JOURNAL OF ECONOMIC THEORY 2, 225–243 (1970)

Increasing Risk: I. A Definition*

MICHAEL ROTHSCHILD

Harvard University, Cambridge, Massachusetts 02138

AND

JOSEPH E. STIGLITZ

Cowles Foundation, Yale University, New Haven, Connecticut 06520 and Gonville and Caius College, Cambridge, England

I. INTRODUCTION

This paper attempts to answer the question: When is a random variable Y "more variable" than another random variable X?

Intuition and tradition suggest at least four plausible—and apparently different—answers to this question. These are:

1. *Y is Equal to X Plus Noise*

If we simply add some uncorrelated noise to a random variable, (r.v.), the new r.v. should be riskier[1] than the original. More formally, suppose Y and X are related as follows:

$$Y \underset{d}{=} X + Z, \tag{1.i}$$

where "$\underset{d}{=}$" means "has the same distribution as" and Z is a r.v. with the property that

$$E(Z \mid X) = 0 \qquad \text{for all } X.^2 \tag{1.ii}$$

* The research described in this paper was carried out under grants from the National Science Foundation and the Ford Foundation.

[1] Throughout this paper we shall use the terms more variable, riskier, and more uncertain synonomously.

[2] David Wallace suggested that we investigate this concept of greater riskiness. Arthur Goldberger has pointed out to us that (1.ii) is stronger than lack of correlation as earlier versions of this paper stated.

That is, Y is equal to X plus a disturbance term (noise.) If X and Y are discrete r.v.'s, condition (1) has another natural interpretation. Suppose X is a lottery ticket which pays off a_i with probability p_i ; $\Sigma p_i = 1$. Then, Y is a lottery ticket which pays b_i with probability p_i where b_i is either a payoff of a_i or a lottery ticket whose expected value is a_i. Note that condition (1) implies that X and Y have the same mean.

2. *Every Risk Averter Prefers X to Y*

In the theory of expected utility maximization, a risk averter is defined as a person with a concave utility function. If X and Y have the same mean, but every risk averter prefers X to Y, i.e., if

$$EU(X) \geqslant EU(Y) \qquad \text{for all concave } U \tag{2}$$

then surely it is reasonable to say that X is less risky than Y.[3]

3. *Y Has More Weight in the Tails Than X*

If X and Y have density functions f and g, and if g was obtained from f by taking some of the probability weight from the center of f and adding it to each tail of f in such a way as to leave the mean unchanged, then it seems reasonable to say that Y is more uncertain than X.

4. *Y Has a Greater Variance Than X*

Comparisons of riskiness or uncertainty are commonly restricted to comparisons of variance, largely because of the long history of the use of the variance as a measure of dispersion in statistical theory.

The major result of this paper is that the first three approaches lead to a single definition of greater riskiness, different from that of the fourth approach. We shall demonstrate the equivalence as follows. In Section II, it is shown that the third approach leads to a characterization of increasing uncertainty in terms of the indefinite integrals of differences of cumulative distribution functions (c.d.f.'s). In Section III it is shown that this indefinite integral induces a partial ordering on the set of distribution functions which is equivalent to the partial ordering induced by the first two approaches.

In Section IV we show that this concept of increasing risk is not equivalent to that implied by equating the risk of X with the variance of X. This suggests to us that our concepts lead to a better definition of increasing risk than the standard one.

It is of course impossible to prove that one definition is better than

[3] It might be argued that we should limit our discussion to increasing concave functions. Imposing this restriction would gain nothing and would destroy the symmetry of some of the results. For example, since $U(X) = X$ and $U(X) = -X$ are both concave functions, condition (2) implies that X and Y have the same mean.

another. This fact is not a license for agnosticism or the suspension of judgment. Although there seems to us no question but that our definition is more consistent with the natural meaning of increasing risk than the variance definition, definitions are chosen for their usefulness as well as their consistency. As Tobin has argued, critics of the mean variance approach "owe us more than demonstrations that it rests on restrictive assumptions. They need to show us how a more general and less vulnerable approach will yield the kind of comparative static results that economists are interested in [8]." In the sequel to this paper we show how our definition may be applied to economic and statistical problems.

Before we begin it will be well to establish certain notational conventions. Throughout this paper X and Y will be r.v.'s with c.d.f.'s, F and G, respectively. When they exist, we shall write the density functions of F and G as f and g. In general we shall adhere to the convention that F is less risky than G.

At present our results apply only to c.d.f.'s whose points of increase lie in a bounded interval, and we shall for convenience take that interval to be [0, 1], that is $F(0) = G(0) = 0$ and $F(1) = G(1) = 1$. The extension (and modification) of the results to c.d.f.'s defined on the whole real line is an open question whose resolution requires the solution of a host of delicate convergence problems of little economic interest. $H(x, z)$ is the joint distribution function of the r.v.'s X and Z defined on $[0, 1] \times [-1, 1]$, the cartesian product of [0, 1] and [−1, 1]. We shall use S to refer to the difference of G and F and let T be its indefinite integral, that is, $S(x) = G(x) - F(x)$ and $T(y) = \int_0^y S(x)\,dx$.

II. The Integral Conditions

In this section we give a geometrically motivated definition of what it means for one r.v. to have more weight in the tails than another (Subsections 1 and 2). A definition of "greater risk" should be transitive. An examination of the consequence of this requirement leads to a more general definition which, although less intuitive, is analytically more convenient (Subsections 3 and 4).

1. *Mean Preserving Spreads: Densities*

Let $s(x)$ be a step function defined by

$$s(x) = \begin{cases} \alpha \geqslant 0 & \text{for } a < x < a + t \\ -\alpha \leqslant 0 & \text{for } a + d < x < a + d + t \\ -\beta \leqslant 0 & \text{for } b < x < b + t \\ \beta \geqslant 0 & \text{for } b + e < x < b + e + t \\ 0 & \text{otherwise,} \end{cases} \tag{3.i}$$

where

$$0 \leqslant a \leqslant a+t \leqslant a+d \leqslant a+d+t \\ \leqslant b \leqslant b+t \leqslant b+e \leqslant b+e+t \leqslant 1 \tag{3.ii}$$

and

$$\beta e = \alpha d. \tag{3.iii}$$

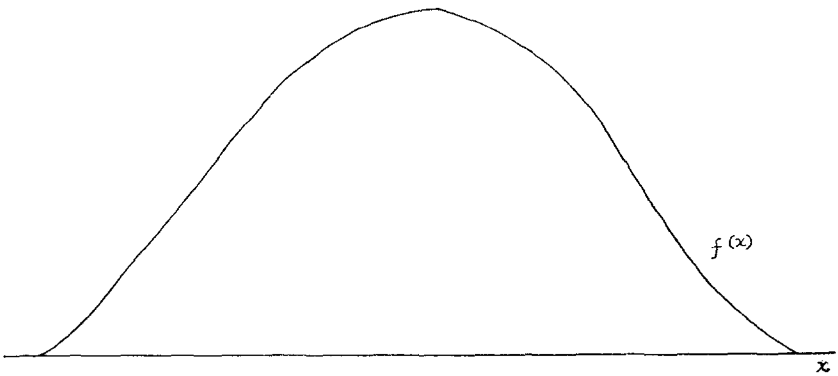

FIGURE 1

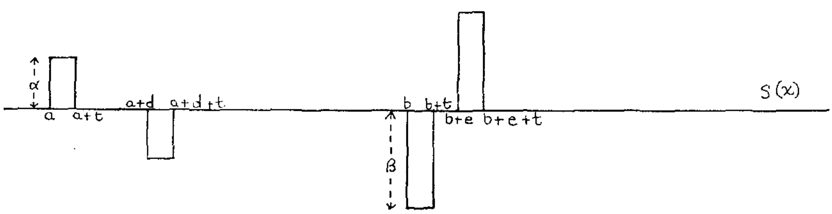

FIGURE 2

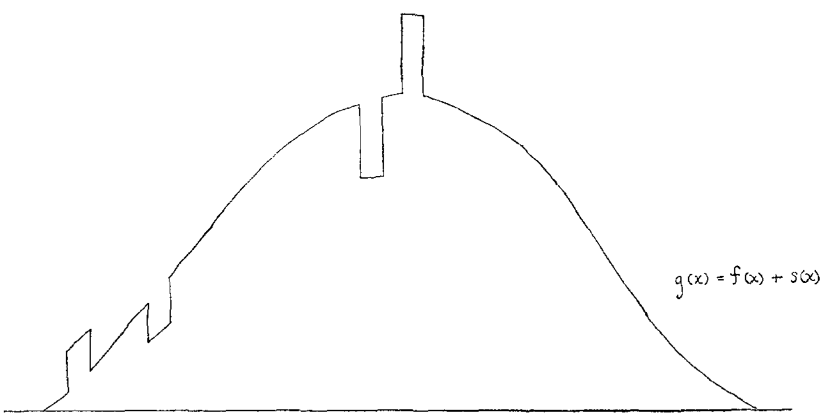

FIGURE 3

Such a function is pictured in Fig. 2. It is easy to verify that $\int_0^1 s(x)\,dx = \int_0^1 xs(x)\,dx = 0$. Thus if f is a density function and if $g = f + s$, then $\int_0^1 g(x)\,dx = \int_0^1 f(x)\,dx + \int_0^1 s(x)\,dx = 1$ and $\int_0^1 xg(x)\,dx = \int_0^1 x(f(x) + s(x))\,dx = \int_0^1 xf(x)\,dx$. It follows then that if $g(x) \geqslant 0$ for all x, g is a density function[4] with the same mean as f. Adding a function like s to f shifts probability weight from the center to the tails. See Figs. 1 and 3. We shall call a function which satisfies conditions (3) a mean preserving spread (MPS) and if f and g are densities and $g - f$ is a MPS we shall say that g differs from f by a single MPS.

2. *Mean Preserving Spreads: Discrete Distributions*

We may define a similar concept for the difference between discrete distributions. Let F and G be the c.d.f.'s of the discrete r.v.'s X and Y. We can describe X and Y completely as follows:

$$\Pr(X = \hat{a}_i) = \hat{f}_i \qquad \text{and} \qquad \Pr(Y = \hat{a}_i) = \hat{g}_i,$$

where $\sum_i \hat{f}_i = \sum_i \hat{g}_i = 1$, and $\{\hat{a}_i\}$ is an increasing sequence of real numbers bounded by 0 and 1. Suppose $\hat{f}_i = \hat{g}_i$ for all but four i, say i_1, i_2, i_3, and i_4 where $i_k < i_{k+1}$. To avoid double subscripts let $a_k = \hat{a}_{i_k}$, $f_k = \hat{f}_{i_k}$, and $g_k = \hat{g}_{i_k}$, and define

$$\gamma_k = g_k - f_k$$

Then if

$$\gamma_1 = -\gamma_2 \geqslant 0 \qquad \text{and} \qquad \gamma_4 = -\gamma_3 \geqslant 0, \tag{4.i}$$

Y has more weight in the tails than X and if

$$\sum_{k=1}^{4} a_k \gamma_k = 0, \tag{4.ii}$$

the means of X and Y will be the same. See Fig. 4. If two discrete r.v.'s X and Y attribute the same weight to all but four points and if their differences satisfy conditions (4) we shall say that Y differs from X by a single MPS.

3. *The Integral Conditions*

If two densities g and f differ by a single MPS, s, the difference of the corresponding c.d.f.'s G and F will be the indefinite integral of s. That is,

[4] That is, if $f(x) \geqslant \alpha$ for $a + d < x < a + d + t$ and $f(x) \geqslant \beta$ for $b < x < b + t$.

$s = g - f$ implies $S = G - F$ where $S(x) = \int_0^x s(u)\, du$. S, which is drawn in Fig. 5, has several interesting properties. The last two of these ((6) and (7) below) will play a crucial role in this paper, and we will refer to them as the integral conditions. First $S(0) = S(1) = 0$. Second, there is a z such that

$$S(x) \geqslant 0 \quad \text{if} \quad x \leqslant z \qquad \text{and} \qquad S(x) \leqslant 0 \quad \text{if} \quad x > z. \tag{5}$$

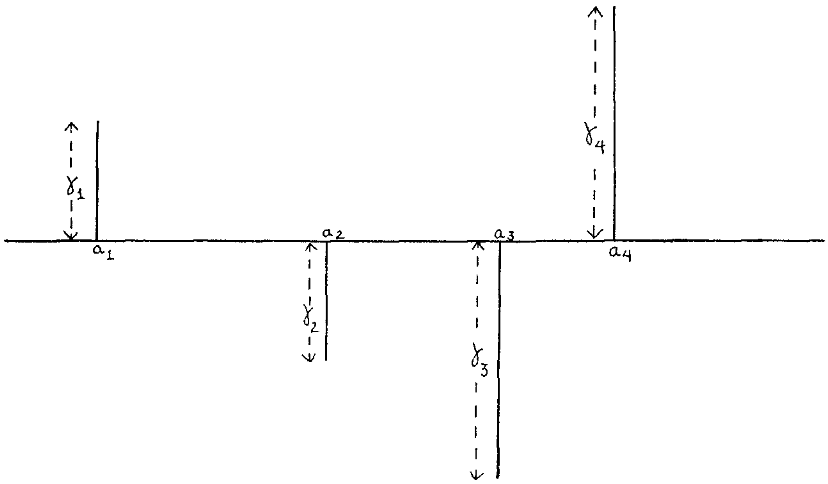

FIGURE 4

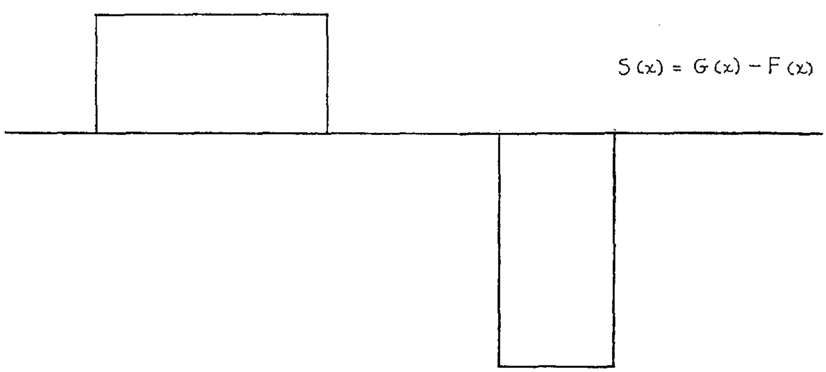

FIGURE 5

Thirdly, if $T(y) = \int_0^y S(x)\, dx$ then

$$T(1) = 0 \tag{6}$$

since $T(1) = \int_0^1 S(x)\, dx = xS(x)]_0^1 - \int_0^1 xs(x)\, dx = 0$.

Finally, conditions (5) and (6) together imply that

$$T(y) \geqslant 0, \qquad 0 \leqslant y < 1. \tag{7}$$

If G and F are discrete distributions differing by a single MPS and if $S = G - F$ then S satisfies (5), (6), and (7). See Fig. 6.

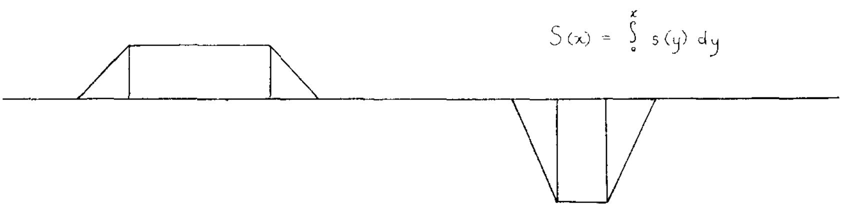

FIGURE 6

4. *Implications of Transitivity*

The concept of a MPS is the beginning, but only the beginning, of a definition of greater variability. To complete it we need to explore the implications of transitivity. That is, for our definition to be reasonable it should be the case that if X_1 is riskier than X_2 which is in turn riskier than X_3, then X_1 is riskier than X_3. Thus, if X and Y are the r.v.'s with c.d.f.'s F and G, we need to find a criterion for deciding whether G could have been obtained from F by a sequence of MPS's. We demonstrate in this section that the criterion is contained in conditions (6) and (7) above.[5]

We will proceed by first stating precisely in Theorem 1(a) the obvious fact that if G is obtained from F by a sequence of MPS's, then $G - F$ satisfies the integral conditions ((6) and (7)). Theorem 1(b) is roughly the converse of that statement: That is, we show that if $G - F$ satisfies the integral conditions, G could have been obtained from F to any desired degree of approximation by a sequence of MPS's.

THEOREM 1(a). *If* (a) *there is a sequence of c.d.f.'s* $\{F_n\}$ *converging* (*weakly*) *to* G, (*written* $F_n \to G$)[6] *and* (b) F_n *differs from* F_{n-1} *by a single* MPS, (*which implies* $F_n = F_{n-1} + S_n = F_0 + \sum_{i=1}^{n} S_i$, *where* $F_0 = F$, *and where each* S_i *satisfies* (6) *and* (7)), *then* $G = F + \sum_{i=1}^{\infty} S_i = F + S$ *and* S *satisfies* (6) *and* (7).

The proof, which is obvious, is omitted.

[5] Condition (5) could not be part of such a criterion for it is easy to construct examples of c.d.f's which differ by two MPS's such that their difference does not satisfy (5).

[6] Let $E(u) = \int_0^1 u(x)\,dG(x)$ and $E_n(u) = \int_0^1 u(x)\,dF_n(x)$. Then $F_n \to G$ if and only if $E_n(u) \to E(u)$ for all continuous u on [0, 1]. See [3, p. 243].

THEOREM 1(b). *If $G - F$ satisfies the integral conditions* (6) *and* (7), *then there exist sequences F_n and G_n, $F_n \to F$, $G_n \to G$, such that for each n, G_n could have been obtained from F_n by a finite number of MPS's.*

The proof is an immediate consequence of the following two lemmas: the first proves the theorem for step functions with a finite number of steps; and the second states that F and G may be approximated arbitrarily closely by step functions which satisfy the integral conditions.

LEMMA 1. *If X and Y are discrete r.v.'s whose c.d.f.'s F and G have a finite number of points of increase, and if $S = G - F$ satisfies* (6) *and* (7), *then there exist c.d.f.'s, $F_0, \ldots, F_n$ such that $F_0 = F$, $F_n = G$, and F_i differs from F_{i-1} by a single MPS.*

Proof. S is a step function with a finite number of steps. Let $I_1 = (a_1, a_2)$ be the first positive step of S. If I_1 does not exist, $S(x) \equiv 0$ implying that $F = G$ and the lemma is trivally true. Let $I_2 = (a_3, a_4)$ be the first negative step of $S(x)$. By (7), $a_2 < a_3$. Let γ_1 be the value of $S(x)$ on I_1 and $-\gamma_2$ be the value of $S(x)$ on I_2.

Either

$$\gamma_1(a_2 - a_1) \geqslant \gamma_2(a_4 - a_3) \tag{8}$$

or

$$\gamma_1(a_2 - a_1) < \gamma_2(a_4 - a_3). \tag{9}$$

If (8) holds, let $\hat{a}_4 = a_4$. There is an $\hat{a}_2$ satisfying $a_1 < \hat{a}_2 \leqslant a_2$ such that

$$\gamma_1(\hat{a}_2 - a_1) = \gamma_2(\hat{a}_4 - a_3). \tag{10}$$

If (9) holds, let $\hat{a}_2 = a_2$; then there is an $\hat{a}_4$ satisfying $a_3 < \hat{a}_4 < a_4$ such that (10) holds. Define $S_1(x)$ by

$$S_1(x) = \begin{cases} \gamma_1 & \text{for } a_1 < x < \hat{a}_2 \\ -\gamma_2 & \text{for } a_3 < x < \hat{a}_4 \\ 0 & \text{otherwise.} \end{cases}$$

Then if $F_1 = F_0 + S_1$, F_1 differs from F by a single MPS and $S^{(1)} = G - F_1$ satisfies (6) and (7).

We use this technique to construct S_2 from $S^{(1)}$ and define F_2 by $F_2 = F_1 + S_2$. Because S is a step function with a finite number of steps, the process terminates after a finite number of iterations.

LEMMA 2. *Let F and G be c.d.f.'s defined on* [0, 1]. *Let $T(y) = \int_0^y (G(x) - F(x))\, dx$. If*

$$T(y) \geqslant 0, \qquad 0 \leqslant y \leqslant 1, \tag{6}$$

and

$$T(1) = 0 \tag{7}$$

then, for each n, there exists F_n and G_n, c.d.f.'s of discrete r.v.'s with a finite number of points of increase, such that if

$$\| F_n - F \| = \int_0^1 | F_n(x) - F(x)| \, dx$$

and

$$\| G_n - G \| = \int_0^1 | G_n(x) - G(x)| \, dx,$$

then[7]

$$\| F_n - F \| + \| G_n - G \| \leqslant \frac{4}{n} \tag{11}$$

and if $T_n(y) = \int_0^y (G_n(x) - F_n(x)) \, dx$ *then*

$$T_n(y) \geqslant 0 \tag{12}$$

and

$$T_n(1) = 0. \tag{13}$$

Proof. We prove this by constructing F_n and G_n for fixed n. For $i = 1,..., n$ let $I_i = ((i - 1)/n, i/n)$. Let $\bar{f}_i = F(i/n)$ and define $\bar{F}_n$ by $\bar{F}_n(x) = \bar{f}_i$ for $x \in I_i$ (see Fig. 7). Since F is monotonic $\bar{F}_n(x) \geqslant F(x)$. It follows also from monotonicity that $\| F_n - F \| \leqslant 1/n$. If $\hat{F}_n(x)$ is any step function constant on each I_i such that $\hat{F}_n(x) \in F(I_i)$ for $x \in I_i$ then $\| \hat{F}_n - \bar{F}_n \| \leqslant 1/n$ and

$$\| \hat{F}_n - F \| \leqslant \| \hat{F}_n - \bar{F}^n \| + \| \bar{F}_n - F \| \leqslant \frac{2}{n} \cdot$$

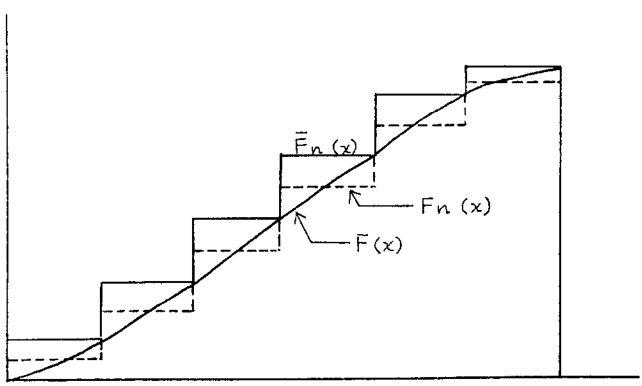

FIGURE 7

[7] Condition (11) implies weak convergence. See [3, p. 243].

Similarly if $\hat{G}_n(x)$ is a step function such that $x \in I_i$ implies $\hat{G}_n(x) \in G(I_i)$ then $\| \hat{G}_n - G \| \leqslant 2/n$.

For every i there exist $f_i \in F(I_i)$ and $g_i \in G(I_i)$ such that $(g_i - f_i)/n = \int_{I_i} (G(x) - F(x))\, dx$. Let $\hat{F}_n(x) = f_i$ and $\hat{G}_n(x) = g_i$, $x \in I_i$. We now show that $\hat{F}_n$ and $\hat{G}_n$ satisfy (11), (12), and (13). We have already shown that (11) is satisfied. Observe that

$$\begin{aligned} \hat{T}_n(1) &= \int_0^1 (\hat{G}_n(x) - \hat{F}_n(x))\, dx \\ &= \sum_{i=1}^{n} \int_{I_i} (\hat{G}_n(x) - \hat{F}_n(x))\, dx \\ &= \sum_{i=1}^{n} \frac{g_i - f_i}{n} = \sum_{i=1}^{n} \int_{I_i} (G(x) - F(x))\, dx \\ &= \int_0^1 (G(x) - F(x))\, dx = T(1) = 0, \end{aligned}$$

so that (13) is satisfied. It remains to show that $\hat{T}_n(y) \geqslant 0$. If $y = j/n$ for $j = 0, 1, \ldots, n$, then $\hat{T}_n(y) = T(j/n) \geqslant 0$ so we need only examine the case where $y = j/n + \alpha$, $0 < \alpha < 1/n$. Then, $\hat{T}_n(x) = T(j/n) + \alpha(g_j - f_j)$. If $g_j > f_j$ both terms of the sum are positive. If $g_j < f_j$ then

$$T\left(\frac{j}{n}\right) + \alpha(g_j - f_j) > T\left(\frac{j}{n}\right) + \frac{1}{n}(g_j - f_j) = T\left(\frac{j+1}{n}\right) \geqslant 0.$$

This completes the proof except for a technical detail. Neither $\hat{F}_n$ nor $\hat{G}_n$ are necessarily c.d.f.'s. We remedy this by defining $F_n(x) = \hat{F}_n(x)$ for $x \in (0, 1)$ and $F_n(0) = 0$ and $F_n(1) = 1$. G_n is defined similarly and if $\hat{F}_n$ and $\hat{G}_n$ satisfy (11), (12), and (13) so do F_n and G_n.

III. Partial Orderings of Distribution Functions

A definition of greater uncertainty is, or should be, a definition of a partial ordering on a set of distribution functions. In this section we formally define the three partial orderings corresponding to the first three concepts of increasing risk set out in Section I and prove their equivalence.

1. *Partial Orderings*

A partial ordering $\leqslant_p$ on a set is a binary, transitive, reflexive and antisymmetric[8] relation. The set over which our partial orderings are defined is the set of distribution functions on [0, 1]. We shall use $F \leqslant_p G$

[8] A relation $\leqslant_p$ is antisymmetric if $A \leqslant_p B$ and $B \leqslant_p A$ implies $A = B$.

interchangeably with $X \leqslant_p Y$ where F and G are the c.d.f.'s of the r.v.'s X and Y.

2. *Definition of* $\leqslant_I$

Following the discussion of the last section we define a partial ordering $\leqslant_I$ as follows: $F \leqslant_I G$ if and only if $G - F$ satisfies the integral conditions (6) and (7).

LEMMA 3. *$\leqslant_I$ is a partial ordering.*

Proof. It is immediate that $\leqslant_I$ is transitive and reflexive. We need only demonstrate antisymmetry. Assume $F \leqslant_I G$ and $G \leqslant_I F$. Define S_1 and S_2 as follows:

$$S_1 = G - F \qquad \text{and} \qquad S_2 = F - G.$$

Thus $S_1 + S_2 = 0$. Furthermore, if $T_i(y) = \int_0^y S_i(x)\,dx$, then $T_i(y) \geqslant 0$, since $F \leqslant_I G$ and $G \leqslant_I F$. Since $0 = \int_0^y (S_1(x) + S_2(x))\,dx = T_1(y) + T_2(y) = 0$ and $T_i(y) \geqslant 0$, $T_i(y) = 0$. We shall prove this implies that $S_1(x) = 0$ a.e. (almost everywhere), or $F(x) = G(x)$ a.e. This will prove the lemma.[9]

Since $S_1(x)$ is of bounded variation (it is the difference of two monotonic functions) its discontinuities form a set of measure zero. Let us call this set N. Define

$$\hat{S}_1(x) = \begin{cases} 0 & \text{for } x \in N \\ S_1(x) & \text{otherwise.} \end{cases}$$

Then $\int_0^y S_1(x)\,dx = \int_0^y \hat{S}_1(x)\,dx = T_1(y)$. Suppose there is an $\hat{x}$ such that $\hat{S}_1(\hat{x}) \neq 0$, say $\hat{S}_1(\hat{x}) > 0$. Then $\hat{S}_1(x) > 0$ for $x \in (\hat{x} - \epsilon, \hat{x} + \epsilon)$ for some $\epsilon > 0$ (since $\hat{S}_1(x)$ is continuous at $\hat{x}$). Then, $T_1(x - \epsilon) < T_1(x + \epsilon)$. This contradiction completes the proof.

3. *Definition of* $\leqslant_u$

We define the partial ordering $\leqslant_u$ corresponding to the idea that X is less risky than Y if every risk averter prefers X to Y as follows. $F \leqslant_u G$ if and only if for every bounded concave function U, $\int_0^1 U(x)\,dF(x) \geqslant \int_0^1 U(x)\,dG(x)$. It is immediate that $\leqslant_u$ is transitive and reflexive. That $\leqslant_u$ is antisymmetric is an immediate consequence of Theorem 2 below.

4. *Definition of* $\leqslant_a$

Corresponding to the notion that X is less risky than Y if Y has the same distribution as X plus some noise is the partial ordering $\leqslant_a$ which

[9] We shall follow the convention of considering two distribution functions to be equal if they differ only on a set of measure zero.

we now define. $F \leqslant_a G$ if and only if there exists a joint distribution function $H(x, z)$ of the r.v.'s X and Z defined on $[0, 1] \times [-1, 1]$ such that if

$$J(y) = \Pr(X + Z \leqslant y),$$

then

$$F(x) = H(x, 1), \qquad 0 \leqslant x \leqslant 1,$$
$$G(y) = J(y), \qquad 0 \leqslant y \leqslant 1,$$

and

$$E(Z \mid X = x) = 0 \qquad \text{for all } x. \tag{14}$$

The equivalent definition in terms of r.v.'s follows: $X \leqslant_a Y$ if there exists an r.v. Z satisfying (14) such that

$$Y \underset{d}{=} X + Z. \tag{15}$$

It is important to realize that (15) does *not* mean that $Y = X + Z$.

For the special case where X and Y are discrete distributions concentrated at a finite number of points, the relation $\leqslant_a$ can be given a useful and tractable characterization. Without loss of generality assume that X and Y are concentrated at the points $a_1, a_2, \ldots, a_n$. Then the c.d.f.'s of X and Y are determined by the numbers

$$f_i = \Pr(X = a_i)$$

and

$$g_i = \Pr(Y = a_i).$$

Then $X \leqslant_a Y$ if and only if there exist n^2 numbers $c_{ij} \geqslant 0$ such that

$$\sum_j c_{ij} = 1, \qquad i = 1,\ldots, n, \tag{16}$$

$$\sum_j c_{ij}(a_j - a_i) = 0, \qquad i = 1,\ldots, n, \tag{14$'$}$$

and

$$g_j = \sum_i f_i c_{ij}, \qquad j = 1,\ldots, n. \tag{15$'$}$$

To see that this is so, define an r.v. Z conditional on X as follows,

$$c_{ij} = \Pr(Z = a_j - a_i \mid X = a_i).$$

Then (16) states that this equation in fact defines a r.v. while (14′) and (15′)

are the analoges of (14) and (15). These conditions can be written in matrix form:

$$Ca = a, \tag{14''}$$

$$g = fC, \tag{15''}$$

$$Ce = e, \tag{16''}$$

where $e = (1,..., 1)$ is the vector composed entirely of 1's. If f^1, f^2, and f^3 are vectors defining the c.d.f.'s of the discrete r.v.'s X^1, X^2, and X^3, ($f_i^k = \Pr(X^k = a_i)$), and if $X^1 \leqslant_a X^2$ and $X^2 \leqslant_a X^3$ then there exist matrices C^1 and C^2 such that $C^1a = C^2a = a$; $C^1e = C^2e = e$, while $f^2 = f^1C^1$ and $f^3 = f^2C^2$. Let $C^* = C^1C^2$. Then $f^3 = f^1C^*$ and $C^*a = C^1C^2a = C^1a = a$ and similarly $C^*e = e$. We have proved

LEMMA 4. *If X^1, X^2, and X^3 are concentrated at a finite number of points, then $X^1 \leqslant_a X^2 \leqslant_a X^3$ implies $X^1 \leqslant_a X^3$.*

5. *Equivalence of* $\leqslant_I$, $\leqslant_a$, $\leqslant_u$

We now state and prove the major result of this paper.

THEOREM 2. *The following statements are equivalent*:

(A) $F \leqslant_u G$;

(B) $F \leqslant_I G$;

(C) $F \leqslant_a G$.

Proof. The proof consists of demonstrating the chain of implications (C) $\Rightarrow$ (A) $\Rightarrow$ (B) $\Rightarrow$ (C). Throughout the proof we adhere to the notational conventions introduced at the end of Section I.

(a) $X \leqslant_a Y \Rightarrow X \leqslant_u Y$.

By hypothesis there is an r.v. Z such that $Y \underset{a}{=} X + Z$ and $E(Z \mid X) = 0$. For every fixed X and concave U we have, upon taking expectations with respect to Z, by Jensen's inequality

$$E_X U(X + Z) \leqslant U(E(X + Z)) = U(X).$$

Taking expectations with respect to X,

$$EE_X U(X + Z) \leqslant EU(X)$$

or

$$EU(Y) \leqslant EU(X).$$

(b) $F \leqslant_u G \Rightarrow F \leqslant_I G$.[10]

If $S = G - F$ then $F \leqslant_u G$ implies $\int_0^1 U(x)\, dS(X) \leqslant 0$ for all concave U. Since the identity function and its negative are both concave we have that $\int_0^1 x\, dS(x) \leqslant 0$ and $\int_0^1 (-x)\, dS(x) \leqslant 0$ so that $\int_0^1 x\, dS(x) = 0$. Integrating by parts we find that $T(1) = 0$. It remains to show that $T(y) \geqslant 0$ for all $y \in [0, 1]$. For fixed y, let $b_y(x) = \mathrm{Max}(y - x, 0)$. Then $-b_y(x)$ is concave and $0 \leqslant \int_0^1 b_y(x)\, dS(x) = \int_0^y (y - x)\, dS(x) = yS(y) - \int_0^y x\, dS(x)$. Integrating the last term by parts we find that

$$\begin{aligned} -\int_0^y x\, dS(x) &= -xS(x)\Big]_0^y + \int_0^y S(x)\, dx \\ &= -yS(y) + T(y). \end{aligned}$$

Thus, $T(y) = \int_0^1 b_y(x)\, dS(x) \geqslant 0$.

(c) $F \leqslant_I G \Rightarrow F \leqslant_a G$.

We prove this implication first for the case where F and G are discrete r.v.'s which differ by a single MPS. Using the notation of Section II.2, let F and G attribute the same probability weight to all but four points $a_1 < a_2 < a_3 < a_4$. Let $\Pr(X = a_k) = f_k$ and $\Pr(Y = a_k) = g_k$. If $\gamma_k = g_k - f_k$, then

$$\gamma_1 = -\gamma_2 \geqslant 0, \qquad \gamma_4 = -\gamma_3 \geqslant 0 \tag{4.i}$$

and

$$\sum_{k=1}^{4} \gamma_k a_k = 0 \tag{4.ii}$$

are the conditions that G differs from F by a single MPS. To prove that $F \leqslant_a G$ we need only show the existence of $c_{ij} \geqslant 0$ $(i, j = 1, 2, 3, 4)$ satisfying (14′), (15′), and (16). Consider,

$$\{c_{ij}\} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \dfrac{\gamma_1(a_4 - a_2)}{f_2(a_4 - a_1)} & \dfrac{g_2}{f_2} & 0 & \dfrac{\gamma_1(a_2 - a_1)}{f_2(a_4 - a_1)} \\ \dfrac{\gamma_4(a_4 - a_3)}{f_3(a_4 - a_1)} & 0 & \dfrac{g_3}{f_3} & \dfrac{\gamma_4(a_3 - a_1)}{f_3(a_4 - a_1)} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{17}$$

[10] We are indebted to David Wallace for the present simplified form of the proof. For continuously differentiable U, the reverse implication may be proved simply by integration by parts.

It is easy to verify that the c_{ij} defined by (17) do satisfy (16) and (14′). Thus if we define Z, as before, by

$$c_{ij} = \Pr(Z = a_j - a_i \mid X = a_i)$$

then Z is a random variable, conditional on X, satisfying $E(Z \mid X) = 0$. It remains to establish (15′) or that $Y \underset{d}{=} X + Z$. Consider $Y^1 = X + Z$. Y^1 is a discrete r.v. which, since $E(Z) = 0$, has the same mean as Y. It can differ from Y only if it attributes different probability weight to the points a_1, a_2, a_3, a_4. But,

$$\begin{aligned}\Pr(Y^1 = a_2) &= \Pr(X = a_2) \cdot \Pr(Z = 0 \mid X = a_2) \\ &= f_2 \cdot \frac{g_2}{f_2} = g_2 = \Pr(Y = a_2).\end{aligned}$$

Similarly, $\Pr(Y^1 = a_3) = \Pr(Y = a_3)$. Then Y and Y^1 can differ in the assignment of probability weight in at most two points. But $\Pr(Y = a_1) > \Pr(Y^1 = a_1)$ implies $\Pr(Y^1 = a_4) > \Pr(Y = a_4)$ which in turn implies that $E(Y^1) > E(Y)$, a contradiction. Thus, $Y \underset{d}{=} Y^1 \underset{d}{=} X + Z$.

Lemmas 1 and 4 allow us to extend this result to all discrete distributions with a finite number of points of increase. We use Theorem 1(b) to extend it to all c.d.f.'s. If $F \leqslant_I G$, there exists sequences $\{F_n\}$ and $\{G_n\}$ of discrete distributions with a finite number of points of increase such that $F_n \to F$ and $G_n \to G$ and $F_n \leqslant_I G_n$. We have just shown $F_n \leqslant_a G_n$. Let X_n and Y_n be the r.v.'s with distributions F_n and G_n. There is for each n an $H_n(x, z)$, the joint distribution function of the r.v.'s X_n and Z_n, such that if $J_n(y) = \Pr(X_n + Z_n \leqslant y)$, then

$$J_n(y) = G_n(y), \tag{18}$$

$$F_n(x) = H_n(x, 1), \tag{19}$$

and

$$E(X_n \mid Z_n) = 0. \tag{20}$$

Since H_n is a discrete distribution function Eq. (20) can be phrased as

$$\int_0^1 \int_{-1}^1 u(x)z \, dH_n(x, z) = 0 \tag{21}$$

for all continuous functions u defined on [0, 1]. Since H_n is stochastically bounded, the sequence $\{H_n\}$ has a subsequence $\{H_{n'}\}$ which converges to a distribution function[11] $H(x, z)$ of the r.v.'s X and Z. Since $H_{n'}(x, 1) = F_{n'}(x) \to F$, $H_{n'}(x, 1) \to F$. Similarly, $J_{n'} \to G$. Let

$$M_{n'} = \int_0^1 \int_{-1}^1 u(x)z \, dH_{n'}(x, z).$$

[11] See [3, pp. 247, 261].

By the definition of weak convergence $M_{n'} \to \int_0^1 \int_{-1}^1 u(x)z\, dH(x, z)$. But $\{M_{n'}\}$ is a sequence all of whose terms are 0 and it must therefore converge to 0. Therefore $\int_0^1 \int_{-1}^1 u(x)z\, dH(x, z) = 0$, which implies $E(Z \mid X) = 0$. This completes the proof.

6. *Further Remarks*

We conclude this section with two remarks about these orderings.

A. *Partial versus Complete Orderings.* In the previous subsection, we established that $\geqslant_a$, $\geqslant_I$, and $\geqslant_u$ define equivalent partial orderings over distributions with the same mean. It should be emphasized that these orderings are only partial, that is, if F and G have the same mean but $\int_0^1 (F(x) - G(x))\, dx = T(y)$ changes sign, F and G cannot be ordered. But this means in turn that there always exist two concave functions, U_1 and U_2, such that $\int_0^1 U_1\, dF(x) > \int_0^1 dG(x)$ while $\int_0^1 U_2\, dF(x) < \int_0^1 U_2\, dG(x)$; that is, there is some risk averse individual who prefers F to G and another who prefers G to F. On the other hand, the ordering $\geqslant_V$ associated with the mean-variance analysis ($X \leqslant_V Y$ if $EX = EY$ and $EX^2 \leqslant EY^2$) is a complete ordering, i.e., if X and Y have the same mean, either $X \leqslant_V Y$ or $X \geqslant_V X$.[12]

B. *Concavity.* We have already noted that if U is concave, $X \leqslant_I Y$ implies $EU(X) \leqslant EU(Y)$. Similarly, given any differentiable function U which over the interval [0, 1] is neither concave nor convex, then there exist distribution functions F, G, and H, $F \geqslant_I G \geqslant_I H$, such that $\int_0^1 U(x)\, dF \leqslant \int_0^1 U(x)\, dG$, but $\int_0^1 U(x)\, dG \geqslant \int_0^1 U(x)\, dH$.

In short, $\geqslant_I$ defines the set of all concave functions: A function U is concave if and only if $X \leqslant_I Y$ implies $EU(X) \leqslant EU(Y)$.

[12] Another way of making this point is to observe that $\geqslant_V$ is stronger than $\geqslant_I$ because many distributions which can be ordered with respect to $\geqslant_V$ cannot be ordered with respect to $\geqslant_I$. Clearly there exist weaker as well as stronger orderings than $\geqslant_I$. One such weaker ordering, to which we drew attention in earlier versions of this paper, is the following. A r.v. X which is a mixture between a r.v. Y and a sure thing with the same mean—a random variable concentrated at the point $E(Y)$—is surely less risky than Y itself. We could use this notion to define a partial ordering $\geqslant_M$. It is obvious that $\geqslant_M$ implies $\geqslant_I$ since the difference between X and Y satisfies the integral conditions. It is also clear that $\geqslant_M$ is a very weak ordering in the sense that very few r.v.'s can be ordered by $\geqslant_M$. In fact if $\bar{Y}$ is the sure thing concentrated at $E(Y)$ than it can be shown that $Y \geqslant_M X$ iff $X \underset{d}{=} aY + (1 - a)\bar{Y}$ for $0 \leqslant a \leqslant 1$. This indicates that $\geqslant_M$ is not a particularly interesting partial ordering. We are indebted to an anonymous referee for pointing out the deficiencies of $\geqslant_M$.

IV. Mean-Variance Analysis

The method most frequently used for comparing uncertain prospects has been mean-variance analysis. It is easy to show that such comparisons may lead to unjustified conclusions. For instance, if X and Y have the same mean, X may have a lower variance and yet Y will be preferred to X by some risk averse individuals. To see this, all we need observe is that, although $F \leqslant_u G \Rightarrow F \geqslant_V G$ (since variance is a convex function), $F \geqslant_V G$ does not imply $F \geqslant_u G$. Indeed by arguments closely analogous to those used earlier, it can be shown that a function U is quadratic if and only if $X \geqslant_V Y$ implies $EU(X) \geqslant EU(Y)$. An immediate consequence of this is that if $U(x)$ is any nonquadratic concave function, then there exists random variables X_i, $i = 1, 2, 3$, all with the same mean such that $EX_1^2 < EX_2^2$ but $EX_2^2 > EX_3^2$ while $EU(X_1) < EU(X_2) < EU(X_3)$, i.e., the ranking by variance and the ranking by expected utility are different.

Tobin has conjectured that mean-variance analysis may be appropriate if the class of distributions—and thus the class of changes in distributions—is restricted. This is true but the restrictions required are, as far as is presently known, very severe. Tobin's proof is—as he implicitly recognizes (in [7, pp. 20–21])—valid only for distributions which differ only by "location parameters." (See [3, p. 144] for a discussion of this classical concept.) That is, Tobin is only willing to consider changes in distributions from F to G if there exist a and b ($a > 0$) such that $F(x) = G(ax + b)$. Such changes amount only to a change in the centering of the distribution and a uniform shrinking or stretching of the distribution—equivalent to a change in units.

There has been some needless confusion along these lines about the concept of a two parameter family of distribution functions. It is undeniable that all distributions which differ only by location parameters form a two parameter family. In general, what is meant by a "two parameter family"? To us a two parameter family of distributions would seem to be any set of distributions such that one member of the set would be picked out be selecting two parameters. As Tobin has put it, it is "one such that it is necessary to know just two numbers in order to describe the whole distribution." Technically that is, a two parameter family is a mapping from E^2 into the space of distribution functions.[13] It is clear that for this broad definition of two parameter family, Tobin's conjecture cannot possibly hold, for nothing restricts the range of this mapping.

Other definitions of two parameter family are of course possible. They involve essentially restrictions to "nice" mappings from E^2 to the space of

[13] Or some subset of E^2; we might restrict one or both of our parameters to be nonnegative.

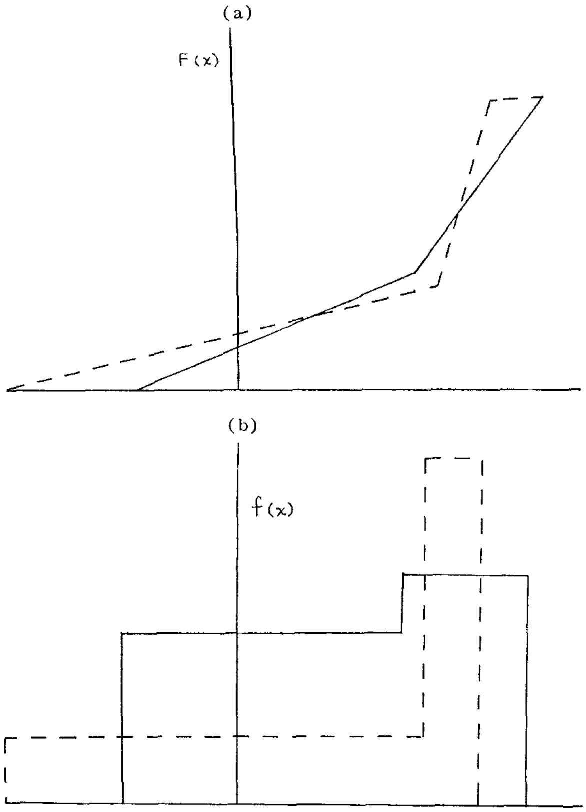

FIGURE 8

distribution functions, e.g., a family of distributions with an explicit algebraic form containing only two parameters which can vary. It is easy, however, to construct examples where if the variance, σ^2, changes with the mean, μ, held constant, $\partial T(y)/\partial\sigma^2$ changes sign, where $T(y, \sigma^2, \mu) = \int_0^y F(x, \sigma^2, \mu)$; that is, there exist individuals with concave utility functions who are better off with an increase in variance.[14]

[14] Consider, for instance, the family of distributions defined as follows: $(a, c > 0)$. (In this example, for expositional clarity we have abandoned our usual convention of defining distributions over [0, 1])

$$F(x; a, c) = \begin{cases} 0 & \text{for } x \leqslant 1 - 0.25/a \\ ax + 0.25 - a & \text{for } 1 - 0.25/a \leqslant x \leqslant 1 + (2c - 0.5)/c - a) \\ cx + 0.75 - 3c & \text{for } 1 + (2c - 0.5)/(c - a) \leqslant x \leqslant 3 + 0.25/c \\ 1 & \text{for } x > 3 + 0.25/c \end{cases}$$

Two members of the family with the same mean but different variances are depicted in Fig. 8(a). They clearly do not satisfy condition (7). The density functions are illustrated in Fig. 8(b).

Acknowledgments

This is a revised version of papers presented at the Cowles Symposium on Capital Theory and Uncertainty (November, 1968), and at the Chicago Growth Symposium (November, 1967).

The authors are deeply indebted to the participants in the symposiums and to David Ragozin, Peter Diamond, David Wallace, and David Grether.

Our problem is not a new one, nor is our approach completely novel; our result is, we think, new. Our interest in this topic was whetted by Peter Diamond [2]. Robert Solow used a device similar to our Mean Preserving Spread (Section II, above) to compare lag structures in [6]. The problem of "stochastic dominance" is a standard one in the (statistics) operations research literature. For other approaches to the problem, see, for instance, [1]. [4, 5] have recently provided an alternative proof to our Theorem 2(b) and its converse (p. 238).

References

1. S. A. Bessler and A. F. Veinott, Jr., Optimal policy for a dynamic multi-echelon inventory model, *Naval Research Logistics Quarterly* **13** (1966), 355–387.
2. P. A. Diamond, Savings decisions under uncertainty, Working paper no. 71, Institute of Business Economic Research, University of California, Berkeley, June 1965.
3. W. Feller, "An Introduction to Probability Theory and Its Applications," Vol. II, Wiley, New York, 1966.
4. J. Hader and W. Russell, Rules for ordering uncertain prospects, *Amer. Econ. Rev.* **59** (1969), 25–34.
5. G. Hanoch and C. Levy, Efficiency Analysis of Choices Involving risk, *Review of Economic Studies* **36** (1969).
6. R. M. Solow, A note on dynamic multipliers, *Econometrica* **19** (1951), 306–316.
7. J. Tobin, The theory of portfolio selection, *in* F. Hahn and F. Brechling, "The Theory of Interest Rates," MacMillan, London, 1965.
8. J. Tobin, Comment on Borch and Feldstein, *Rev. Econ. Studies* **36** (1969), 13–14.