Journal of Risk and Uncertainty, 14:103–127 (1997)

Increasing Risk: Some Direct Constructions

MARK J. MACHINA
Department of Economics, University of California, San Diego, La Jolla, CA 92093

JOHN W. PRATT
Graduate School of Business Administration, Harvard University, Boston, MA 02163

Abstract

This article extends the classic Rothschild–Stiglitz characterization of comparative risk ("increasing risk") in two directions. By adopting a more general definition of "mean preserving spread" (MPS), it provides a direct construction of a sequence of MPS's linking any pair of distributions that are ranked in terms of comparative risk. It also provides a direct, explicit construction of a zero-conditional-mean "noise" variable for any such pair of distributions. Both results are extended to the case of second order stochastic dominance.

Key words: risk, comparative risk, increasing risk, mean preserving spread, noise, stochastic dominance

JEL Classification: D81

This article offers some contributions to the theory of comparative risk ("increasing risk"), arising from the work of Hardy, Littlewood, and Pólya (1929: 1934, pp.49,89), Blackwell (1951, 1953), Strassen (1965), Hadar and Russell (1969), Hanoch and Levy (1969), and others, culminating in the four-way characterization of Rothschild and Stiglitz (1970, 1971). Sparked especially by the latter, this topic has received widespread attention and application in the economics of uncertainty, and at this point it is hard to envision what the field would look like without it.[1]

Rothschild and Stiglitz present four notions of what it might mean for a univariate cumulative distribution function $G(\cdot)$ to be "riskier" than a distribution $F(\cdot)$ with the same mean, and show that each is equivalent to the others. Loosely speaking, these are:

- $G(\cdot)$ can be obtained from $F(\cdot)$ by a sequence of one or more "mean preserving spreads."
- $G(\cdot)$ can be obtained from $F(\cdot)$ by the addition of zero-conditional-mean "noise."
- $G(\cdot)$ and $F(\cdot)$ satisfy an "integral condition" over their combined support.
- All risk averse expected utility maximizers weakly prefer $F(\cdot)$ to $G(\cdot)$.

However, the original Rothschild–Stiglitz characterization is not as general or direct as it might be. The original definition of a mean preserving spread only included spreads from one discrete distribution to another, or from one density function to another, and a more general definition has since appeared, for which we argue below. Concerning the characterization itself:[2]

- *Sequences of MPS's* are explicitly constructed only between pairs of finite-outcome probability distributions. For more general $F(\cdot)$ and $G(\cdot)$, the formal result consists of constructing two sequences of finite-outcome distributions, $F_n(\cdot) \to F(\cdot)$ and $G_n(\cdot) \to G(\cdot)$, such that for each n, $G_n(\cdot)$ can be obtained from $F_n(\cdot)$ by a finite number of MPS's.
- The *noise variable* is also explicitly constructed only for pairs of finite-outcome distributions. For general $F(\cdot)$ and $G(\cdot)$, the formal result consists of constructing a noise variable for each $(F_n(\cdot),G_n(\cdot))$ pair in the above sequences, and invoking a limit theorem to establish the existence of a noise variable linking $F(\cdot)$ and $G(\cdot)$.

The purpose of this article is to present more general and direct versions of these results. Specifically:

- Given arbitrary distributions $F(\cdot)$ and $G(\cdot)$ satisfying the integral condition, we explicitly construct a sequence of MPS's, beginning at $F(\cdot)$ and converging directly to $G(\cdot)$.
- Given arbitrary distributions $F(\cdot)$ and $G(\cdot)$ satisfying the integral condition, we explicitly construct a zero-conditional-mean noise variable $\tilde{\varepsilon}$, such that if $\tilde{x}$ has distribution $F(\cdot)$, then $\tilde{x} + \tilde{\varepsilon}$ will have distribution $G(\cdot)$.

We then incorporate these two results to obtain a more general and direct statement of the four-way Rothschild–Stiglitz characterization of comparative risk, and extend them to the case of second order stochastic dominance. We conclude with mention of related work.

1. Argument for a more general definition of MPS

We adopt the Rothschild–Stiglitz setting of all cumulative distribution functions $F(\cdot)$, $G(\cdot)$, … over a closed bounded interval, say $[0,M]$.[3] In (1970, Section II), they define $G(\cdot)$ to differ from $F(\cdot)$ by a mean preserving spread under either of the following conditions:

For pairs of discrete distributions: $F(\cdot)$ and $G(\cdot)$ assign identical probabilities except to the points $x_1 < x_2 < x_3 < x_4$, where the differences in their probabilities satisfy:

$$g_1 - f_1 = \alpha \qquad g_2 - f_2 = -\alpha \qquad g_3 - f_3 = -\beta \qquad g_4 - f_4 = \beta. \tag{1}$$

For pairs of distributions with density functions: $F(\cdot)$ and $G(\cdot)$ have identical densities except over the nonoverlapping intervals $(x_1, x_1 + t)$, $(x_2, x_2 + t)$, $(x_3, x_3 + t)$ and $(x_4, x_4 + t)$ with $x_1 < x_2 < x_3 < x_4$, where the difference in their densities satisfies:

$$\underbrace{g(x) - f(x) \equiv \alpha}_{\text{for all } x \in (x_1, x_1 + t)} \qquad \underbrace{g(x) - f(x) \equiv -\alpha}_{\text{for all } x \in (x_2, x_2 + t)}$$

$$\underbrace{g(x) - f(x) \equiv -\beta}_{\text{for all } x \in (x_3, x_3 + t)} \qquad \underbrace{g(x) - f(x) \equiv \beta}_{\text{for all } x \in (x_4, x_4 + t).} \tag{2}$$

In each case we require $\alpha, \beta \geq 0$ (to ensure a spread) and $\alpha \cdot (x_2 - x_1) = \beta \cdot (x_4 - x_3)$ (to ensure equal means).

Both these cases exemplify the Rothschild–Stiglitz idea of a *single* (or "basic") spread of mass, from the center of a distribution toward its tails. But there are other types of "spreads," arguably just as "basic," that do not take the form of (1) or (2). For example:

- a spread of the degenerate distribution at 1/2 to the uniform distribution on [0,1]
- a spread of the uniform distribution on [0,1] to a 50:50 chance of 0:1.

However, the most restrictive feature of the Rothschild–Stiglitz definition is that it only applies to distributions that are discrete, or have density functions. Yet we would also want to classify the following as basic or "single" mean preserving spreads:

- a spread of the degenerate distribution at 1/2 to the Cantor distribution[4] $C(\cdot)$ on [0,1]
- a spread of the Cantor distribution on [0,1] to a 50:50 chance of 0:1 even though the Cantor distribution possesses neither mass points *nor* a density.

This last example illustrates why a more general notion of mean preserving spread would be required to obtain a direct "sequence of mean preserving spreads" result for general univariate probability distributions: Although the two distributions in the example clearly satisfy each of the *other three* notions of comparative risk (the integral condition, addition of noise,[5] and unanimity of preference), there can be *no* sequence $C(\cdot) = F_0(\cdot)$, $F_1(\cdot)$, $F_2(\cdot)$ … in which each step takes the form (1) or (2). The reason is that $C(\cdot)$ has neither positive mass points nor any positive density, so there can be no first distribution $F_1(\cdot)$ that differs from $C(\cdot)$ by either (1) or (2). In light of this, it seems that the original Rothschild–Stiglitz "double-sequence" result is probably the strongest one that can be obtained under their definition of mean preserving spread.

We accordingly adopt the following definition, which essentially consists of any mean preserving transfer of some or all of the probability mass within a finite interval, out to or beyond its end points. It is worded so that it may be applied to any pair of finite-mean distributions over $(-\infty,+\infty)$:[6]

Definition. A cumulative distribution function $G(\cdot)$ is said to differ from $F(\cdot)$ by a *mean preserving spread* if $F(\cdot)$ and $G(\cdot)$ have the same finite expected value, and there exist outcome levels $x' \leq x''$ such that:[7]

$G(\cdot)$ assigns *at least as much* probability as $F(\cdot)$ to every subinterval of $(-\infty, x')$,
$G(\cdot)$ assigns *no more* probability than $F(\cdot)$ to every subinterval of (x', x''),
$G(\cdot)$ assigns *at least as much* probability as $F(\cdot)$ to every subinterval of $(x'', +\infty)$.

Some remarks:

1. Note that the probabilities assigned to the points x' and/or x'' could either rise, remain unchanged, or drop. This follows because x' and/or x'' could be either part of the central region that loses probability, or else one of the tail regions that gain probability.
2. Like (1) and (2), this definition implies that each distribution $F(\cdot)$ differs from *itself* by a mean preserving spread, which we term a *null* spread.
3. The case of $x' = x''$, which implies that the interval (x', x'') is empty, does not necessarily imply a null spread. It also includes the case of spreading a strictly positive amount of probability mass from the point x' to the intervals $(-\infty, x')$ and $(x', +\infty)$, because the condition that $G(\cdot)$ assign at least as much probability as $F(\cdot)$ to every subinterval of $(-\infty, x')$ and $(x', +\infty)$ implies that $G(\cdot)$ assigns no greater probability than $F(\cdot)$ to x'.

All subsequent references to mean preserving spread are to this more general definition,[8] and we refer to cases (1) and (2) as *R-S mean preserving spreads*.

2. A direct sequence of mean preserving spreads

The original Rothschild–Stiglitz characterization linked the "sequence of MPS's" condition and the other notions of comparative risk to the following "integral conditions" on a pair of distributions $F(\cdot)$ and $G(\cdot)$:[9]

$$\int_0^x [G(\omega) - F(\omega)]d\omega \geq 0 \text{ for all } x \in [0, M) \quad \text{and} \quad \int_0^M [G(\omega) - F(\omega)]d\omega = 0. \tag{3}$$

It is straightforward to show that if a sequence of distributions $F(\cdot) = H_0(\cdot), H_1(\cdot), H_2(\cdot), \ldots$ converges to $G(\cdot)$, and each $H_{n+1}(\cdot)$ differs from $H_n(\cdot)$ by a single mean preserving spread (R-S or more general), then $F(\cdot)$ and $G(\cdot)$ will satisfy the integral condition (3).[10] As Rothschild and Stiglitz (1970, p. 231) note, however, they did not establish the exact converse of this statement, but rather that, if $F(\cdot)$ and $G(\cdot)$ satisfy the integral condition, then there exist two sequences of discrete distributions, $F_n(\cdot) \to F(\cdot)$ and $G_n(\cdot) \to G(\cdot)$, such that each $G_n(\cdot)$ differs from $F_n(\cdot)$ by a finite number of (discrete) R-S mean preserving spreads.

However, the above more general definition of an MPS does allow us to establish the full converse, by constructing a sequence $H_0(\cdot), H_1(\cdot), H_2(\cdot), \ldots$ directly from $F(\cdot)$ to $G(\cdot)$:[11]

Theorem 1 (Construction of a Sequence of MPS's). If the cumulative distribution functions $F(\cdot)$ and $G(\cdot)$ on $[0,M]$ satisfy the integral condition (3), then one can construct a sequence of distributions $F(\cdot) = H_0(\cdot), H_1(\cdot), H_2(\cdot) \ldots$, converging to $G(\cdot)$, such that for every n, $H_{n+1}(\cdot)$ differs from $H_n(\cdot)$ by a mean preserving spread.[12]

We sketch the construction here. Each step takes a distribution $H(\cdot)$ that is "less risky" than $G(\cdot)$ (i.e., satisfies the integral condition with respect to $G(\cdot)$), and generates a mean preserving spread to construct a new distribution $H_{+1}(\cdot)$ that is still less risky than $G(\cdot)$, but is now "closer" to $G(\cdot)$. Each spread in the construction consists of shifting all the probability mass of an interval out to its end points, and the spreads are chosen so that the resulting sequence of distributions converges to $G(\cdot)$.

Consider Figure 1, which plots the two *integrated cumulative functions*

$$\mathcal{H}(x) \equiv \int_0^x H(\omega)\, d\omega \quad \text{and} \quad \mathcal{G}(x) \equiv \int_0^x G(\omega)\, d\omega. \tag{4}$$

By the integral condition (3), we know that $\mathcal{H}(0) = \mathcal{G}(0) = 0$, $\mathcal{H}(M) = \mathcal{G}(M)$, and $\mathcal{H}(\cdot)$ lies on or below $\mathcal{G}(\cdot)$ everywhere else on $[0,M]$. Note that because $\mathcal{H}(\cdot)$ and $\mathcal{G}(\cdot)$ are integrals of nondecreasing nonnegative functions, they are both nondecreasing convex functions.

Pick any point on the curve $\mathcal{G}(\cdot)$ and consider its tangent line $\mathcal{L}(\cdot)$, which by convexity must lie everywhere on or below $\mathcal{G}(\cdot)$. We use this tangent to "slice" the curve of $\mathcal{H}(\cdot)$, in the sense that we extend $\mathcal{L}(\cdot)$ northeast and southwest until it meets $\mathcal{H}(\cdot)$ at the outcome levels x' and x'', then define $\mathcal{H}_{+1}(\cdot)$ to be the pointwise maximum (upper envelope) of $\mathcal{H}(\cdot)$ and the line $\mathcal{L}(\cdot)$.

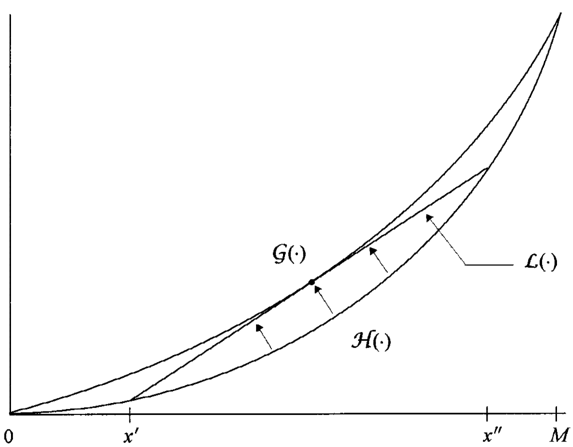

Figure 1. "Slicing" the integrated cumulative function $\mathcal{H}(\cdot)$ to generate a mean preserving spread.

"Slicing" the integrated cumulative in this manner shifts all the probability mass from the interval (x',x'') out to its end points, without changing the mean of the distribution. This can be seen as follows: Because $H_{+1}(\cdot)$ is convex, it is the integrated cumulative of a distribution $H_{+1}(\cdot)$ that is identical to $H(\cdot)$ below x' and above x'', but is constant over (x',x''). Accordingly, $H_{+1}(\cdot)$ assigns the same probability as $H(\cdot)$ to every interval below x' and every interval above x'', but assigns *zero* probability to every subinterval of (x',x''). This implies that $H_{+1}(\cdot)$ is obtained from $H(\cdot)$ by shifting all the mass from (x',x'') out to its end points x' and x''. The positive probability that $H_{+1}(\cdot)$ assigns to each point x' and x''—that is, its discontinuous jumps at these points—is reflected in the figure by the convex kinks in its integrated cumulative function $H_{+1}(\cdot)$ at x' and x''. In addition, because $H_{+1}(M) = H(M)$, $H_{+1}(\cdot)$ and $H(\cdot)$ have the same mean, so $H_{+1}(\cdot)$ differs from $H(\cdot)$ by a mean preserving spread. Finally, because $H_{+1}(\cdot)$ lies on or below $G(\cdot)$, $H_{+1}(\cdot)$ still satisfies the integral condition with respect to $G(\cdot)$.

Of course, this slice only generates a single mean preserving spread. However, as illustrated in Figure 2, we can then slice $H_{+1}(\cdot)$ to obtain a new integrated cumulative $H_{+2}(\cdot)$, closer to $G(\cdot)$, whose corresponding distribution $H_{+2}(\cdot)$ differs from $H_{+1}(\cdot)$ by another mean preserving spread. Continuing in this way, the successive slices can be chosen to generate a sequence of integrated cumulatives $H(\cdot)$, $H_{+1}(\cdot)$, $H_{+2}(\cdot)$ … approaching $G(\cdot)$, whose corresponding distributions $H(\cdot)$, $H_{+1}(\cdot)$, $H_{+2}(\cdot)$ … converge to the distribution $G(\cdot)$.

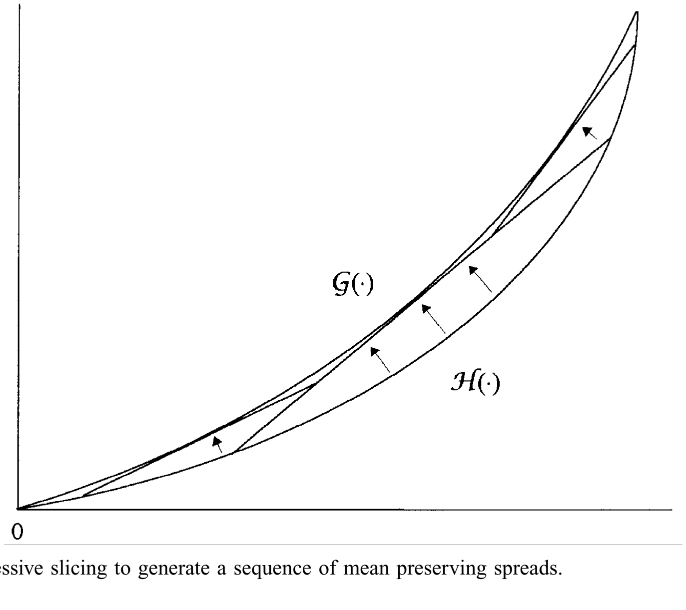

Figure 2. Successive slicing to generate a sequence of mean preserving spreads.

3. Direct construction of "noise"

The Rothschild–Stiglitz characterization implies that if the distributions $F(\cdot)$ and $G(\cdot)$ satisfy the integral condition (3), then there exists a joint distribution of the random variables $(\tilde{x},\tilde{\varepsilon})$, such that the marginal distribution of $\tilde{x}$ is $F(\cdot)$, $E[\tilde{\varepsilon} \mid x] \equiv 0$, and the distribution of $\tilde{z} = \tilde{x} + \tilde{\varepsilon}$ is $G(\cdot)$. Thus, $G(\cdot)$ can be thought of as being obtained from $F(\cdot)$ by the addition to $\tilde{x}$ of a zero-conditional-mean "noise" variable $\tilde{\varepsilon}$.

Rothschild and Stiglitz (1970, pp.238-240) construct such a noise variable $\tilde{\varepsilon}$, and thus a joint distribution $K(\tilde{x},\tilde{\varepsilon})$, for any pair of finite-outcome distributions that differ by a single R-S mean preserving spread. They extend this result to distributions that differ by a finite number of mean preserving spreads by proving that finite compositions of such noise variables are also noise.[13] To extend the result to arbitrary distributions $F(\cdot)$ and $G(\cdot)$ satisfying (3), they: (a) construct sequences $F_n(\cdot) \to F(\cdot)$ and $G_n(\cdot) \to G(\cdot)$ of finite-outcome distributions such that each $G_n(\cdot)$ differs from $F_n(\cdot)$ by a finite number of MPS's, so that by the earlier result there will be a joint distribution $K_n(\cdot,\cdot)$ linking each pair; (b) use a limit theorem[14] to show that the sequence $K_n(\cdot,\cdot)$ has a subsequence converging to some joint distribution $K^*(\cdot,\cdot)$; and (c) prove that $K^*(\cdot,\cdot)$ must satisfy the "addition of noise" condition with respect to $F(\cdot)$ and $G(\cdot)$.

The following result constructs the noise variable directly from any pair of distributions $F(\cdot)$ and $G(\cdot)$ that satisfy the integral condition (3):

Theorem 2 (Construction of Zero-Conditional-Mean Noise). If the cumulative distribution functions $F(\cdot)$ and $G(\cdot)$ on $[0,M]$ satisfy the integral condition (3), then one can construct a set of random variables $\{\tilde{\varepsilon}(x) \mid x \in [0,M]\}$, with $E[\tilde{\varepsilon}(x)] = 0$ for each x, such that if $\tilde{x}$ has distribution $F(\cdot)$ then $\tilde{y} = \tilde{x} + \tilde{\varepsilon}(\tilde{x})$ has distribution $G(\cdot)$.[15]

Figure 3 illustrates our construction in the case of a pair of strictly increasing, differentiable distributions $F(\cdot)$ and $G(\cdot)$. In the figure, "+" indicates the area between the functions where $G(\cdot)$ lies to the left of $F(\cdot)$, and "−" indicates the area where $G(\cdot)$ lies to the right of $F(\cdot)$. The conditional noise variables are defined a pair at a time. Specifically, pick probability levels u and v such that the *horizontal* "+" area up to u equals the horizontal "−" area up to v, and let x_u, x_v, y_u, and y_v satisfy $u = F(x_u) = G(y_u)$ and $v = F(x_v) = G(y_v)$. Define the noise variable $\tilde{\varepsilon}(x_u)$ so that it has mean 0 and takes the outcome x_u to the outcomes $\{y_u, y_v\}$, and similarly for $\tilde{\varepsilon}(x_v)$. In other words, define[16]

$$\tilde{\varepsilon}(x_u) = \begin{cases} y_u - x_u \text{ with probability } \dfrac{y_v - x_u}{y_v - y_u} \\ y_v - x_u \text{ with probability } \dfrac{x_u - y_u}{y_v - y_u} \end{cases}$$

$$\tilde{\varepsilon}(x_v) = \begin{cases} y_u - x_v \text{ with probability } \dfrac{y_v - x_v}{y_v - y_u} \\ y_v - x_v \text{ with probability } \dfrac{x_v - y_u}{y_v - y_u} \end{cases} \tag{5}$$

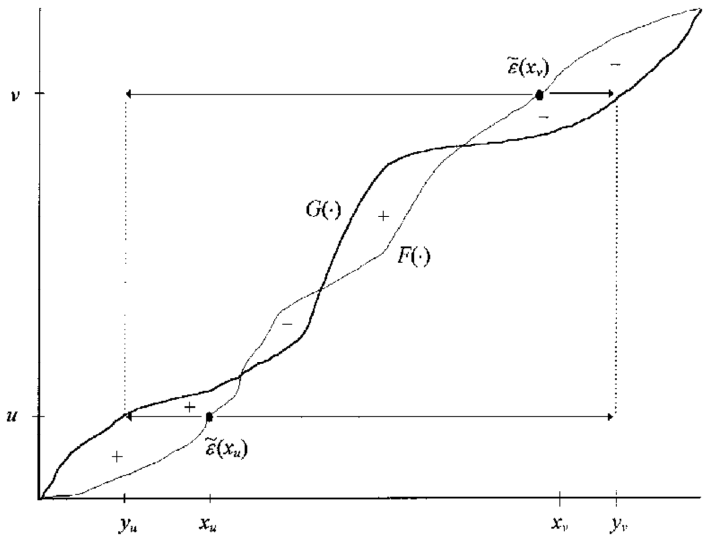

Figure 3. Construction of the conditional noise variables $\tilde{\varepsilon}(x_u)$ and $\tilde{\varepsilon}(x_v)$.

so that $\tilde{y}(x) = x + \tilde{\varepsilon}(x)$ will satisfy

$$\tilde{y}(x_u) = \begin{cases} y_u \text{ with probability } \dfrac{y_v - x_u}{y_v - y_u} \\ y_v \text{ with probability } \dfrac{x_u - y_u}{y_v - y_u} \end{cases}$$

$$\tilde{y}(x_v) = \begin{cases} y_u \text{ with probability } \dfrac{y_v - x_v}{y_v - y_u} \\ y_v \text{ with probability } \dfrac{x_v - y_u}{y_v - y_u} \end{cases} \tag{6}$$

Adding the random variable $\tilde{\varepsilon}(x_u)$ to x_u thus serves to split the probability (density) originally assigned to x_u by the distribution $F(\cdot)$, sending proportion $(y_v - x_u)/(y_v - y_u)$ of it down to the outcome value y_u, and the rest up to y_v. Similarly, adding $\tilde{\varepsilon}(x_v)$ to x_v splits the original probability at x_v sending proportion $(y_v - x_v)/(y_v - y_u)$ down to y_u and the rest up to y_v. This construction yields a zero-conditional-mean noise variable $\tilde{\varepsilon}(x)$ for all $x \in [0,M]$.

To see that the aggregate effect is to yield a random variable $\tilde{y} = \tilde{x} + \tilde{\varepsilon}$ with distribution $G(\cdot)$, consider Figure 4, where the horizontal "+" area up to u again equals the horizontal "−" area up to v, and similarly for u' and v'. It is clear that the probability ultimately

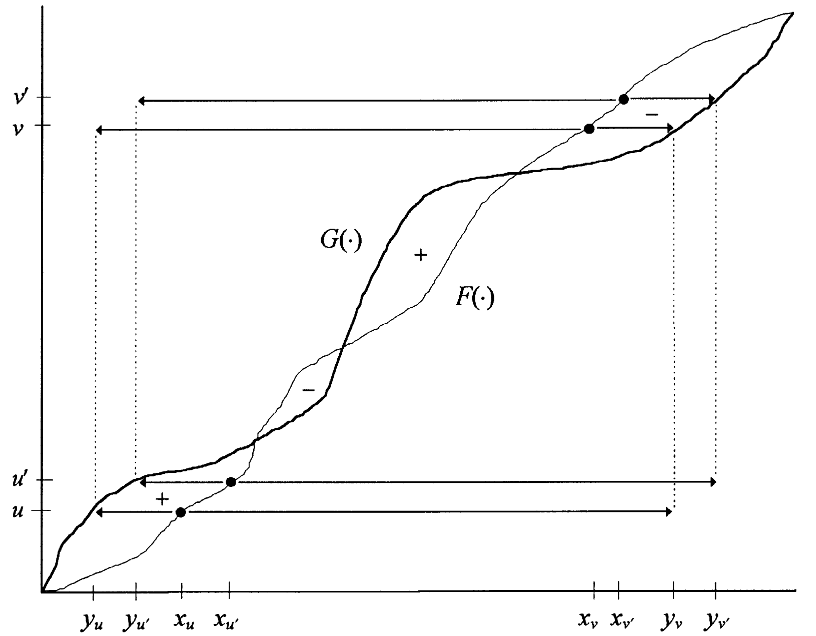

Figure 4. Demonstration that $\tilde{y} = \tilde{x} + \tilde{\varepsilon}$ has distribution $G(\cdot)$.

assigned to the interval $[y_u, y_{u'}]$—in other words, $Pr\{\tilde{y} \in [y_u, y_{u'}]\}$—consists of the amount of probability sent there from $[x_u, x_{u'}]$ plus the amount sent there from $[x_v, x_{v'}]$, so that we have

$$Pr\{\tilde{y} \in [y_u, y_{u'}]\} = Pr\{\tilde{y} \in [y_u, y_{u'}] \,|\, \tilde{x} \in [x_u, x_{u'}]\} \cdot Pr\{\tilde{x} \in [x_u, x_{u'}]\} + Pr\{\tilde{y} \in [y_u, y_{u'}] | \tilde{x} \in [x_v, x_{v'}]\} \cdot Pr\{\tilde{x} \in [x_v, x_{v'}]\}. \tag{7}$$

The above discussion implies that when the intervals $[x_u, x_{u'}]$ and $[x_v, x_{v'}]$ are very small, we can approximate the above two conditional probabilities by $(y_v - x_u)/(y_v - y_u)$ and $(y_v - x_v)/(y_v - y_u)$, respectively. Indicating first-order approximations[17] by $\approx$, we accordingly have

$$Pr\{\tilde{y} \in [y_u, y_{u'}]\} \approx \frac{y_v - x_u}{y_v - y_u} \cdot Pr\{\tilde{x} \in [x_u, x_{u'}]\} + \frac{y_v - x_v}{y_v - y_u} \cdot Pr\{\tilde{x} \in [x_v, x_{v'}]\}$$

$$= \frac{y_v - x_u}{y_v - y_u} \cdot (u' - u) + \frac{y_v - x_v}{y_v - y_u} \cdot (v' - v)$$

$$\approx \frac{y_v - x_u}{y_v - y_u} \cdot (u' - u) + \frac{x_u - y_u}{y_v - y_u} \cdot (u' - u)$$
$$= u' - u = G(y_{u'}) - G(y_u), \tag{8}$$

where the second approximation follows because the "+" area between the levels u and u' equals the "−" area between the levels v and v', so that $(y_v - x_v) \cdot (v' - v) \approx (x_u - y_u) \cdot (u' - u)$. A corresponding argument establishes the first-order approximation $Pr\{\tilde{y} \in [y_v, y_{v'}]\} \approx G(y_{v'}) - G(y_v)$. Therefore, the random variable $\tilde{y} = \tilde{x} + \tilde{\varepsilon}$ has distribution $G(\cdot)$. The formal proof establishes this result for arbitrary (e.g., not necessarily differentiable, continuous, or strictly increasing) distributions over $[0,M]$.[18]

4. A strengthened characterization of comparative risk

Theorems 1 and 2 can be combined with classical results to fully extend the four-way R-S characterization from finite-outcome distributions and/or densities to arbitrary pairs of equal-mean probability distributions over $[0,M]$:

Theorem 3 (Characterization of Comparative Risk). The following conditions on a pair of cumulative distribution functions $F(\cdot)$ and $G(\cdot)$ over an outcome interval $[0,M]$ are equivalent, and each implies that $F(\cdot)$ and $G(\cdot)$ have the same mean:

(a) $G(\cdot)$ can be obtained from $F(\cdot)$ by a sequence of mean preserving spreads; that is, there exists a sequence $F(\cdot) = H_0(\cdot), H_1(\cdot), H_2(\cdot) \ldots$ converging to $G(\cdot)$, such that for every n, $H_{n+1}(\cdot)$ differs from $H_n(\cdot)$ by a mean preserving spread.[19]

(b) There exists a pair of jointly distributed random variables $(\tilde{x}, \tilde{\varepsilon})$ with $E[\tilde{\varepsilon} \mid x] = 0$ for all x, such that $F(\cdot)$ and $G(\cdot)$ are the cumulative distribution functions of $\tilde{x}$ and $\tilde{x} + \tilde{\varepsilon}$, respectively.

(c) $\int_0^x [G(\omega) - F(\omega)]d\omega \geq 0$ for all $x \in [0,M)$, and $\int_0^M [G(\omega) - F(\omega)]d\omega = 0$.

(d) $\int_0^M U(x)dG(x) \leq \int_0^M U(x)dF(x)$ for every concave function $U(\cdot)$ over $[0,M]$.[20]

In such a case we say that $G(\cdot)$ differs from $F(\cdot)$ by a *mean preserving increase in risk* (MPIR), and that $F(\cdot)$ differs from $G(\cdot)$ by a *mean preserving reduction in risk* (MPRR).

The implications (c) ⇒ (a) and (c) ⇒ (b) in this characterization are Theorems 1 and 2 above. The implications (a) ⇒ (c) and (b) ⇒ (d) are straightforward (Rothschild and Stiglitz's arguments apply to general probability distributions), and (d) ⇒ (c) can be proved via integration by parts and proper selection of concave functions $U(\cdot)$ (Rothschild and Stiglitz (1970, p. 238)).

5. Extension to second order stochastic dominance

The notion of comparative risk is the equal-means case of *second order stochastic dominance* (SSD) over pairs of probability distributions. Hadar and Russell (1969), Hanoch and Levy (1969), and others have shown that a distribution $F(\cdot)$ is weakly preferred to $G(\cdot)$ by all risk averse expected utility maximizers if and only if

$$\int_0^x [G(\omega) - F(\omega)]d\omega \geq 0 \quad \text{for all } x \in [0,M] \tag{9}$$

in which case $F(\cdot)$ is said to *second order stochastically dominate* $G(\cdot)$. Note that (9) implies that the mean of $F(\cdot)$ is greater than or equal to the mean of $G(\cdot)$, and (3) is stronger only in requiring equality at $x = M$, that is, equal means. Second order stochastic dominance can be thought of as a combination of a mean preserving reduction in risk and/or first order stochastic dominance, where $F(\cdot)$ is said to *first order stochastically dominate* $G(\cdot)$ if $F(x) \leq G(x)$ for all $x \in [0,M]$ (e.g., Quirk and Saposnik (1962)). A simple special case of a first order stochastically dominating shift is:

Definition. A distribution $F(\cdot)$ is said to differ from $G(\cdot)$ by a *rightward shift of probability mass* if there exists an outcome level x' such that:

$F(\cdot)$ assigns *at least as much* probability as $G(\cdot)$ to every subinterval of $(x',+\infty)$,
$F(\cdot)$ assigns *no more* probability than $G(\cdot)$ to every subinterval of $(-\infty,x')$.

In such a case, $G(\cdot)$ is said to differ from $F(\cdot)$ by a *leftward shift of probability mass*.[21]

Each of the equal-mean constructions of Theorems 1 and 2 can be extended to the case of second order stochastic dominance. Specifically, given arbitrary distributions $F(\cdot)$ and $G(\cdot)$ over $[0,M]$ satisfying the SSD integral condition (9):

- One can construct a sequence of distributions $F(\cdot) = H_0(\cdot), H_1(\cdot), H_2(\cdot) \ldots$ converging to $G(\cdot)$, such that for each n, $H_{n+1}(\cdot)$ differs from $H_n(\cdot)$ either by a mean preserving spread or by a leftward shift of probability mass.
- One can construct a family $\{\tilde{\varepsilon}(x) \mid x \in [0,M]\}$ of nonpositive mean noise variables, such that if $\tilde{x}$ has distribution $F(\cdot)$, then $\tilde{x} + \tilde{\varepsilon}$ will have distribution $G(\cdot)$.

Each of these constructions involves a simple adjustment to its equal-mean counterpart, which we can illustrate here.[22] In each case, assume that $F(\cdot)$ has a greater mean than $G(\cdot)$, so that $\int_0^M F(\omega)d\omega < \int_0^M G(\omega)d\omega$.[23] The SSD sequence $F(\cdot) = H_0(\cdot), H_1(\cdot), H_2(\cdot) \ldots$ involves an initial leftward shift of probability mass to obtain a distribution $H_1(\cdot)$, which differs from $G(\cdot)$ by the equal-mean integral condition (3). Figure 5 shows how $H_1(\cdot)$ can be constructed from the integrated cumulative functions $\mathcal{F}(x) \equiv \int_0^x F(\omega)d\omega$ and $\mathcal{G}(x) \equiv \int_0^x G(\omega)d\omega$. Because $\mathcal{F}(M) < \mathcal{G}(M)$ and $\mathcal{G}(\cdot)$ is a convex function whose slope is never greater than one, the straight line $W(\cdot)$ that passes through the point $(M, \mathcal{G}(M))$ and

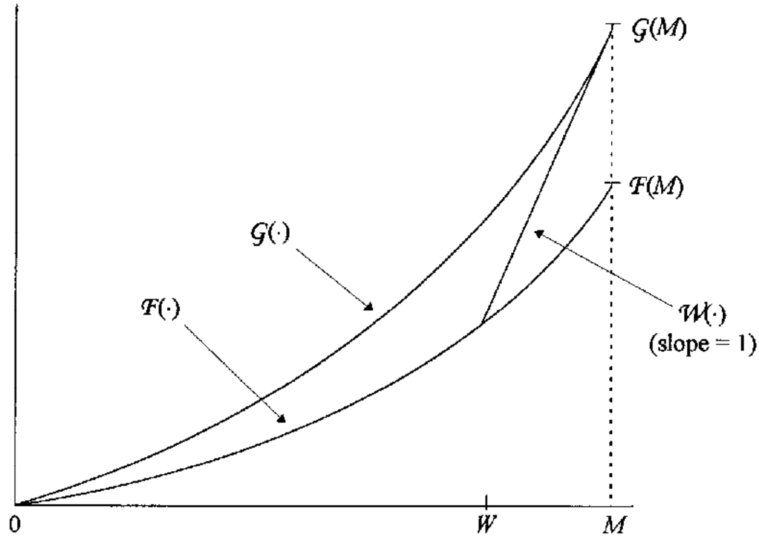

Figure 5. Generating a leftward shift of probability mass.

has unit slope must lie everywhere on or below $G(\cdot)$, and must intersect $F(\cdot)$ at some point $W \in [0,M)$. Let $H_1(\cdot)$ be the pointwise maximum of $F(\cdot)$ and $W(\cdot)$, and let $H_1(\cdot)$ be the distribution corresponding to $H(\cdot)$. The figure reveals that $H_1(\cdot)$ differs from $F(\cdot)$ by a leftward shift of probability mass, namely a shifting of all of $F(\cdot)$'s mass on the interval $[W,M]$ to the point W.[24] Because the figure also shows that $H_1(\cdot)$ satisfies the equal-mean integral condition (3) with respect to $G(\cdot)$, the distributions $H_2(\cdot)$, $H_3(\cdot)$ … can be obtained by mean preserving spreads as in the earlier construction.

The SSD noise variable $\tilde{\varepsilon}(x)$ differs from its equal-mean counterpart in that it is degenerate and negative for some outcome levels x. Because the greater mean of $F(\cdot)$ implies $\int_0^M [F(x) - G(x)]dx < 0$, so that the entire "+" area between $F(\cdot)$ and $G(\cdot)$ strictly exceeds the entire "−" area, there is some probability level $\bar{u}$ such that the "+" area up to $\bar{u}$ equals the entire "−" area, as illustrated in Figure 6. Thus, in addition to outcome pairs such as x_u, x_v and their corresponding equal-mean noise variables $\tilde{\varepsilon}(x_u)$, $\tilde{\varepsilon}(x_v)$ (as in Figure 3), there will be outcomes such as $x_{\hat{u}}$ with degenerate noise variable $\tilde{\varepsilon}(x_{\hat{u}}) \equiv y_{\hat{u}} - x_{\hat{u}} < 0$. The discussion following Figure 4 can be extended to this case by noting that, while the random variables such as $\tilde{\varepsilon}(x_u)$ and $\tilde{\varepsilon}(x_v)$ continue to shift probability mass both down and up (leftward and rightward), random variables such as $\tilde{\varepsilon}(x_{\hat{u}})$ shift *all* of the probability mass at $x_{\hat{u}}$ leftward to the outcome $y_{\hat{u}}$ so that the aggregate effect again yields a random variable $\tilde{x} + \tilde{\varepsilon}(x)$ with distribution $G(\cdot)$.

We can use these extended constructions to obtain the following four-way characterization of second order stochastic dominance:

Theorem 3′ (Characterization of Second Order Stochastic Dominance). The following conditions on a pair of cumulative distribution functions $F(\cdot)$ and $G(\cdot)$ over an outcome interval $[0,M]$ are equivalent, and each implies that the mean of $F(\cdot)$ is greater than or equal to the mean of $G(\cdot)$:

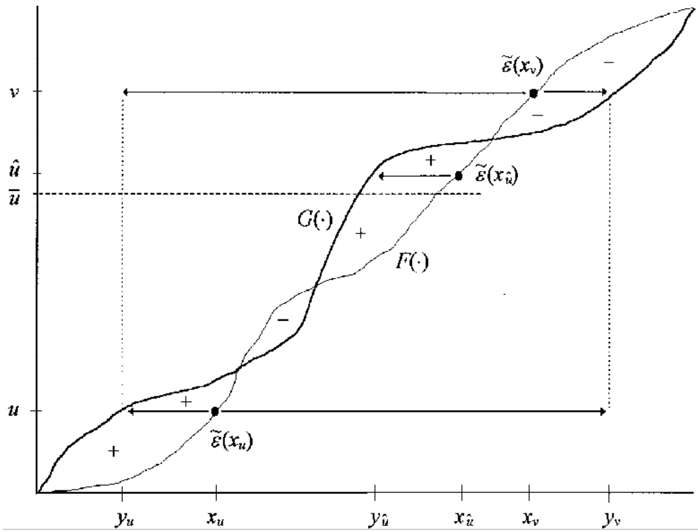

Figure 6. Construction of the conditional noise variables $\tilde{\varepsilon}(x_u)$, $\tilde{\varepsilon}(x_v)$, and $\tilde{\varepsilon}(x_{\hat{u}})$.

(a′) $G(\cdot)$ can be obtained from $F(\cdot)$ by a sequence of mean preserving spreads and/or leftward shifts of probability mass; that is, there exists a sequence $F(\cdot) = H_0(\cdot), H_1(\cdot), H_2(\cdot) \ldots$ converging to $G(\cdot)$, such that for every n, $H_{n+1}(\cdot)$ differs from $H_n(\cdot)$ by a mean preserving spread or a leftward shift of probability mass.

(b′) There exists a pair of jointly distributed random variables $(\tilde{x}, \tilde{\varepsilon})$ with $\mathrm{E}[\tilde{\varepsilon}|x] \leq 0$ for all x, such that $F(\cdot)$ and $G(\cdot)$ are the cumulative distribution functions of $\tilde{x}$ and $\tilde{x} + \tilde{\varepsilon}$ respectively.

(c′) $\int_0^x [G(\omega) - F(\omega)]d\omega \geq 0$ for all $x \in [0,M]$.

(d′) $\int_0^M U(x)dG(x) \leq \int_0^M U(x)dF(x)$ for every increasing concave function $U(\cdot)$ over $[0,M]$.

Implications $(c') \Rightarrow (a')$ and $(c') \Rightarrow (b')$, proven in the Appendix, are the SSD sequence and noise constructions. Implications $(a') \Rightarrow (c')$, $(b') \Rightarrow (d')$ and $(d') \Rightarrow (c')$ are straightforward extensions of their counterparts from theorem 3.

6. Related work

Landsberger and Meilijson (1990, Note 1) and Scarsini (1994, p. 356) attribute the more general definition of MPS to Machina, although we would not be surprised to find a prior appearance in the vast literature on this topic. In the course of completing this article, we discovered versions of the equal-mean and SSD sequence of spreads constructions among the results of Kaas, van Heerwaarden, and Goovaerts (1994, chapter 4; forthcoming) and Müller (1996).

As far as we know, our direct noise constructions have not appeared anywhere except in Pratt (1990). Fishburn and Vickson (1978, Theorem 2.6), Hadar and Seo (1990, Lemma 2), Kaas, van Heerwaarden, and Goovaerts (1994, Section IV.4), and others have shown that if $F(\cdot)$ and $G(\cdot)$ satisfy condition (c') (i.e. (9)), then there exists a joint distribution $(\tilde{x}, \tilde{\varepsilon})$, such that $E[\tilde{\varepsilon}|x] \leq 0$ for all x, and $\tilde{x}$ and $\tilde{x} + \tilde{\varepsilon}$ have distributions $F(\cdot)$ and $G(\cdot)$. However, unlike our construction $(c') \Rightarrow (b')$, those results have been nonconstructive, or restricted to pairs of distributions that are finite-outcome/have densities/cross a finite number of times, with extension to general distributions again represented by noise variables linking each pair $(F_n(\cdot), G_n(\cdot))$ in the finite-outcome sequences $F_n(\cdot) \to F(\cdot)$ and $G_n(\cdot) \to G(\cdot)$.

Appendix: proofs of theorems

Proof of Theorem 1. The proof has three steps. Step 1 shows that if two distributions over $[0,M]$ have common $1/n$-, $2/n$-, …, $(n-1)/n$-quantiles, then they differ vertically by at most $1/n$. Step 2 shows how to take a distribution $H(\cdot)$ that satisfies the integral condition with respect to $G(\cdot)$, and any probability u, and construct a mean preserving spread of $H(\cdot)$ to obtain a distribution $H^*(\cdot)$ that still satisfies the integral condition with respect to $G(\cdot)$, and has a common u-quantile with $G(\cdot)$. Step 3 combines these results to generate the sequence $F(\cdot) = H_0(\cdot), H_1(\cdot), H_2(\cdot)$ …, and shows that it converges to $G(\cdot)$.

Step 1 (common i/n quantiles implies $|H(\cdot) - G(\cdot)| \leq 1/n$). For any distribution $H(\cdot)$ over $[0,M]$, x_u is said to be a *u-quantile* of $H(\cdot)$ if $H(x-) \leq u \leq H(x)$, where $H(x-)$ denotes the left-hand limit of $H(\cdot)$ at x. Say there exist $\{x_{1/n}, x_{2/n}, \ldots, x_{(n-1)/n}\}$ such that $x_{i/n}$ is an i/n-quantile of both $H(\cdot)$ and $G(\cdot)$. Define $x_0 = 0$ (which is a 0-quantile of both $H(\cdot)$ and $G(\cdot)$) and $x_1 = M$ (a 1-quantile of $H(\cdot)$ and $G(\cdot)$). By the definition of a quantile, we must have $x_0 \leq x_{1/n} \leq \ldots \leq x_{(n-1)/n} \leq x_1$, so the intervals

$$[x_0, x_{1/n}), [x_{1/n}, x_{2/n}), \ldots, [x_{(n-1)/n}, x_1), [x_1, x_1]$$

form a partition of $[0,M]$.[25] If $x \in [x_{i/n}, x_{(i+1)/n})$, then $i/n \leq H(x) \leq (i+1)/n$ and $i/n \leq G(x) \leq (i+1)/n$. If $x \in [x_1, x_1]$, then $H(x) = G(x) = 1$. Thus, $|H(x) - G(x)| \leq 1/n$ for all $x \in [0,M]$.

Step 2 (construction of the spread).[26] Let $H(\cdot)$ be any distribution over $[0,M]$ that satisfies the integral condition with respect to $G(\cdot)$, and pick any probability level $u \in (0,1)$. Defining the functions $\mathcal{G}(x) \equiv \int_0^x G(\omega)d\omega$ and $\mathcal{H}(x) \equiv \int_0^x H(\omega)d\omega$, the integral condition implies $\mathcal{G}(x) \geq \mathcal{H}(x)$ for all $x \in [0,M)$, and $\mathcal{G}(M) = \mathcal{H}(M)$. As illustrated in Figure 1, $\mathcal{G}(\cdot)$ and $\mathcal{H}(\cdot)$ are nondecreasing, continuous, convex functions of x, with $\mathcal{G}(0) = \mathcal{H}(0) = 0$. The left and right derivatives of $\mathcal{G}(\cdot)$ at x are given by $G(x-)$ and $G(x)$ respectively, and similarly for $\mathcal{H}(\cdot)$.

Let $x_u = \min\{x \in [0,M] \mid u \leq G(x)\}$, so that $G(x_u-) \leq u \leq G(x_u)$, x_u is a u-quantile of $G(\cdot)$, and u lies between the left and right derivatives of $G(\cdot)$ at x_u. Thus the linear (affine) function $L(\cdot)$ defined by

$$L(x) \equiv G(x_u) + u\cdot(x - x_u) \text{ for all } x \in [0,M]$$

passes through the point $(x_u, G(x_u))$, has slope u, and is subtangent to $G(\cdot)$ at x_u. This implies

$$L(0) \leq G(0) = H(0)$$

$$L(x_u) = G(x_u) \geq H(x_u)$$

$$L(M) \leq G(M) = H(M).$$

Continuity of $H(\cdot)$ implies that there exist values $x' \in [0,x_u]$ and $x'' \in [x_u,M]$ such that $L(x') = H(x')$ and $L(x'') = H(x'')$, in which case convexity of $H(\cdot)$ implies

$$L(x) \leq H(x) \quad \text{for } x \in [0,x']$$

$$L(x) \geq H(x) \quad \text{for } x \in [x',x'']$$

$$L(x) \leq H(x) \quad \text{for } x \in [x'',M].$$

Define $H_{+1}(\cdot)$ as the pointwise maximum of the functions $L(\cdot)$ and $H(\cdot)$:

$$H_{+1}(x) \equiv H(x) \quad \text{for } x \in [0,x']$$

$$H_{+1}(x) \equiv L(x) \quad \text{for } x \in [x',x'']$$

$$H_{+1}(x) \equiv H(x) \quad \text{for } x \in [x'',M]$$

(so $H_{+1}(x_u) = G(x_u)$). Because $H_{+1}(\cdot)$ is convex, its right derivative $H_{+1}(\cdot)$ over $[0,M)$ (with $H_{+1}(M) = 1$) is nondecreasing, right continuous, and satisfies

$$H_{+1}(x) \equiv H(x) \geq 0 \quad \text{for } x \in [0,x')$$

$$H_{+1}(x) \equiv u \quad \text{for } x \in [x',x'')$$

$$H_{+1}(x) \equiv H(x) \leq 1 \quad \text{for } x \in [x'',M]$$

thus $H_{+1}(\cdot)$ is a cumulative distribution function on $[0,M]$, with u-quantile x_u.[27]

Because $\int_0^M H_{+1}(\omega)d\omega = H_{+1}(M) = H(M) = \int_0^M H(\omega)d\omega$, the distribution $H_{+1}(\cdot)$ has the same mean as $H(\cdot)$ (see Note 9), and hence the same mean as $G(\cdot)$. Because the above identities imply

$$H_{+1}(b) - H_{+1}(a) \equiv H(b) - H(a) \quad \text{for all } (a, b] \subset (-\infty, x')$$

$$H_{+1}(b) - H_{+1}(a) \leq H(b) - H(a) \quad \text{for all } (a, b] \subset (x', x'')$$

$$H_{+1}(b) - H_{+1}(a) \equiv H(b) - H(a) \quad \text{for all } (a, b] \subset (x'', +\infty).$$

it follows that $H_{+1}(\cdot)$ differs from $H(\cdot)$ by a single mean preserving spread. In this case, the spread removes all of the probability mass from the open interval (x',x'') (because $H_{+1}(\cdot)$ is constant on $[x',x'')$), and shifts it to the boundary points x' and x''. Because

$$\int_0^x H_{+1}(\omega)d\omega \equiv H_{+1}(x) \equiv \max\{L(x), H(x)\} \leq G(x)$$

$$\equiv \int_0^x G(\omega)d\omega \quad \text{for all } x \in [0,M)$$

it follows that $H_{+1}(\cdot)$ also satisfies the integral condition with respect to $G(\cdot)$.

In preparation for Step 3, we show that x_u will remain a u-quantile of any *subsequent* such spread of the distribution $H_{+1}(\cdot)$: Given any subsequent probability level $u_{+1} \gtrless u$ and observing that the integrated cumulative of $H_{+1}(\cdot)$ is $H_{+1}(\cdot)$, repeating the procedure generates a new convex function $H_{+2}(\cdot)$ lying on or above $H_{+1}(\cdot)$ (and on or below $G(\cdot)$), with $H_{+1}(x_u) = H_{+2}(x_u) = G(x_u)$. This implies the right derivatives of $H_{+1}(\cdot)$ and $H_{+2}(\cdot)$, namely $H_{+1}(\cdot)$ and $H_{+2}(\cdot)$, satisfy $H_{+2}(x_u-) \leq H_{+1}(x_u-) \leq u = H_{+1}(x_u) \leq H_{+2}(x_u)$, so that x_u is a u-quantile of $H_{+2}(\cdot)$.

Step 3 (construction of the sequence and proof of convergence). We construct the sequence $H_0(\cdot), H_1(\cdot), H_2(\cdot), \ldots$ by setting $H_0(\cdot) = F(\cdot)$ and generating Step 2 type spreads for the successive u values:

$$u_1, u_2, u_3, \ldots = \underbrace{\tfrac{1}{2}}_{group\ 1}\ \underbrace{\tfrac{1}{4}, \tfrac{3}{4},}_{group\ 2}\ \underbrace{\tfrac{1}{8}, \tfrac{3}{8}, \tfrac{5}{8}, \tfrac{7}{8}}_{group\ 3}\ \underbrace{\tfrac{1}{16}, \tfrac{3}{16}, \tfrac{5}{16}, \tfrac{7}{16}, \tfrac{9}{16}, \ldots, \ldots}_{group\ 4}, \ldots$$

For $i > 2^n$, $H_i(\cdot)$ follows all of the "group 1" through "group n" spreads, so by the last paragraph of Step 2, $H_i(\cdot)$ and $G(\cdot)$ will have a common u-quantile for $u = 1/2^n, 2/2^n, 3/2^n, \ldots, (2^n - 1)/2^n$. By Step 1, this implies $|H_i(x) - G(x)| \leq 1/2^n$ for all $x \in [0,M]$, which establishes that the sequence $\{H_i(\cdot)\}$ converges weakly (in fact, converges uniformly) to $G(\cdot)$. ■

Proof of Theorem 2. The proof has two steps. Step 1 defines several auxiliary functions and establishes some relationships between them. Step 2 takes a random variable $\tilde{x}$ with distribution $F(\cdot)$, constructs a noise variable $\tilde{\varepsilon}$ whose expected value given x is identically zero, and shows that $\tilde{y} = \tilde{x} + \tilde{\varepsilon}$ has distribution $G(\cdot)$.

Step 1 (preliminaries). For each probability $p \in [0,1]$, define the inverse cumulative (i.e., minimum quantile) functions $F^{-1}(p) = \min\{x \mid F(x) \geq p\}$ and $G^{-1}(p) = \min\{x \mid G(x) \geq p\}$. Define $[F^{-1}(\cdot) - G^{-1}(\cdot)]^{+}$ and $[F^{-1}(\cdot) - G^{-1}(\cdot)]^{-}$ as the absolute values of the positive and the negative parts of $F^{-1}(\cdot) - G^{-1}(\cdot)$, and define $P^{+}, P^{-}, P^{=} \subseteq [0, 1]$ as the sets over which the left continuous function $F^{-1}(\cdot) - G^{-1}(\cdot)$ is positive, negative, and zero.

For each $p \in [0,1]$, define the horizontal "+" and "−" areas up to p by

$$A^{+}(p) = \int_0^p [F^{-1}(s) - G^{-1}(s)]^{+}\, ds \qquad A^{-}(p) = \int_0^p [F^{-1}(s) - G^{-1}(s)]^{-}\, ds.$$

$A^{+}(\cdot)$ and $A^{-}(\cdot)$ are nondecreasing over [0,1], and strictly increasing over P^{+} and P^{-}, respectively. The equality of the means of $F(\cdot)$ and $G(\cdot)$ implies $A^{+}(1) = A^{-}(1) \overset{\text{def}}{=} A_{max}$.

For each area $a \in [0, A_{max}]$, define $u(a)$ and $v(a)$ as the smallest probabilities that solve

$$A^{+}(u(a)) = a \quad \text{and} \quad A^{-}(v(a)) = a;$$

$u(\cdot)$ and $v(\cdot)$ are strictly increasing functions from $[0, A_{max}]$ into $P^{+} \cup P^{=}$ and $P^{-} \cup P^{=}$, respectively, and the integral condition (3) implies $u(a) \leq v(a)$ for all a. The functions $u(\cdot)$ and $A^{+}(\cdot)$, as well as $v(\cdot)$ and $A^{-}(\cdot)$, are pseudo-inverses in that, in addition to the previous equations,

$$u(A^{+}(p)) = p \quad \text{for} \quad p \in P^{+} \qquad v(A^{-}(p)) = p \quad \text{for} \quad p \in P^{-}.$$

Define $A(p) = A^{+}(p)$ for $p \in P^{+}$ and $A(p) = A^{-}(p)$ for $p \in P^{-}$. Because $A^{+}(\cdot)$ and $A^{-}(\cdot)$ are absolutely continuous and nondecreasing, a double application of the change of variable theorem[28] yields that for any $a \in [0, A_{max}]$ and any integrable function $k(\cdot)$

$$\int_{\substack{p \leq u(a) \\ p \in P^{+}}} k(A(p))[F^{-1}(p) - G^{-1}(p)]dp = \int_{p \leq u(a)} k(A^{+}(p))[F^{-1}(p) - G^{-1}(p)]^{+}dp =$$

$$\int_{\alpha \leq A^{+}(u(a))} k(\alpha)d\alpha = \int_{\alpha \leq A^{-}(v(a))} k(\alpha)d\alpha = \int_{p \leq v(a)} k(A^{-}(p))[F^{-1}(p) - G^{-1}(p)]^{-}dp =$$

$$\int_{\substack{p \leq v(a) \\ p \in P^{-}}} k(A(p))[G^{-1}(p) - F^{-1}(p)]dp.$$

Step 2 (construction of the noise variable). Let $\tilde{x}$ be a random variable with distribution $F(\cdot)$. The distribution of noise at each outcome value x of $\tilde{x}$ will be based on the corresponding value(s) of its induced *cumulative probability variable* $\tilde{p}$, which equals $F(x)$ if x is a continuity point of $F(\cdot)$, and has a uniform distribution over $[F(x-), F(x)]$ if x is a discontinuity point of $F(\cdot)$. For every x, $F^{-1}(\cdot)$ maps the realized value(s) of $\tilde{p}$ back to x, so that we have $\tilde{x} \equiv F^{-1}(\tilde{p})$. A standard result is that $\tilde{p}$ is uniform over [0, 1].[29] For each $p \in P^+ \cup P^-$ define the probabilities[30]

$$\pi(p) = \frac{G^{-1}(v(A(p))) - F^{-1}(p)}{G^{-1}(v(A(p))) - G^{-1}(u(A(p)))} \quad \text{and}$$

$$1 - \pi(p) = \frac{F^{-1}(p) - G^{-1}(u(A(p)))}{G^{-1}(v(A(p))) - G^{-1}(u(A(p)))}.$$

Applying the final equation of Step 1 to the function[31]

$$k(\alpha) = \begin{cases} \dfrac{1}{G^{-1}(v(\alpha)) - G^{-1}(u(\alpha))} & \text{if } u(\alpha) \in P^+ \text{ or } v(\alpha) \in P^- \\ 0 & \text{if } u(\alpha) \in P^= \text{ and } v(\alpha) \in P^= \end{cases}$$

and replacing $G^{-1}(p)$ by $G^{-1}(u(A(p)))$ for $p \in P^+$, and by $G^{-1}(v(A(p)))$ for $p \in P^-$, yields

$$\int\limits_{\substack{p \le u(a) \\ p \in P^+}} (1 - \pi(p))dp = \int\limits_{\substack{p \le v(a) \\ p \in P^-}} \pi(p)dp \quad \text{for each } a \in [0, A_{max}].$$

For each $p \in P^+ \cup P^-$, define the noise variable

$$\tilde{\varepsilon}(p) = \begin{cases} G^{-1}(u(A(p))) - F^{-1}(p) \text{ with probability } \pi(p) \\ G^{-1}(v(A(p))) - F^{-1}(p) \text{ with probability } 1 - \pi(p) \end{cases}$$

with $\tilde{\varepsilon}(p) \equiv 0$ for $p \in P^=$. Because $E[\tilde{\varepsilon}(p)] = 0$ for all p, $E[\tilde{\varepsilon} \mid x] = E[\tilde{\varepsilon}(\tilde{p}(x))] = 0$ for all $x \in [0,M]$.

Define $\tilde{y}(p) = x(p) + \tilde{\varepsilon}(p) = F^{-1}(p) + \tilde{\varepsilon}(p)$ for all $p \in [0,1]$, so that for $p \in P^+ \cup P^-$

$$\tilde{y}(p) = \begin{cases} G^{-1}(u(A(p))) \text{ with probability } \pi(p) \\ G^{-1}(v(A(p))) \text{ with probability } 1 - \pi(p) \end{cases}$$

and $\tilde{y}(p) = x(p) = F^{-1}(p) = G^{-1}(p)$ for $p \in P^=$. The results of Step 1 imply the following relations:

$$\text{for } p \in P^+: \begin{cases} G^{-1}(u(A(p))) = G^{-1}(u(A^+(p))) \le y \Leftrightarrow G^{-1}(p) \le y \Leftrightarrow p \le G(y), \\ \\ G^{-1}(v(A(p))) = G^{-1}(v(A^+(p))) \le y \Leftrightarrow v(A^+(p)) \le G(y) \\ \qquad \Leftrightarrow u(A^-(v(A^+(p)))) \le u(A^-(G(y))) \Leftrightarrow p \le u(A^-(G(y))), \end{cases}$$

$$\text{for } p \in P^-: \begin{cases} G^{-1}(u(A(p))) = G^{-1}(u(A^-(p))) \leq y \Leftrightarrow u(A^-(p)) \leq G(y) \\ \qquad \Leftrightarrow v(A^+(u(A^-(p)))) \leq v(A^+(G(y))) \Leftrightarrow p \leq v(A^+(G(y))), \\ \\ G^{-1}(v(A(p))) = G^{-1}(v(A^-(p))) \leq y \Leftrightarrow G^{-1}(p) \leq y \Leftrightarrow p \leq G(y), \end{cases}$$

$$\text{for } p \in P^=: G^{-1}(p) \leq y \Leftrightarrow p \leq G(y).$$

Because $\tilde{p}$ is uniform over [0,1], it follows that for any $y \in [0,M]$:

$$Pr\{\tilde{y} \leq y\} = I_1 + I_2 + I_3 + I_4 + I_5 = G(y)$$

where

$$I_1 = \int_{\substack{G^{-1}(u(A(p)))\leq y \\ p\in P^+}} \pi(p)dp = \int_{\substack{p\leq G(y) \\ p\in P^+}} \pi(p)dp$$

$$I_2 = \int_{\substack{G^{-1}(v(A(p)))\leq y \\ p\in P^+}} (1-\pi(p))dp = \int_{\substack{p\leq u(A^-(G(y))) \\ p\in P^+}} (1-\pi(p))dp = \int_{\substack{p\leq v(A^-(G(y))) \\ p\in P^-}} \pi(p)dp$$
$$= \int_{\substack{p\leq G(y) \\ p\in P^-}} \pi(p)dp$$

$$I_3 = \int_{\substack{G^{-1}(u(A(p)))\leq y \\ p\in P^-}} \pi(p)dp = \int_{\substack{p\leq v(A^+(G(y))) \\ p\in P^-}} \pi(p)dp = \int_{\substack{p\leq u(A^+(G(y))) \\ p\in P^+}} (1-\pi(p|M))dp$$
$$= \int_{\substack{p\leq G(y) \\ p\in P^+}} (1-\pi(p))dp$$

$$I_4 = \int_{\substack{G^{-1}(v(A(p)))\leq y \\ p\in P^-}} (1-\pi(p))dp = \int_{\substack{p\leq G(y) \\ p\in P^-}} (1-\pi(p))dp$$

$$I_5 = \int_{\substack{G^{-1}(p)\leq y \\ p\in P^=}} 1dp = \int_{\substack{p\leq G(y) \\ p\in P^=}} 1dp$$

which implies that $\tilde{y}$ has distribution $G(\cdot)$ ■

Proof of (c′) ⇒ (a′) in Theorem 3′. We first show that if $F(\cdot)$ and $G(\cdot)$ satisfy the SSD integral condition (c')/(9), then there exists a distribution $H_1(\cdot)$ that differs from $F(\cdot)$ by a (possibly null) leftward shift of probability mass, and that satisfies the equal-mean integral condition (3) with respect to $G(\cdot)$. Define the function $W(x) \equiv \int_0^M G(\omega)d\omega + x - M$, which has unit slope and passes through the point $(M, G(M))$. Because

$$W(0) = \int_0^M G(\omega)d\omega - M \leq 0 = F(0) \quad \text{and} \quad W(M) = G(M) \geq F(M)$$

there exists some $W \in [0,M]$ such that $W(W) = F(W)$. Define the distribution

$$H_1(\cdot) = \begin{cases} F(\cdot) & \text{over } [0,W) \\ 1 & \text{over } [W,M] \end{cases}$$

$H_1(\cdot)$ is seen to differ from $F(\cdot)$ by a (possibly null) leftward shift of probability mass. In addition:

$$\int_0^x [G(\omega) - H_1(\omega)]d\omega$$

$$= \begin{cases} \int_0^M G(\omega)d\omega - \int_W^M H_1(\omega)d\omega - \int_0^W H_1(\omega)d\omega = & \text{for } x = M \\ \int_0^M G(\omega)d\omega - (M - W) - F(W) = W(W) - F(W) = 0 & \\ \int_0^x [G(\omega) - F(\omega)]d\omega \geq 0 & \text{for } x \in [0,W) \\ -\int_x^M [G(\omega) - H_1(\omega)]d\omega = -\int_x^M [G(\omega) - 1]d\omega \geq 0 & \text{for } x \in [W,M) \end{cases}$$

so $H_1(\cdot)$ and $G(\cdot)$ satisfy condition (3). We then can use Theorem 1 to construct a sequence of mean preserving spreads, yielding distributions $H_2(\cdot), H_3(\cdot) \ldots$ which converge to $G(\cdot)$.■

Proof of (c′) ⇒ (b′) in Theorem 3′. Assume $F(\cdot)$ and $G(\cdot)$ satisfy the SSD integral condition (c′) (condition (9)). Define $\bar{u}$ as the smallest probability that satisfies

$$\int_0^{\bar{u}} [F^{-1}(s) - G^{-1}(s)]^+ ds \geq \int_0^1 [F^{-1}(s) - G^{-1}(s)]^- ds$$

Note that $\bar{u} < 1$ unless $F(\cdot)$ and $G(\cdot)$ have the same mean. Define $P^{++} = \{p \in [\bar{u}, 1] | F^{-1}(p) - G^{-1}(p) > 0\}$, and define $\hat{G}(\cdot)$ as the distribution over $[0,M]$ whose inverse takes the form[32]

$$\hat{G}^{-1}(p) = \begin{cases} F^{-1}(p) \text{ for } p \in P^{++} \\ G^{-1}(p) \text{ for } p \notin P^{++} \end{cases} \quad \text{or equivalently} \quad \begin{cases} \max(F^{-1}(p), G^{-1}(p)) & \text{for } p > \bar{u} \\ G^{-1}(p) & \text{for } p \leq \bar{u} \end{cases}$$

In Figure 6, P^{++} is the set of probability levels corresponding to the "+" region(s) at or above $\bar{u}$, and the graph of $\hat{G}(\cdot)$ corresponds to $G(\cdot)$ at or below $\bar{u}$, and to the right envelope of $F(\cdot)$ and $G(\cdot)$ above $\bar{u}$.

Because the total "+" and "−" areas between $\hat{G}(\cdot)$ and $F(\cdot)$ are equal, and the "+" area up to each height is always greater than or equal to the "−" area, $\hat{G}(\cdot)$ and $F(\cdot)$ satisfy the equal-mean integral condition (3). Thus, the construction of Theorem 2 yields a family of zero-mean noise variables $\{\tilde{\upsilon}(p) \mid p \in [0,1]\}$ such that if $\tilde{p}$ is uniform over [0,1], and $\tilde{y}(p) \equiv x(p) + \tilde{\upsilon}(p) \equiv F^{-1}(p) + \tilde{\upsilon}(p)$, then $\tilde{y}(\tilde{p})$ has distribution $\hat{G}(\cdot)$. Because $\hat{G}^{-1}(\cdot)$ and $F^{-1}(\cdot)$ coincide over P^{++}, the construction will imply $\tilde{\upsilon}(p) \equiv 0$ for $p \in P^{++}$, so that $\tilde{y}(p) \equiv F^{-1}(p) \equiv \hat{G}^{-1}(p)$ for all $p \in P^{++}$.

Define the nonpositive mean SSD noise variable $\tilde{\varepsilon}(p)$, hence the variable $\tilde{z}(p) \equiv x(p) + \tilde{\varepsilon}(p) \equiv F^{-1}(p) + \tilde{\varepsilon}(p)$, by

$$\tilde{\varepsilon}(p) = \begin{cases} G^{-1}(p) - F^{-1}(p) & \text{for } p \in P^{++} \\ \tilde{\upsilon}(p) & \text{for } p \notin P^{++} \end{cases} \quad \text{so that} \quad \tilde{z}(p) = \begin{cases} G^{-1}(p) & \text{for } p \in P^{++} \\ \tilde{y}(p) & \text{for } p \notin P^{++} \end{cases}.$$

Because $\tilde{y}(\tilde{p})$ and $\hat{G}^{-1}(\tilde{p})$ both have unconditional distribution $\hat{G}(\cdot)$, we have that, for all $z \in [0,M]$:

$$\int_{p \notin P^{++}} Pr\{\tilde{y}(p) \le z\}dp = Pr\{\tilde{y}(\tilde{p}) \le z\} - \int_{p \in P^{++}} Pr\{\tilde{y}(p) \le z\}dp$$
$$= Pr\{\hat{G}^{-1}(\tilde{p}) \le z\} - \int_{p \in P^{++}} Pr\{\hat{G}^{-1}(p) \le z\}dp =$$
$$\int_{p \notin P^{++}} Pr\{\hat{G}^{-1}(p) \le z\}dp = \int_{p \notin P^{++}} Pr\{G^{-1}(p) \le z\}dp$$

So for all $z \in [0,M]$:

$$Pr\{\tilde{z}(\tilde{p}) \le z\} = \int_{p \in P^{++}} Pr\{\tilde{z}(p) \le z\}dp + \int_{p \notin P^{++}} Pr\{\tilde{z}(p) \le z\}dp$$
$$= \int_{p \in P^{++}} Pr\{G^{-1}(p) \le z\}dp + \int_{p \notin P^{++}} Pr\{\tilde{y}(p) \le z\}dp =$$
$$\int_{p \in P^{++}} Pr\{G^{-1}(p) \le z\}dp + \int_{p \notin P^{++}} Pr\{G^{-1}(p) \le z\}dp = Pr\{G^{-1}(\tilde{p}) \le z\} = G(z)$$

which implies that $\tilde{z}(\tilde{p})$ has unconditional distribution $G(\cdot)$.[33] ■

Acknowledgments

We would like to thank Persi Diaconis, Ted Groves, Josef Hadar, Walter Heller, David Kreps, Jack Meyer, Robert Nau, Joel Sobel, Shaun Wang, and Hal White for helpful comments on this material. Machina is grateful to the National Science Foundation (Grant SES-9209012) for financial support. All errors are our own responsibility.

Notes

1. See Sherman (1951), Rothschild and Stiglitz (1972), Fishburn and Vickson (1978), Schmeidler (1979), Bawa (1982), and Levy (1992) for literature summaries and/or bibliographies, as well as the more recent work cited in Section 6 below.
2. Rothschild and Stiglitz (1970), Theorems 1b and 2c, respectively. Unless stated otherwise, convergence of probability distributions ($\rightarrow$) refers to weak convergence (e.g., Billingsley (1968; 1986, Section 25)).
3. Rothschild and Stiglitz use the interval [0,1].
4. The Cantor function $C(\cdot)$ over [0,1] is continuous, nondecreasing, and satisfies $C(0) = 0$ and $C(1) = 1$, so it is a valid cumulative distribution function that assigns probability one to the interval [0,1]. But because $C(\cdot)$ is continuous it has no mass points, and because $C'(\cdot) = 0$ almost everywhere on [0,1], $C(\cdot)$ cannot be represented as the integral of any density function (e.g., Kolmogorov and Fomin (1970, pp. 334-336), Feller (1971, pp. 35-36)).
5. For each $x \in [0,1]$, let $\tilde{\varepsilon}(x)$ be a $(1 - x){:}x$ chance of $-x{:}(1 - x)$. Clearly $\mathrm{E}[\tilde{\varepsilon}(x) \mid x] \equiv 0$, and if $\tilde{x}$ has the distribution $C(\cdot)$ (which is symmetric about the outcome 1/2) then $\tilde{x} + \tilde{\varepsilon}$ is a 50:50 chance of 0:1.
6. Throughout this article we assume the natural extensions $F(x) \equiv 0$ on $(-\infty,0)$ and $F(x) \equiv 1$ on $(M, +\infty)$ for all cumulative distributions functions $F(\cdot)$ defined "over $[0,M]$."
7. Because the Borel subsets of $[0,M]$ (or $(-\infty,+\infty)$) can be countably generated from the set of half-open intervals of the form $(a,b]$, the following conditions can equivalently be written as:

   $G(b) - G(a) \ \geq / \leq / \geq \ F(b) - F(a)$ for all $(a,b] \subset (-\infty,x')/(x',x'')/(x'', +\infty)$.
8. Like the R-S spreads (1) and (2), this definition of a mean preserving spread implies that $F(\cdot)$ and $G(\cdot)$ differ by a "single crossing", that is, there exists some x_0 such that $G(x) \geq F(x)$ for $x < x_0$ and $G(x) \leq F(x)$ for $x \geq x_0$. To see that the single-crossing condition is strictly more general, however, note that the distributions assigning probabilities {0 , 2/10 , 0 , 8/10 , 0} and {3/10 , 0 , 1/10 , 0 , 6/10} to the outcomes {1 , 2 , 3 , 4 , 5} differ by a mean preserving single crossing, but not by a mean preserving spread.
9. To see that the condition $\int_0^M [G(\omega) - F(\omega)]d\omega = 0$ is equivalent to the condition that $F(\cdot)$ and $G(\cdot)$ have the same mean, apply the integration-by-parts formula for Stieltjes integrals (e.g., Billingsley (1986, p. 240), Feller (1971, p.150)) to obtain that it is equivalent to $0 = -\int_0^M \omega[dG(\omega) - dF(\omega)] = -\int_0^M \omega dG(\omega) + \int_0^M \omega dF(\omega)$. (See also Hanoch and Levy (1969, Lemma 1)).
10. As in Rothschild and Stiglitz (1970), this follows because every mean preserving spread satisfies condition (3), and this condition is transitive over distributions. This same argument and result hold for mean preserving single crossings.
11. Full proofs of theorems 1 and 2 are given in the Appendix.
12. By Note 8, we can replace "mean preserving spread" by "mean preserving single crossing" in the statement of this theorem.
13. Cf. Note 18.
14. See, for example, Feller (1971, p. 267, Section VIII.6), who says the theorem is usually ascribed to Helly.
15. The joint distributions of $(\tilde{x},\tilde{\varepsilon})$ and $(\tilde{x},\tilde{z})$ can thus be derived directly from $F(\cdot)$ and the distributions of $\tilde{\varepsilon}(x)$.
16. If $x_u = y_u$ (whether or not a crossing occurs) we define $\tilde{\varepsilon}(x_u) \equiv 0$, and similarly if $x_v = y_v$.

17. That is, first order as the difference between the primed and nonprimed values of each variable goes to zero. For purposes of this informal version of the argument, we assume that the slopes of $F(\cdot)$ and $G(\cdot)$ are bounded away from both zero and infinity, so that all these differences go to zero at a common rate.
18. Each of the successive slices in the construction of Theorem 1 above can be equivalently represented as the addition of a zero-conditional-mean noise variable, so that their sum provides a noise variable linking $F(\cdot)$ and $G(\cdot)$, as noted independently by Robert Nau. The partial sums of this series form a bounded martingale, which implies both convergence and the zero-conditional-mean property of the infinite sum. Persi Diaconis tells us that this is a case of "balayage" (Choquet (1969, Index)). The noise variable in Theorem 2 differs from this and is constructed in a single step (equation (5)).
19. See Note 2 for the specific type of convergence used here. As in Theorem 1, "mean preserving spread" can be replaced by "mean preserving single crossing" (see Notes 8, 10, and 12).
20. A standard restatement of (d), more suited to the case when x denotes income or wealth, is "$F(\cdot)$ and $G(\cdot)$ have the same mean, and $\int_0^M U(x)dG(x) \leq \int_0^M U(x)dF(x)$ for every *increasing* concave function $U(\cdot)$ over $[0,M]$."
21. As with mean preserving spreads: (a) this condition can be written as $G(b) - G(a) \geq$ (resp. $\leq$) $F(b) - F(a)$ for all $(a, b] \subset (-\infty,x')$ (resp. $(x', +\infty)$); (b) the probability assigned to the point x' itself can rise, remain unchanged, or drop; and (c) each distribution differs from itself by *null* rightward and leftward shifts of probability mass.
22. Proofs of the extended constructions appear in the Appendix, as parts of the proof of Theorem 3′.
23. Otherwise, the equal-mean constructions apply directly.
24. Family loyalty and royal protocol require one author to point out that, in data analysis, replacing the tail of a distribution by equal mass at the cutoff point is termed "Winsorization." (The other author's middle name is Joseph.)
25. An interval $[x_{i/n}, x_{(i+1)/n})$ will be empty if $x_{i/n} = x_{(i+1)/n}$, which can happen if $H(\cdot)$ and $G(\cdot)$ have a jump of size $1/n$ or greater at that outcome level. This will have no bearing on the argument.
26. In this step we invoke the following results from convex analysis: If $H(\cdot)$ is a convex function, then its right derivative exists at each point, is nondecreasing and right continuous and integrates back to $H(\cdot)$. Conversely, if $H(\cdot)$ is nondecreasing and right continuous, then its integral $H(\cdot)$ is convex, with left and right derivatives at x given by $H(x-)$ and $H(x)$. Corresponding results hold for left derivatives. These results are implied in our context by Rockafellar (1970, Theorems 23.1, 24.1, 24.2, Corollary 24.2.1).
27. If $x' = x''$ the interval $[x',x'')$ is empty, so $H_{+1}(\cdot) = H(\cdot)$. This has no bearing on the argument, which will still demonstrate (somewhat redundantly) that $H_{+1}(\cdot)$ is a (null) mean preserving spread of $H(\cdot)$, that $H_{+1}(\cdot)$ satisfies the integral condition with respect to $G(\cdot)$, and (not so redundantly) that x_u is a u-quantile of $H_{+1}(\cdot) = H(\cdot)$.
28. The change of variable theorem invoked here (Klambauer, 1973, p. 168, Proposition 27) does not require differentiability of $A^+(\cdot)$ or $A^-(\cdot)$, merely that each be the integral of some nonnegative integrable function.
29. For any value $p \in [0,1]$, if $x = F^{-1}(p)$ is a continuity point of $F(\cdot)$, then $Pr\{\tilde{p} \leq p\} = Pr\{F(\tilde{x}) \leq p\} = Pr\{\tilde{x} \leq F^{-1}(p)\} = F(F^{-1}(p)) = p$. If $x = F^{-1}(p)$ is a discontinuity point of $F(\cdot)$, then $Pr\{\tilde{p} \leq p\} = Pr\{\tilde{p} < F(x-)\} + Pr\{F(x-) \leq \tilde{p} \leq p\} = Pr\{F(\tilde{x}) < F(x-)\} + Pr\{\tilde{x} = x\}$. $Pr\{F(x-) \leq \tilde{p} \leq p \mid \tilde{x} = x\} = Pr\{\tilde{x} < x\} + [F(x) - F(x-)]\cdot[(p - F(x-))/(F(x) - F(x-))] = F(x-) + (p - F(x-)) = p$. Thus, $Pr\{\tilde{p} \leq p\} = p$, for all $p \in [0,1]$.
30. The condition $p \in P^+$ (resp. $p \in P^-$) implies $G^{-1}(u(A(p))) <$ (resp. $\leq$) $F^{-1}(u(A(p))) \leq F^{-1}(p) \leq F^{-1}(v(A(p))) \leq$ (resp. $<$) $G^{-1}(v(A(p))$, which implies $\pi(p)$ and $1 - \pi(p)$ are in $[0,1]$ for all $p \in P^+ \cup P^-$.
31. To see that the denominator of $k(\alpha)$ is nonzero when $u(\alpha) \in P^+$ (resp. $v(\alpha) \in P^-$), note that these conditions imply $G^{-1}(u(\alpha)) <$ (resp. $\leq$) $F^{-1}(u(\alpha)) \leq F^{-1}(v(\alpha)) \leq$ (resp. $<$) $G^{-1}(v(\alpha))$.
32. Given any nondecreasing, left continuous function $\Phi(\cdot)$ from $[0,1]$ into $[0,M]$ with $\Phi(0) = 0$ (such as the function specified in the display), the function $\hat{G}(x) \equiv \max\{p \in [0,1] \mid \Phi(p) \leq x\}$ will be a distribution function over $[0,M]$ with inverse $\Phi(\cdot)$.
33. An alternative, two-component SSD noise variable can be derived as follows: Assume $F(\cdot)$ and $G(\cdot)$ satisfy (c'), let $\tilde{x}$ have distribution $F(\cdot)$, and define W and $H_1(\cdot)$ as in the proof of $(c') \Rightarrow (a')$. Defining $\alpha(x) \equiv \min(0, W - x) \leq 0$, it follows that $\tilde{z} \equiv \tilde{x} + \alpha(\tilde{x}) \equiv \min(\tilde{x}, W)$ has distribution $H_1(\cdot)$. Because $H_1(\cdot)$ and $G(\cdot)$ satisfy condition (3), Theorem 2 yields a zero-mean variable $\tilde{\eta}(z)$, such that if $\tilde{z}$ has distribution $H_1(\cdot)$ then

$\tilde{z} + \tilde{\eta}(\tilde{z})$ has distribution $G(\cdot)$. Define the nonpositive-mean SSD noise variable $\tilde{\varepsilon}(x) \equiv \alpha(x) + \tilde{\eta}(x + \alpha(x))$. Because $\tilde{x} + \alpha(\tilde{x})$ has distribution $H_1(\cdot)$, we have that $\tilde{x} + \tilde{\varepsilon}(\tilde{x}) \equiv \tilde{x} + \alpha(\tilde{x}) + \tilde{\eta}(x + \alpha(\tilde{x})) \equiv \tilde{z} + \tilde{\eta}(\tilde{z})$ has distribution $G(\cdot)$.

References

Bawa, Vijay. (1982). "Stochastic Dominance: A Research Bibliography," *Management Science* 28, 698–712.

Billingsley, Patrick. (1968). *Convergence of Probability Measures*. New York: John Wiley and Sons.

Billingsley, Patrick. (1986). *Probability and Measure* (2nd ed.). New York: John Wiley and Sons.

Blackwell, David. (1951). "Comparison of Experiments." In Jerzy Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press.

Blackwell, David. (1953). "Equivalent Comparisons of Experiments." *Annals of Mathematical Statistics* 24, 265–272.

Choquet, G. (1969). *Lectures on Analysis* (J. Marsden, T. Lance, and S. Gelbert, eds.). New York: W. A. Benjamin.

Feller, William. (1971). *An Introduction to Probability Theory and Its Applications* (Vol. II, 2nd ed.). New York: John Wiley and Sons.

Fishburn, Peter, and Raymond Vickson. (1978). "Theoretical Foundations of Stochastic Dominance." In G. A. Whitmore and M. Chapman Findlay (eds.), *Stochastic Dominance*: *An Approach to Decision Making Under Risk*. Lexington, MA: Heath.

Hadar, Josef, and William Russell. (1969). "Rules for Ordering Uncertain Prospects," *American Economic Review* 59, 25–34.

Hadar, Josef, and Tae Kun Seo. (1990). "The Effects of Shifts in a Return Distribution on Optimal Portfolios," *International Economic Review* 31, 721–736.

Hanoch, Giora, and Haim Levy. (1969). "The Efficiency Analysis of Choices Involving Risk," *Review of Economic Studies* 36, 335–346.

Hardy, G. H., John Littlewood, and Gyorgy Pólya. (1929). "Some Simple Inequalities Satisfied by Convex Functions," *Messenger of Mathematics* 58, 145–152.

Hardy, G. H., J. Littlewood, and G. Pólya. (1934). *Inequalities*. Cambridge, England: Cambridge University Press.

Kaas, R., A. E. van Heerwaarden, and M. J. Goovaerts. (1994). *Ordering of Actuarial Risks*. Brussels: Education Series 1, CAIRE. (Second edition, forthcoming).

Klambauer, Gabriel. (1973). *Real Analysis*. New York: American Elsevier.

Kolmogorov, Andrei, and Sergei Fomin. (1970). *Introductory Real Analysis*. New York: Dover Publications. (English translation by Richard A. Silverman.)

Landsberger, Michael, and Isaac Meilijson. (1990). "A Tale of Two Tales: An Alternative Characterization of Increasing Risk," *Journal of Risk and Uncertainty* 3, 65–82.

Levy, Haim. (1992). "Stochastic Dominance and Expected Utility: Survey and Analysis," *Management Science* 38, 555–593.

Müller, Alfred. (1996). "Ordering of Risks: A Comparative Study via Stop-Loss Transforms," *Insurance*: *Mathematics and Economics* 17, 215–222.

Pratt, J. (1990). "From Dominance to Noise in One Step," manuscript, Harvard University. Cambridge, MA.

Quirk, James, and Rubin Saposnik. (1962). "Admissibility and Measurable Utility Functions," *Review of Economic Studies* 29, 140–146.

Rockafellar, Ralph. (1970). *Convex Analysis*. Princeton, NJ: Princeton University Press.

Rothschild, Michael, and Joseph Stiglitz. (1970). "Increasing Risk: I. A Definition," *Journal of Economic Theory* 2, 225–243.

Rothschild, Michael, and Joseph Stiglitz. (1971). "Increasing Risk: II. Its Economic Consequences," *Journal of Economic Theory* 3, 66–84.

Rothschild, Michael, and Joseph Stiglitz. (1972). "Addendum to 'Increasing Risk: I. A Definition'," *Journal of Economic Theory* 5, 306.

Scarsini, Marco. (1994). "Comparing Risk and Risk Aversion." In Moshe Shaked and J. George Shanthikumar (eds.), *Stochastic Orders and Their Applications*. San Diego: Academic Press.

Schmeidler, David. (1979). "A Bibliographical Note on a Theorem of Hardy, Littlewood, and Pólya," *Journal of Economic Theory* 20, 125–128.

Sherman, S. (1951). "On a Theorem of Hardy, Littlewood, Pólya, and Blackwell," *Proceedings of the National Academy of Sciences* 37, 826–831; "Errata," *Proc. Nat. Ac. Sci.* 38, 382.

Strassen, V. (1965). "The Existence of Probability Measures with Given Marginals," *Annals of Mathematical Statistics* 36, 423–439.