001), Rosenberg et al. (2007), or Ambrus et al. (2013)). The most common notion of monotonicity in stochastic dynamic models is defined in terms of stochastic dominance (e.g., Stokey and Lucas (1989)). There is a large literature on econometric tests for stochastic dominance. See, for example, Linton et al. (2005) and Lee et al. (2009).

Further discussion of economic and financial applications can be found in standard textbooks such as Stokey and Lucas (1989) or Mas-Colell et al. (1995). More detailed treatments can be found in Levy (2006) and Föllmer and Schied (2011).

One difficulty with stochastic dominance is that the measure is fragile, in the sense that it typically fails to be invariant to small changes in the distributions under consideration. For example, it is well known that two normal distribution are ranked by stochastic dominance if and only if their variances are identical. As a consequence, any arbitrarily small change in variance for either of two ranked distributions will break the ranking. Given these facts, it is not surprising that normal distributions estimate from data are almost never ranked. Similar issues afflict estimation with other kinds of distributions, a point we return to below.

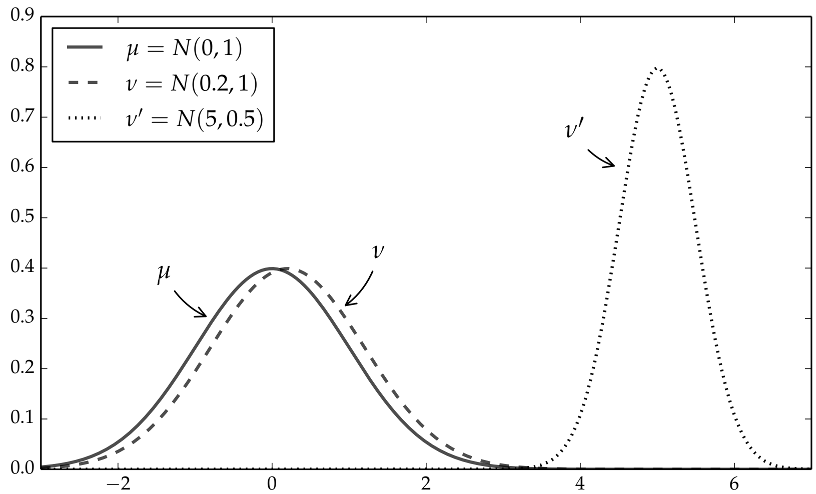

FIGURE 1.

To help illustrate ideas, consider the normal distributions μ, ν and ν' shown in figure 1. Let $\preceq_{sd}$ indicate stochastic dominance. Since ordering normal distributions requires identical variance, the ranking $\mu \preceq_{sd} \nu$ holds while $\mu \preceq_{sd} \nu'$ fails. Intuitively, $\mu \preceq_{sd} \nu'$ fails because ν' has smaller variance, and hence holds less mass in its tails. These thin tails mean that, for very large c, the probability that a draw from μ exceeds c is larger than the probability that a draw from ν' exceeds c. This violates stochastic dominance. It should be noted, however, that the smallest such c is larger than 10, and the amount of mass that either of these distributions puts on any of these sets is numerically indistinguishable from zero, even with 64 bit precision.

These issues translate into various difficulties in applied settings. While stochastic dominance is a popular tool in the theory of portfolio choice, in practice it often fails to provide effective analysis. As phrased by Levy (1992), "the difficulty with stochastic dominance rules is that, in general, [they result] in a large efficient set of investments, i.e. in many cases this framework is unable to rank the two risky options under consideration." On the empirical side, Davidson and Duclos (2013) show that, as a consequence of the sensitivity of the ordering, it is in fact *not possible* to reject non-dominance in favor of dominance over the entire supports of two distributions when these distributions are continuous. This ”necessitates a recasting of the usual theoretical links between stochastic dominance relationships and orderings in terms of poverty, social welfare, and expected utility. It also highlights better why a failure to reject the usual null hypothesis of unrestricted dominance *cannot* be interpreted as an acceptance of unrestricted dominance, since unrestricted dominance can never be inferred from continuous data” (Davidson and Duclos, 2013, p. 87).

Because of the relative fragility of stochastic dominance, several notions of *partial* stochastic dominance have been introduced into the literature. An early one is ”restricted” first order stochastic dominance up to some point c, which is defined (Atkinson, 1987, p. 751) as holding for one-dimensional cumulative distributions F and G whenever

$$G(x) \leq F(x) \quad \text{for all} \quad x \leq c. \tag{1}$$

(Recall that, for cdfs F and G on the real line, stochastic domination of F by G can equivalently be defined as $G(x) \leq F(x)$ at every x.) Restricted first order stochastic dominance is a specialized measure intended for situations where interest is focused on what happens below some threshold, such as a poverty line. It is not suitable as a general measure of approximate stochastic dominance.[2]

Another measure of partial stochastic dominance is the notion of “almost stochastic dominance” from the finance literature (see, e.g., Leshno and Levy (2002); Levy (2009)). This measure also compares cdfs, and is described in detail below. Several

[2]One issue is that in general the appropriate choice of c might not be clear. A second issue is illustrated by the following example. Let $F = N(0, 0.1)$ and $G = N(10, 1)$. The distribution G ”approximately” dominates F in the sense that the probability that a draw from F exceeds a draw from G is essentially indistinguishable from zero. At the same time, condition (1) fails for any $c \in \mathbb{R}$.

other measures are considered in Stoyanov et al. (2012), who analyze degrees of violation of stochastic dominance using quasi-semidistances.

The aim of this paper is to make the case that there is a single measure of extent of stochastic dominance, denoted below by σ, that can arguably be considered as the default measure from the perspective of economic analysis (without claiming to be the most suitable in every individual setting). For one dimensional distributions with cdfs F and G, the measure can be expressed by the simple representation

$$\sigma(F, G) = 1 - \sup_x \{G(x) - F(x)\}. \tag{2}$$

This value measures the extent to which F is dominated by G. It is not difficult to verify that $0 \leq \sigma(F, G) \leq 1$ and $\sigma(F, G) = 1$ if and only if F is stochastic dominated by G.

While the measure $\sigma(F, G)$ seems almost naïve, we show that it has a number of different and useful representations, can be motivated from several entirely different perspectives, possesses attractive continuity properties, and extends naturally to multidimensional or more abstract settings. As an example of an alternative representation presented in the paper, we show that $\sigma(F, G)$ can also be expressed as

$$\sigma(F, G) = \max_{X,Y} \mathbb{P}\{X \leq Y\}$$

where the supremum is over all pairs of random variables (X, Y) such that $X \sim F$ and $Y \sim G$. One way to think of this expression is as providing the "maximal regret probability" for a decision maker with monotone preferences who chooses F over G. Other interpretations and alternative expressions are discussed below.

Applying the measure σ to the distributions in figure 1, we calculate that $\sigma(\mu, \nu) = 1$ as expected (since μ is dominated by ν). Comparing μ and ν', it can be shown analytically that $\sigma(\mu, \nu')$ is strictly less than one. However, from the perspective of 64 bit floating point arithmetic, $\sigma(\mu, \nu')$ is indistinguishable from 1.

The main part of our paper proceeds as follows. First we present several alternative measures of extent of stochastic dominance, all motivated from different perspectives, and then show that all are equivalent to σ (sections 2 and 3). Next we show that σ has desirable continuity and robustness properties (section 4). All of these results are stated and proved in a general setting, for probability measures on an arbitrary space with arbitrary closed partial order. Hence our analysis accommodates a variety of different state spaces and state space orderings, including multivariate and vector-valued state spaces with standard partial orders, infinite dimensional state spaces (e.g., sets of distributions) and so on. Section 5 discusses two applications. Proofs are given in section 7.

2. A Measure of Stochastic Dominance

We begin with some preliminaries and then define a measure of degree of stochastic dominance.

2.1. Definitions.

Let S be a Polish space with closed partial order $\preceq$ and Borel sets $\mathscr{B}$.[3] As usual, we say that $h\colon S \to \mathbb{R}$ is an **increasing function** if $x \preceq y$ implies $h(x) \leq h(y)$. We say that h is **bounded** if there exists a $K \in \mathbb{R}$ with $|h(x)| \leq K$ for all $x \in S$. We say that a set $B \subset S$ is an **increasing set** if $x \in B$ and $x \preceq y$ implies $y \in B$. Equivalently, B is an increasing set if its indicator function $\mathbb{1}_B$ is an increasing function. We write

- $\mathscr{M}$ for the finite measures[4] on $(S, \mathscr{B})$
- $\mathscr{P} \subset \mathscr{M}$ for the probability measures
- $is\mathscr{B}$ for the increasing sets in $\mathscr{B}$
- $ib\mathscr{B}$ for the increasing, bounded, Borel measurable functions from S to $\mathbb{R}$.

[3] A **Polish space** is a completely metrizable topological space. A **closed partial order** on S is a binary relation $\preceq$ on $S \times S$ that is reflexive, transitive and antisymmetric, and such that its graph $\{(x, y) \in S \times S : x \preceq y\}$ is closed in the product topology.

[4] A **finite measure** on $(S, \mathscr{B})$ is a countably additive function $\mu\colon \mathscr{B} \to \mathbb{R}_+$ satisfying $\mu(S) < \infty$.

Two natural partial orders on $\mathscr{M}$ are defined as follows: Given μ and ν in $\mathscr{M}$, we say that μ is **stochastically dominated** by ν and write $\mu \preceq_{sd} \nu$ if

$$\mu(S) = \nu(S) \quad \text{and} \quad \int h \, d\mu \leq \int h \, d\nu \quad \forall \, h \in ib\mathscr{B}.$$

In our setting, the condition $\int h \, d\mu \leq \int h \, d\nu$ for all $h \in ib\mathscr{B}$ can be replaced by $\mu(I) \leq \nu(I)$ for all $I \in is\mathscr{B}$, or even $\mu(I) \leq \nu(I)$ for all closed $I \in is\mathscr{B}$ (Kamae et al., 1977, theorem 1). We say that μ is **setwise dominated** by ν and write $\mu \leq \nu$ if

$$\mu(B) \leq \nu(B) \quad \forall \, B \in \mathscr{B}.$$

For each $\mu, \nu \in \mathscr{P}$, we define

$$\Phi(\mu, \nu) := \{(\mu', \nu') \in \mathscr{M} \times \mathscr{M} : \mu' \leq \mu, \ \nu' \leq \nu, \ \mu' \preceq_{sd} \nu'\}.$$

If μ and μ' are measures with $\mu' \leq \mu$, then μ' is sometimes called a **component** of μ. In this terminology, we can think of $\Phi(\mu, \nu)$ as the set of "ordered component pairs" corresponding to (μ, ν), where "ordered" means ordered by stochastic dominance. Note that $(\mu', \nu') \in \Phi(\mu, \nu)$ implies that $\mu'(S) = \nu'(S)$.

It is clear that if (μ', ν') is any ordered component pair, then the total mass $\mu'(S) = \nu'(S)$ must be no more than one; and that if μ is stochastically dominated by ν, then the original pair (μ, ν) is itself an ordered component pair, with each element of the pair having total mass that attains the maximum of unity. It is also intuitive that if μ is "almost" dominated by ν, then we can choose relatively large components, so that the mass of each element of the ordered component pair approaches one. These ideas suggest regarding the mass of the largest ordered component pair as a measure of the extent to which μ is dominated by ν. Thus, for each μ, ν in $\mathscr{P}$, we define $\sigma(\mu, \nu)$ by

$$\sigma(\mu, \nu) := \sup\{\mu'(S) : (\mu', \nu') \in \Phi(\mu, \nu)\}, \tag{3}$$

and regard it as a measure of degree of first order stochastic domination of μ by ν. Its basic properties are as follows:

Proposition 2.1. *For any $\mu, \nu \in \mathscr{P}$,*

1. *$\sigma(\mu, \nu)$ is well defined and contained in* $[0,1]$.
2. *$\mu \preceq_{sd} \nu$ if and only if $\sigma(\mu, \nu) = 1$* .
3. *$\sigma(\mu, \nu) = \sigma(\nu, \mu) = 1$ if and only if $\mu = \nu$.*

In particular, $0 \leq \sigma(\mu, \nu) \leq 1$, with higher values representing ”greater” domination of μ by ν, and with $\sigma(\mu, \nu) = 1$ attained if and only if stochastic dominance is complete.

Proposition 2.2. *For any $\mu, \nu \in \mathscr{P}$, there exists a pair $(\mu', \nu') \in \Phi(\mu, \nu)$ such that*

$$\mu'(S) = \nu'(S) = \sigma(\mu, \nu).$$

In other words, the sup in (3) can be replaced by max. Like most of our results, the proofs of propositions 2.1 and 2.2 are deferred to section 7.

Two particularly important properties of σ are as follows. First, this measure of extent of stochastic dominance turns out to be equivalent to several other natural measures of the same quantity, each of which can be motivated from considerations related to choice under uncertainty. Second, σ exhibits “continuity” or “robustness,” in the sense that, when μ is “almost” dominated by ν, the value of $\sigma(\mu, \nu)$ is almost 1. These topics are treated below, beginning with equivalences.

3. Equivalences

In this section we introduce several other measures of extent of stochastic dominance, all of which have some claim to be natural measures of this concept. We then show that each of these measures is equivalent to σ.

3.1. **A Coupling Measure.**

For agents with monotone preferences, stochastic dominance ranks distributions. Let us now also consider choosing between arbitrary distributions that may or may not be ordered. We will consider a setting where preferences are monotone but make no assumptions beyond that. For example, we can consider a decision maker evaluating distributions on behalf of agents whose preferences are known only to be monotone (consider, for example, a fund manager or a policy maker).

Given that the agents have monotone preferences, one way for the decision maker to think about distribution μ being less desirable than distribution ν is to consider the probability that choosing prospect μ turns out to be worse than ν *ex post*—that is, the probability that a draw X from μ is dominated by a draw Y from ν. This probability depends on the joint distribution of X and Y. In the current setting we have no information on the joint distribution. Absent such information, one approach is to make the best decision under the worst case scenario, which is the joint distribution making $X \preceq Y$ most likely. This leads us to the value $\sup_{X,Y} \mathbb{P}\{X \preceq Y\}$, where the supremum is over all pairs (X, Y) with marginals μ and ν. The larger this value, the more μ can be regarded as ”dominated” by ν.

To state this idea more formally, recall that, given $(\mu, \nu) \in \mathscr{P} \times \mathscr{P}$, a pair of S-valued random variables (X, Y) defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called a **coupling** of (μ, ν) if $\mu(B) = \mathbb{P}\{X \in B\}$ and $\nu(B) = \mathbb{P}\{Y \in B\}$ for all $B \in \mathscr{B}$. In other words, X and Y live on the same probability space and have marginals μ and ν. The set of all couplings of $(\mu, \nu) \in \mathscr{P} \times \mathscr{P}$ will be denoted by $\mathscr{C}(\mu, \nu)$. Given this notation, the reasoning in the preceding paragraph suggests

$$s(\mu, \nu) := \sup_{(X,Y) \in \mathscr{C}(\mu,\nu)} \mathbb{P}\{X \preceq Y\}. \tag{4}$$

as a measure of the extent to which μ is stochastically dominated by ν.

To help understand this idea better, let's consider the case where $\mu \preceq_{sd} \nu$. For such a pair we can draw on a well known result due to Strassen (1965), which, in the current setting, tells us that

$$\mu \preceq_{sd} \nu \quad \iff \quad \exists\, (X, Y) \in \mathscr{C}(\mu, \nu) \text{ s.t. } \mathbb{P}\{X \preceq Y\} = 1. \tag{5}$$

It follows that the measure $s(\mu, \nu)$ defined in (4) attains its maximum (i.e., equals one) when μ is stochastically dominated by ν. In fact, as we will see, the two statements are equivalent, in the sense that if μ is not stochastically dominated by

ν, then $s(\mu, \nu)$ will be strictly less than one. Moreover, as will also be shown below, when μ is "almost" dominated by ν, the value will be close to one.

Before going further, it is worth noting that we can replace sup in (4) with max because, somewhat surprisingly, the maximum always exists:

Proposition 3.1. *For all $(\mu, \nu) \in \mathscr{P} \times \mathscr{P}$, there exists a pair $(X, Y) \in \mathscr{C}(\mu, \nu)$ such that*

$$\mathbb{P}\{X \preceq Y\} = \sup_{(X', Y') \in \mathscr{C}(\mu, \nu)} \mathbb{P}\{X' \preceq Y'\}. \tag{6}$$

While it might at first seem that s defined in (4) is a new and entirely different measure of degree of stochastic dominance, we show below that it coincides with σ.

3.2. **Deviation Measures.** There is another class of functions on $\mathscr{P} \times \mathscr{P}$ that have a strong claim to providing a measure of extent of stochastic dominance. To construct them, let μ and ν be any two probability measures, and observe that, by definition, $\mu \preceq_{sd} \nu$ if and only if

$$\sup_{h \in ib\mathscr{B}} \left\{ \int h \, d\mu - \int h \, d\nu \right\} \leq 0. \tag{7}$$

From a social welfare or choice-under-uncertainty perspective, this inequality states that every agent with monotone preferences will achieve at least the same expected utility under ν as under μ.

Note that zero can always be achieved in (7) by choosing $h = 0$ or $h = 1$. Moreover, if μ is not completely dominated by ν, then a strictly positive value can be obtained by the supremum. This value gives some indication of the extent to which μ *fails to be dominated* by ν. Hence we can consider

$$\theta(\mu, \nu) := \sup_{h \in H} \left\{ \int h \, d\mu - \int h \, d\nu \right\} \quad \text{where} \quad H := \{h \in ib\mathscr{B} : 0 \leq h \leq 1\} \tag{8}$$

as an *inverse* measure of the degree to which μ is dominated by ν, taking smaller values when domination by ν is greater, and zero when domination is complete. A measure similar in spirit to θ is mentioned in Stoyanov et al. (2012), with the only difference being the class of functions over which the supremum is taken.

In the definition of θ, the restriction of $ib\mathscr{B}$ to functions with range in $[0,1]$ is a convenient normalization. In fact we can go a step further and restrict to indicator functions, which leads to the alternative measure

$$\theta_I(\mu,\nu) := \sup_{B \in is\mathscr{B}} \{\mu(B) - \nu(B)\}.$$

3.3. **Equivalence.** It turns out that all of the preceding measures are equivalent in the sense made precise by the following theorem:

Theorem 3.1. *For any* $(\mu,\nu) \in \mathscr{P} \times \mathscr{P}$ *we have*

$$\sigma(\mu,\nu) = s(\mu,\nu) = 1 - \theta(\mu,\nu) = 1 - \theta_I(\mu,\nu).$$

As with other results, the proof of theorem 3.1 can be found in section 7.

For some calculations it can be convenient to consider the variation

$$(9) \qquad \theta_I^{\mathscr{O}}(\mu,\nu) := \sup_{B \in i\mathscr{O}} \{\mu(B) - \nu(B)\} \quad \text{where} \quad i\mathscr{O} := \{\text{all open } B \in is\mathscr{B}\}.$$

Evidently $\theta_I^{\mathscr{O}} \leq \theta_I$. In fact, as the proof of theorem 3.1 shows, we have $\theta_I^{\mathscr{O}} = \theta_I$ on all of $\mathscr{P} \times \mathscr{P}$, and hence $\theta_I^{\mathscr{O}}(\mu,\nu) = 1 - \sigma(\mu,\nu)$ is also valid. Moreover, it is trivial to confirm that

$$\sup_{B \in i\mathscr{O}} \{\mu(B) - \nu(B)\} = \sup_{D \in d\mathscr{C}} \{\nu(D) - \mu(D)\},$$

where $d\mathscr{C}$ is the closed decreasing sets in $\mathscr{B}$. This is useful, particularly in the one-dimensional case, where $d\mathscr{C}$ is the set of all intervals $(-\infty, x]$ over $x \in \mathbb{R}$. In this setting it implies that for cdfs F and G on $\mathbb{R}$, we have $\sigma(F,G) = 1 - \sup_x\{G(x) - F(x)\}$. This is the expression for σ given in the introduction.

Remark 3.1. One benefit of the last expression is that the measure σ is straightforward to compute in one dimension. Aside from numerical optimization, if G and F

are both absolutely continuous, then the first order condition for $\max_{x \in \mathbb{R}}\{G(x) - F(x)\}$ is just $f(x) = g(x)$, where f and g are the densities of F and G respectively.

Remark 3.2. For other calculations, alternative expressions for σ can be easier to work with. To illustrate, let S be a general space—not necessarily one dimensional—and suppose that μ is degenerate, putting all mass on $\bar{x} \in S$. Let ν be arbitrary. If we wish to compute $\sigma(\mu, \nu)$, it is enough to observe that if (X, Y) is any coupling of (μ, ν), then, since X puts all mass on $\bar{x}$, we must have $\mathbb{P}\{X \preceq Y\} = \mathbb{P}\{\bar{x} \preceq Y\}$. We conclude that $\sigma(\mu, \nu) = \nu(I_{\bar{x}})$ where $I_{\bar{x}} := \{x \in S : \bar{x} \preceq x\}$.

4. Continuity

Since we wish to use $\sigma(\mu, \nu)$ as a measure of the extent to which μ is dominated by ν, one desirable characteristic is that $\sigma(\mu, \nu)$ is "almost" 1 whenever μ is "almost" dominated by ν. This turns out to be valid in many cases, although some care is required when defining "almost." To state a precise result, let's write $\mu_n \xrightarrow{u} \mu$ to mean that $\|\mu_n - \mu\| \to 0$ as $n \to \infty$, where $\|\cdot\|$ is the total variation norm

$$\|\mu_n - \mu\| := \sup_{B \in \mathscr{B}} |\mu_n(B) - \mu(B)|.$$

Now, given sequences of probability measures $\{\mu_n\}$ and $\{\nu_n\}$, the following result holds:

Proposition 4.1. *If $\mu_n \xrightarrow{u} \mu$ and $\nu_n \xrightarrow{u} \nu$, then $\sigma(\mu_n, \nu_n) \to \sigma(\mu, \nu)$.*

Put differently, $\sigma\colon \mathscr{P} \times \mathscr{P} \to [0,1]$ is continuous in the product topology induced by the total vairation norm. In fact the proof of proposition 4.1 gives the bound

$$|\sigma(\mu_n, \nu_n) - \sigma(\mu, \nu)| \leq \|\mu_n - \mu\| + \|\nu_n - \nu\|.$$

Specializing this bound to the case where $\mu \preceq_{sd} \nu$ and $\nu_n = \nu$, we get $\sigma(\mu_n, \nu) \geq 1 - \|\mu_n - \mu\|$. This quantifies the idea that $\sigma(\mu_n, \nu)$ is almost 1 when μ_n almost equal to a distribution dominated by ν.

The total variation convergence used in proposition 4.1 can be checked relatively easily when the probability measures in question have densities, since pointwise convergence of densities implies total variation convergence by Scheffé's lemma (Williams, 1991, p. 55). Also, in view of Pinsker's inequality, proposition 4.1 remains valid when comparison is made based on measures of relative entropy. For example, if $\mu \preceq_{sd} \nu$ and either the Kullback-Leibler divergence or the Jensen-Shannon distance from μ_n to μ and ν_n to ν goes to zero, then $\sigma(\mu_n, \nu_n) \to \sigma(\mu, \nu)$. See Devroye and Lugosi (2001).

By comparison, consider the concept of "almost stochastic dominance" mentioned in the introduction. Let S be a bounded interval of $\mathbb{R}$ and let F and G be cdfs on S. Let $\epsilon \geq 0$. Leshno and Levy (2002) define F to be **ϵ-almost stochastically dominated** by G if

$$\int_{\{F<G\}} (G(x) - F(x))dx \leq \epsilon \int |G(x) - F(x)|dx. \tag{10}$$

Here $\{F < G\} := \{x \in S : F(x) < G(x)\}$. We can convert this to a measure of extent of stochastic dominance by considering the smallest ϵ such that (10) is valid. That is,

$$a(F, G) = \inf \left\{ \epsilon \geq 0 \, : \, \int_{\{F<G\}} (G(x) - F(x))dx \leq \epsilon \int |G(x) - F(x)|dx \right\}.$$

Smaller values of $a(F, G)$ are associated with a greater degree of dominance of G over F. Moreover $a(F, G) = 0$ if and only if $F \preceq_{sd} G$.

Unfortunately, the measure a fails to have the form of continuity discussed in proposition 4.1. For example, let $F = G = G_n = N(0, 1)$, the standard normal distribution, and let $F_n = N(1/n, 1)$. Here $G_n \xrightarrow{u} G$ is trivial and $F_n \xrightarrow{u} F$ is valid because the densities converge pointwise everywhere on $\mathbb{R}$ (which is sufficient for total variation convergence as discussed above). At the same since $F_n(x) < G_n(x)$ for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$, we have $a(F_n, G_n) = 1$ for all n. This compares to $a(F, G) = 0$ (because $F = G$).

Note that the claim in proposition 4.1 may fail if we replace total variation convergence with weak convergence. (Recall that $\mu_n \to \mu$ weakly if $\int g d\mu_n \to \int g d\mu$ for all continuous bounded $g\colon S \to \mathbb{R}$. Symbolically we write $\mu_n \stackrel{w}{\to} \mu$.) The reason the proposition is not generally valid in this case is that weak convergence is a topological notion, while the stochastic dominance measure σ is defined only by the ordering on S.

Note however that total variation convergence can be weakened while still maintaining the conclusions of proposition 4.1. For example, in the one-dimensional setting, it is sufficient that convergence takes place in the Kolmogorov metric. That is, if F, G, F_n and G_n are cdfs on $\mathbb{R}$ with

$$\lim_{n\to\infty} \sup_x |F_n(x) - F(x)| = \lim_{n\to\infty} \sup_x |G_n(x) - G(x)| = 0,$$

then $\sigma(F_n, G_n) \to \sigma(F, G)$. A sufficient condition for convergence of cdfs $\{H_n\}$ to H in the Kolmogorov metric is that $H_n \to H$ weakly and H is continuous (see, e.g., Dudley (2002), p. 389). This is obviously the case when H has a density. This leads to the following corollary:

Corollary 4.1. *Let F_n, G_n, F and G be cdfs on $\mathbb{R}$. The following statements are true:*

1. *If $F_n \stackrel{w}{\to} F$ and F is continuous, then $\sigma(F_n, G) \to \sigma(F, G)$.*
2. *If $G_n \stackrel{w}{\to} G$ and G is continuous, then $\sigma(F, G_n) \to \sigma(F, G)$.*
3. *If $F_n \stackrel{w}{\to} F$, $G_n \stackrel{w}{\to} G$ and both F and G are continuous, then $\sigma(F_n, G_n) \to \sigma(F, G)$.*

In other words, weak convergence is enough when the limiting distributions are sufficiently regular.

5. Applications

The concepts discussed in this paper can be applied to any setting where distributions are compared, from industrial organization (e.g., distributions of firm size or productivity) to labor economics (wage distributions), development and growth (wealth distributions), macroeconomics (parameter uncertainty represented via distributions), network analysis (degree distributions), social welfare (income distributions) and so on. Here we consider a two problems, one related to wealth distribution and the other to portfolio choice. Our discussion illustrates the point that some properties only hold "in the limit." The measure σ allows us to make this idea precise in the context of stochastic dominance.

5.1. **Wealth Distributions.** Let W_0 be an initial wealth or asset distribution across some set of agents, and suppose that individual wealth grows according to some specified dynamics. Interesting examples can be found in Piketty (1997), Benhabib et al. (2011) and many other places. Since the main points change little, here we simply suppose that wealth grows geometrically and deterministically via $W_{t+1} = aW_t$ for some $a \geq 1$. As a result, $W_t = a^t W_0$, or, with lower case representing logs, $w_t = t \ln a + w_0$.

Now consider an alternative economy with wealth distribution $w_0' = \ln W_0'$ and growth rate $a' > a$. Despite this higher growth rate, one cannot in general claim that $w_t' = t \ln a' + w_0'$ will stochastically dominate w_t for any finite t. An obvious counterexample is as follows: Suppose that W_0 and W_0' are lognormal, so that w_t and w_t' are normal. If w_0 and w_0' do not have identical variance, then the same will be true of w_t and w_t', and hence they will fail to be ordered for any finite t.

On the other hand, it seems intiutive that w_t' will "almost" stochastically dominate w_t when t is large enough. We can formalize this using σ. In particular, we can show that $\sigma(F_t, F_t') \to 1$ as $t \to \infty$, where F_t and F_t' are the distributions of w_t and w_t' respectively. To do this, observe that, by (4) and theorem 3.1, the statement $\sigma(F_t, F_t') \geq \mathbb{P}\{w_t \leq w_t'\}$ is always valid. Hence

$$\sigma(F_t, F_t') \geq \mathbb{P}\{t \ln a + w_0 \leq t \ln a' + w_0'\} = \mathbb{P}\{w_0 - w_0' \leq t(\ln a' - \ln a)\}.$$

Letting G be the cdf of $w_0 - w_0'$ and $\kappa := \ln a' - \ln a$, we can express this as $\sigma(F_t, F_t') \geq G(t\kappa)$ for all t. Since G is a cdf and $\kappa > 0$, we have shown that $\sigma(F_t, F_t') \to 1$ as claimed.

5.2. **A Portfolio Problem.** Next we turn to the portfolio problem. To begin, consider a set of securities indexed by $j \in \mathbb{N}$. Let $x_j = g + \epsilon_j$ be the payoff of the j-th security. Here g is a common factor with distribution function G, and $\{\epsilon_j\}_{j \in \mathbb{N}}$ is a sequence of uncorrelated shocks with zero mean and finite variance σ^2. A portfolio is a sequence $\theta = \{\theta_1, \theta_2 \ldots\}$ such that $\sum_{j=1}^{\infty} \theta_j = 1$. Each portfolio is identified with a payoff $\sum_{j=1}^{\infty} \theta_j x_j$. A variance minimizing finite diversified portfolio is a portfolio that puts weight $1/n$ on n securities. More formally, it is a portfolio $\{\theta_j\}$ of the form $\theta_j = n^{-1} \mathbb{1}\{j \in I_n\}$ where I_n is a finite set containing n elements of $\mathbb{N}$. Since it changes nothing substantial, we always take $I_n := \{1, \ldots, n\}$. Let θ_n^* denote this portfolio. The payoff of θ_n^* is $g + \frac{1}{n} \sum_{j=1}^{n} \epsilon_j$.

The diversified portfolio θ_n^* uses the law of large numbers to reduce idiosyncratic risk. Each increase in n decreases the variance of the payoff. In the limit only factor risk remains. However, the limit of the sequence of diversified portfolios is not trivial to describe, since the share $\theta_j = 1/n$ converges to zero, while the constraint $\sum_j \theta_j = 1$ must be respected. Werner (1997) places the limiting portfolio on a firm theoretical foundation by representing it as a density charge (i.e., a purely finitely additive measure) on the power set of $\mathbb{N}$. Here we investigate limiting behavior from a different angle, at least as it pertains to stochastic dominance.

Suppose we are interested in comparing the diversified portfolio against some alternative portfolio (e.g., a portfolio containing only a risk free asset). For the sake of generality we will consider an arbitrary alternative, with payoff distribution F. Let G_n be the distribution function for the payoff $g + \frac{1}{n} \sum_{j=1}^{n} \epsilon_j$ of θ_n^*. Regarding G_n and F we have the following result:

Proposition 5.1. *If G is continuous, then $\sigma(F, G_n) \to \sigma(F, G)$.*

The proof is a simple consequence of the results in section 4. A useful implication of proposition 5.1 is that if the distribution G of the factor g stochastically dominates F, then $\sigma(F, G_n) \to 1$. In other words, the payoff of the diversified portfolio G_n "stochastically dominates F in the limit," as continued diversification reduces idiosyncratic risk towards zero. This is true even if F fails to be dominated by G_n for all n, as would be the case if the shocks ϵ_j have sufficiently large support.[5]

6. Conclusion

This paper shows that there is a single measure of extent of stochastic dominance that can arguably be treated as the default measure of this property from the perspective of economic analysis. The main results of the paper demonstrate that this measure has a number of different and informative representations that are in fact equivalent. As a consequence, the measure can be motivated from several very different perspectives. In addition, the measure is shown to possess important continuity properties, and extends naturally to multidimensional or more abstract settings. Two applications are given.

As a caveat we reiterate that, as with many problems, the existence of a default measure does not preclude the optimality of other measures in certain specific settings. On the other hand, it does provide a common frame of reference for theoretical work. Moreover, since the measure can be estimated, the preceding analysis also opens up new avenues for econometric investigation of stochastic dominance. Finally, the question of whether analogous measures and results exist for the case of higher order stochastic dominance (second order, third order, etc.) is left to future research.

[5]To see why continuity of G matters, suppose that F and G are both degenerate distributions concentrated on some risk free rate R. Let ϵ_j be standard normal, so that $G_n = N(R, 1/n)$. Then $\sigma(F, G) = 1$ but $\sigma(F, G_n) = 1/2$ for all n (see remark 3.2 on how to obtain the last equality). The intuitive reason for failure of "stochastic dominance in the limit" here is as follows: Suppose we have an agent with preferences defined by von Neumann-Morgenstern utility function $u(x) = \mathbb{1}\{R \leq x\}$. This agent cares only about equaling or beating the risk free rate. Despite monotone increasing preferences, the expected payoff of G_n for this agent is always half of that of F, regardless of how large n is.

7. Proofs

7.1. **Preliminaries.** Throughout the proofs we let $\mathscr{O}$ and $\mathscr{C}$ be the open and closed subsets of S respectively. In addition, let $d\mathscr{C}$ be the decreasing sets in $\mathscr{C}$, and let $i\mathscr{O}$ be the increasing sets in $\mathscr{O}$. As in (8), let $H := \{h \in ib\mathscr{B} : 0 \leq h \leq 1\}$. For our partial order $\preceq$ on S, let $\mathbb{G}$ be the graph. That is

$$\mathbb{G} := \{(x, y) \in S \times S : x \preceq y\}.$$

Let $S \times S$ be the Cartesian product of S with itself, and let $\mathscr{B} \otimes \mathscr{B}$ be the product σ-algebra. Let π_i be the i-th coordinate projection, so that $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$ for any $(x, y) \in S \times S$. As usual, given $Q \subset S \times S$, we let $\pi_1(G)$ be all $x \in S$ such that $(x, y) \in Q$, and similarly for π_2. A probability measure λ on $(S \times S, \mathscr{B} \otimes \mathscr{B})$ is said to **have marginals** μ and ν if

$$\mu(B) = \lambda(B \times S) \quad \text{and} \quad \nu(B) = \lambda(S \times B) \quad \forall\, B \in \mathscr{B}.$$

Equivalently, $\mu = \lambda \circ \pi_1^{-1}$ and $\nu = \lambda \circ \pi_2^{-1}$. We will apply the following deep theorem due to Strassen (1965), slightly specialized to the current setting:

Theorem 7.1 (Strassen's theorem)**.** *Let $\mu, \nu \in \mathscr{P}$. For any $\epsilon \geq 0$ and any closed set $K \subset S \times S$, there exists a probability measure λ on $(S \times S, \mathscr{B} \otimes \mathscr{B})$ with marginals μ and ν such that $\lambda(K) \geq 1 - \epsilon$ if and only if*

$$\nu(F) \leq \mu(\pi_1(K \cap (S \times F))) + \epsilon, \qquad \forall\, F \in \mathscr{C}.$$

By applying Strassen's theorem and some other more elementary results, we will now show that the following chain of inequalities holds:

$$\sup_{h\in H}\left\{\int h\,d\mu - \int h\,d\nu\right\} \le 1-\sigma(\mu,\nu) \tag{11}$$

$$\le 1 - \sup_{(X,Y)\in\mathscr{C}(\mu,\nu)} \mathbb{P}\{X \preceq Y\} \tag{12}$$

$$\le \sup_{I\in i\mathscr{O}}\{\mu(I)-\nu(I)\} \tag{13}$$

$$\le \sup_{h\in H}\left\{\int h\,d\mu - \int h\,d\nu\right\}. \tag{14}$$

Fix $\mu, \nu \in \mathscr{P}$. Regarding inequality (11), fix $h \in H$. Let (μ', ν') be an element of $\Phi(\mu,\nu)$ with $\mu'(S) = \nu'(S) = \sigma(\mu,\nu)$, the existence of which follows from proposition 2.2. Let $\mu^r := \mu - \mu'$ and $\nu^r := \nu - \nu'$. We have

$$\begin{aligned}
\int hd\mu - \int hd\nu &= \int hd\mu^r + \int hd\mu' - \int hd\nu^r - \int hd\nu' \\
&= \int hd\mu^r - \int hd\nu^r + \left\{\int hd\mu' - \int hd\nu'\right\} \\
&\le \int hd\mu^r - \int hd\nu^r \\
&\le \int hd\mu^r \le \mu^r(S) = 1-\sigma(\mu,\nu).
\end{aligned}$$

Next we prove inequality (12). To prove this inequality it is enough to show that if (X,Y) is a coupling of (μ,ν), then $\mathbb{P}\{X \preceq Y\} \le \sigma(X,Y)$. To see that this is so, let (X,Y) be a coupling of (μ,ν), and define

$$\mu'(B) := \mathbb{P}\{X \in B,\, X \preceq Y\} \quad \text{and} \quad \nu'(B) := \mathbb{P}\{Y \in B,\, X \preceq Y\}.$$

Clearly $\mu' \le \mu$, $\nu' \le \nu$ and $\mu'(S) = \mathbb{P}\{X \preceq Y\} = \nu'(S)$. Moreover, for any increasing set $I \in \mathscr{B}$ we have

$$\mu'(I) = \mathbb{P}\{X \in I,\, X \preceq Y\} \le \mathbb{P}\{Y \in I,\, X \preceq Y\} = \nu'(I).$$

Hence $(\mu',\nu') \in \Phi(\mu,\nu)$ and we obtain $\mathbb{P}\{X \preceq Y\} = \mu'(S) \le \sigma(\mu,\nu)$.

We now turn to inequality (13). For any $F \subset S$, let F^d be the smallest decreasing set containing F. In other words, F^d is all $x \in S$ such that $x \preceq y$ for some $y \in F$. Note that if F is closed then, since $\preceq$ is a closed partial order, so is the set F^d (i.e., $F \in \mathscr{C} \implies F^d \in d\mathscr{C}$). Let

$$\epsilon := \sup_{D \in d\mathscr{C}} \{\nu(D) - \mu(D)\}.$$

Evidently

$$\epsilon \geq \sup_{F \in \mathscr{C}} \{\nu(F^d) - \mu(F^d)\} \geq \sup_{F \in \mathscr{C}} \{\nu(F) - \mu(F^d)\}.$$

Noting that F^d can be expressed as $\pi_1(\mathbb{G} \cap (S \times F))$, it follows that, for any $F \in \mathscr{C}$,

$$\nu(F) \leq \mu(\pi_1(\mathbb{G} \cap (S \times F))) + \epsilon.$$

Since $\preceq$ is a closed partial order, the set $\mathbb{G}$ is closed, and Strassen's theorem (theorem 7.1) applies. From this theorem we obtain a probability measure λ on the product space $S \times S$ such that (a) $\lambda(\mathbb{G}) \geq 1 - \epsilon$ and (b) λ has marginals μ and ν.

Because complements of increasing sets are decreasing and vice versa, we have

$$\sup_{I \in i\mathscr{O}} \{\mu(I) - \nu(I)\} = \sup_{D \in d\mathscr{C}} \{\nu(D) - \mu(D)\} = \epsilon \geq 1 - \lambda(\mathbb{G}). \tag{15}$$

Now consider the probability space $(\Omega, \mathscr{F}, \mathbb{P}) = (S \times S, \mathscr{B} \otimes \mathscr{B}, \lambda)$, and let $X = \pi_1$ and $Y = \pi_2$. We then have

$$\lambda(\mathbb{G}) = \lambda\{(x, y) \in S \times S : x \preceq y\} = \mathbb{P}\{X \preceq Y\}.$$

Combining this equality with (15), we have shown the existence of a $(X, Y) \in \mathscr{C}(\mu, \nu)$ with

$$\sup_{I \in i\mathscr{O}} \{\mu(I) - \nu(I)\} \geq 1 - \mathbb{P}\{X \preceq Y\}.$$

Inequality (13) now follows.

Since inequality (14) is trivial we have now established the chain (11)–(14). It follows that all the terms in the chain are equal.

7.2. **Remaining Proofs.** We now turn to the remaining proofs from the paper. We prove propostion 2.2 first, since it is independent of other claims and will be used below.

Proof of propostion 2.2. Let $\{\tilde{\mu}_i\}_{i\in\mathbb{N}}$ and $\{\tilde{\nu}_i\}_{i\in\mathbb{N}}$ be sequences in $\mathscr{M}$ such that

$$\tilde{\mu}_i(S) \uparrow \sigma(\mu,\nu), \quad \forall i \in \mathbb{N},\ (\tilde{\mu}_i, \tilde{\nu}_i) \in \Phi(\mu,\nu). \tag{16}$$

Since $\tilde{\mu}_i \leq \mu$ and $\tilde{\nu}_i \leq \nu$ for all $i \in \mathbb{N}$, each of these sequences has a weakly convergent subsequence by Prohorov's theorem (Dudley, 2002, theorem 11.5.4). By taking further subsequences, we can assume that $\tilde{\mu}_i \to \tilde{\mu}$ and $\tilde{\nu}_i \to \tilde{\nu}$ weakly for some $\tilde{\mu}, \tilde{\nu} \in \mathscr{M}$. Since $\tilde{\mu}_i \leq \mu$ and $\tilde{\nu}_i \leq \nu$ for all $i \in \mathbb{N}$, it follows from Hernández-Lerma and Lasserre (2003), theorem 1.5.5, that for all $B \in \mathscr{B}$, $\tilde{\mu}_i(B) \to \tilde{\mu}(B)$ and $\tilde{\nu}_i(B) \to \hat{\nu}(B)$. Hence, for any $B \in \mathscr{B}$, we have $\tilde{\mu}(B) \leq \mu(B)$ and $\tilde{\nu}(B) \leq \nu(B)$ since $\tilde{\mu}_i(B) \leq \mu(B)$ and $\tilde{\nu}_i(B) \leq \nu(B)$ for all $i \in \mathbb{N}$. Thus $\tilde{\mu} \leq \mu$ and $\tilde{\nu} \leq \nu$. Since $\tilde{\mu}_i(S) = \tilde{\nu}_i(S)$ for all $i \in \mathbb{N}$, we also have $\tilde{\mu}(S) = \tilde{\nu}(S)$. Moreover, $\sigma(\mu,\nu) = \tilde{\mu}(S)$, since $\tilde{\mu}_i(S) \uparrow \sigma(\mu,\nu)$. For any increasing $I \in \mathscr{B}$, the fact that $\tilde{\mu}_i(I) \leq \tilde{\nu}_i(I)$ for all $i \in \mathbb{N}$ gives us $\tilde{\mu}(I) \leq \tilde{\nu}(I)$. Thus $\tilde{\mu} \preceq_{sd} \tilde{\nu}$. We have shown that the pair $(\tilde{\mu}, \tilde{\nu})$ satisfies the claim in proposition 2.2. □

Now we can return to the proof of proposition 2.1.

Proof of proposition 2.1. The first claim is trivial. Regarding the second, it is clear that if $\mu \preceq_{sd} \nu$, then $(\mu,\nu) \in \Phi(\mu,\nu)$, and hence $\sigma(\mu,\nu) = 1$. To see the converse, suppose that $\sigma(\mu,\nu) = 1$. Then, in view of proposition 2.2, there exists an ordered component pair $(\mu',\nu') \in \Phi(\mu,\nu)$ with $\mu'(S) = \nu'(S) = 1$. Since $\mu' \leq \mu$ and $\mu'(S) = 1$, it must be that $\mu' = \mu$. (Otherwise, there exists a $B \in \mathscr{B}$ with $\mu'(B) < \mu(B)$, from which we obtain $\mu'(B^c) > \mu(B^c)$. This contradicts $\mu' \leq \mu$.) For the same reason we have, $\nu' = \nu$. Because $\mu' \preceq_{sd} \nu'$, we have shown that $\mu \preceq_{sd} \nu$.

Finally, for the third claim, suppose first that $\mu = \nu$. In that case we have $\mu \preceq_{sd} \nu$ and hence $\sigma(\mu,\nu) = 1$. Similarly, $\nu \preceq_{sd} \mu$, and hence $\sigma(\nu,\mu) = 1$. Conversely, suppose that $\sigma(\mu, \nu) = \sigma(\nu, \mu) = 1$. In that case both $\mu \preceq_{sd} \nu$ and $\nu \preceq_{sd} \mu$. Since $\preceq_{sd}$ is a partial order (Kamae and Krengel (1978)) it is antisymmetric. Hence $\mu = \nu$. $\square$

Proof of theorem 3.1. The theorem follows directly from inequalities (11)–(14). $\square$

Proof of proposition 3.1. Let μ and ν be given. It view of theorem 3.1, it suffices to construct a coupling $(X, Y) \in \mathscr{C}(\mu, \nu)$ such that $\mathbb{P}\{X \preceq Y\} = \sigma(\mu, \nu)$. In constructing the coupling, we can and do assume that $\sigma := \sigma(\mu, \nu) > 0$.[6]

To begin, observe that, by proposition 2.2, there exists a pair $(\mu', \nu') \in \Phi(\mu, \nu)$ with $\mu'(S) = \nu'(S) = \sigma$. Let

$$\mu^r := \frac{\mu - \mu'}{1 - \sigma} \quad \text{and} \quad \nu^r := \frac{\nu - \nu'}{1 - \sigma}.$$

It is easy to check that both μ^r and ν^r are probability measures, as are μ'/σ and ν'/σ. Moreover, μ and ν can be expressed as the convex combinations

$$\mu = (1 - \sigma)\mu^r + \sigma\frac{\mu'}{\sigma} \quad \text{and} \quad \nu = (1 - \sigma)\nu^r + \sigma\frac{\nu'}{\sigma}.$$

Thus, to draw from μ, one can draw from μ^r with probability $(1 - \sigma)$ and μ'/σ with probability σ. A similar statement holds for ν. We construct a coupling (X, Y) as follows. Let U, X', Y', X^r and Y^r be random variables on a common probability space such that

(a) $X' \stackrel{\mathscr{D}}{=} \mu'/\sigma$, $Y' \stackrel{\mathscr{D}}{=} \nu'/\sigma$, $X^r \stackrel{\mathscr{D}}{=} \mu^r$ and $Y^r \stackrel{\mathscr{D}}{=} \nu^r$
(b) U is uniform on $[0, 1]$
(c) U is independent of (X', Y', X^r, Y^r) and
(d) $\mathbb{P}\{X' \preceq Y'\} = 1$.

The only nontrivial existence problem here is (d), but such a pair can be constructed via Strassen's theorem since $\mu'/\sigma \preceq_{sd} \nu'/\sigma$. Now let

$$X := \mathbb{1}\{U \leq \sigma\}X' + \mathbb{1}\{U > \sigma\}X^r \quad \text{and} \quad Y := \mathbb{1}\{U \leq \sigma\}Y' + \mathbb{1}\{U > \sigma\}Y^r.$$

[6] If not, then for any $(X, Y) \in \mathscr{C}(\mu, \nu)$ we have $0 \leq \mathbb{P}\{X \preceq Y\} \leq \sigma(\mu, \nu) = 0$.

Evidently (X, Y) is a coupling of (μ, ν). Moreover, for this pair, we have

$$\mathbb{P}\{X \preceq Y\} \geq \mathbb{P}\{X \preceq Y,\ U \leq \sigma\} = \mathbb{P}\{X' \preceq Y',\ U \leq \sigma\}.$$

In view of independence the right hand side is equal to $\mathbb{P}\{X' \preceq Y'\}\mathbb{P}\{U \leq \sigma\}$. But this is just $\sigma = \sigma(\mu, \nu)$, so $\mathbb{P}\{X \preceq Y\} \geq \sigma(\mu, \nu)$. Theorem 3.1 tells us that the reverse is also true, so $\mathbb{P}\{X \preceq Y\} = \sigma(\mu, \nu)$ as claimed. □

Proof of proposition 4.1. Let $\{\mu_n\}$, $\{\nu_n\}$, μ and ν be as in the statement of the proposition. Applying theorem 3.1, we have

$$\begin{aligned}
|\sigma(\mu_n, \nu_n) - \sigma(\mu, \nu)| &= |1 - \sigma(\mu_n, \nu_n) - (1 - \sigma(\mu, \nu))| \\
&= \left| \sup_{I \in is\mathscr{B}} \{\mu_n(I) - \nu_n(I)\} - \sup_{I \in is\mathscr{B}} \{\mu(I) - \nu(I)\} \right| \\
&\leq \sup_{I \in is\mathscr{B}} |\mu_n(I) - \nu_n(I) - \{\mu(I) - \nu(I)\}| \\
&\leq \sup_{B \in \mathscr{B}} |\mu_n(B) - \mu(B)| + \sup_{B \in \mathscr{B}} |\nu_n(B) - \nu(B)| .
\end{aligned}$$

Since $\mu_n \xrightarrow{u} \mu$ and $\nu_n \xrightarrow{u} \nu$, we now have $\sigma(\mu_n, \nu_n) \to \sigma(\mu, \nu)$. □

Proof of proposition 5.1. Under our assumptions on $\{\epsilon_j\}$, the weak law of large numbers gives $g + \frac{1}{n}\sum_{j=1}^{n} \epsilon_j \to g$ in probability as $n \to \infty$. Hence $G_n \xrightarrow{w} G$. It follows from corollary 4.1 that $\sigma(F, G_n) \to \sigma(F, G)$. □

References

Ambrus, Attila, Eduardo M Azevedo, and Yuichiro Kamada (2013) "Hierarchical cheap talk," *Theoretical Economics*, Vol. 8, pp. 233–261.

Atkinson, Anthony B (1987) "On the measurement of poverty," *Econometrica*, Vol. 55, pp. 749–764.

Benhabib, Jess, Alberto Bisin, and Shenghao Zhu (2011) "The distribution of wealth and fiscal policy in economies with finitely lived agents," *Econometrica*, Vol. 79, pp. 123–157.

Blackorby, Charles and David Donaldson (1980) "Ethical indices for the measurement of poverty," *Econometrica*, pp. 1053–1060.

Border, Kim C (1992) "Revealed preference, stochastic dominance, and the expected utility hypothesis," *Journal of Economic Theory*, Vol. 56, pp. 20–42.

Brock, William A and Steven N Durlauf (2001) "Discrete choice with social interactions," *The Review of Economic Studies*, Vol. 68, pp. 235–260.

Davidson, Russell and Jean-Yves Duclos (2013) "Testing for restricted stochastic dominance," *Econometric Reviews*, Vol. 32, pp. 84–125.

Devroye, L. and G. Lugosi (2001) *Combinatorial Methods in Density Estimation*, Springer Series in Statistics: Springer New York.

Dudley, Richard M (2002) *Real Analysis and Probability*, Vol. 74: Cambridge University Press.

Föllmer, H. and A. Schied (2011) *Stochastic Finance: An Introduction in Discrete Time*, De Gruyter Textbook Series: De Gruyter.

Foster, James E and Anthony F Shorrocks (1988) "Poverty orderings," *Econometrica*, pp. 173–177.

Hadar, Josef and William R Russell (1971) "Stochastic dominance and diversification," *Journal of Economic Theory*, Vol. 3, pp. 288–305.

Halevy, Yoram and Vincent Feltkamp (2005) "A Bayesian approach to uncertainty aversion," *The Review of Economic Studies*, Vol. 72, pp. 449–466.

Hernández-Lerma, Onésimo and Jean B Lasserre (2003) *Markov Chains and Invariant Probabilities*: Springer.

Jarrow, Robert (1986) "The relationship between arbitrage and first order stochastic dominance," *The Journal of Finance*, Vol. 41, pp. 915–921.

Kamae, Teturo and Ulrich Krengel (1978) "Stochastic partial ordering," *The Annals of Probability*, Vol. 6, pp. 1044–1049.

Kamae, Teturo, Ulrich Krengel, and George L O'Brien (1977) "Stochastic inequalities on partially ordered spaces," *The Annals of Probability*, pp. 899–912.

Lee, Sokbae, Oliver Linton, and Yoon-Jae Whang (2009) "Testing for stochastic monotonicity," *Econometrica*, Vol. 77, pp. 585–602.

Lehmann, Erich L (1955) "Ordered families of distributions," *The Annals of Mathematical Statistics*, pp. 399–419.

Leshno, Moshe and Haim Levy (2002) "Preferred by all and preferred by most decision makers: Almost stochastic dominance," *Management Science*, Vol. 48, pp. 1074–1085.

Levy, H. (2006) *Stochastic Dominance: Investment Decision Making under Uncertainty*, Studies in Risk and Uncertainty: Springer.

Levy, Haim (1992) "Stochastic dominance and expected utility: survey and analysis," *Management Science*, Vol. 38, pp. 555–593.

Levy, Moshe (2009) "Almost stochastic dominance and stocks for the long run," *European Journal of Operational Research*, Vol. 194, pp. 250–257.

Linton, Oliver, Esfandiar Maasoumi, and Yoon-Jae Whang (2005) "Consistent testing for stochastic dominance under general sampling schemes," *The Review of Economic Studies*, Vol. 72, pp. 735–765.

Machina, Mark J (1982) "'Expected utility' analysis without the independence axiom," *Econometrica*, pp. 277–323.

Mas-Colell, A., M.D. Whinston, and J.R. Green (1995) *Microeconomic Theory*: Oxford University Press.

Maskin, Eric and John Riley (2000) "Asymmetric auctions," *Review of Economic studies*, pp. 413–438.

Piketty, Thomas (1997) "The dynamics of the wealth distribution and the interest rate with credit rationing," *The Review of Economic Studies*, Vol. 64, pp. 173–189.

Quirk, James P and Rubin Saposnik (1962) "Admissibility and measurable utility functions," *The Review of Economic Studies*, pp. 140–146.

Rosenberg, Dinah, Eilon Solan, and Nicolas Vieille (2007) "Social learning in one-arm bandit problems," *Econometrica*, Vol. 75, pp. 1591–1611.

Safra, Zvi and Uzi Segal (2008) "Calibration results for non-expected utility theories," *Econometrica*, Vol. 76, pp. 1143–1166.

Stokey, Nancy and Robert E Lucas (1989) *Recursive Methods in Economic Dynamics (with EC Prescott)*: Harvard University Press.

Stoyanov, Stoyan V, Svetlozar T Rachev, and Frank J Fabozzi (2012) "Metrization of stochastic dominance rules," *International Journal of Theoretical and Applied Finance*, Vol. 15.

Strassen, Volker (1965) "The existence of probability measures with given marginals," *The Annals of Mathematical Statistics*, pp. 423–439.

Tversky, Amos and Daniel Kahneman (1992) "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, Vol. 5, pp. 297–323.

Werner, Jan (1997) "Diversification and equilibrium in securities markets," *Journal of Economic Theory*, Vol. 75, pp. 89–103.

Williams, David (1991) *Probability with Martingales*: Cambridge University Press.