Cowles Foundation Paper 341a

Reprinted from JOURNAL OF ECONOMIC THEORY
Vol. 2, No. 3, September 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

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

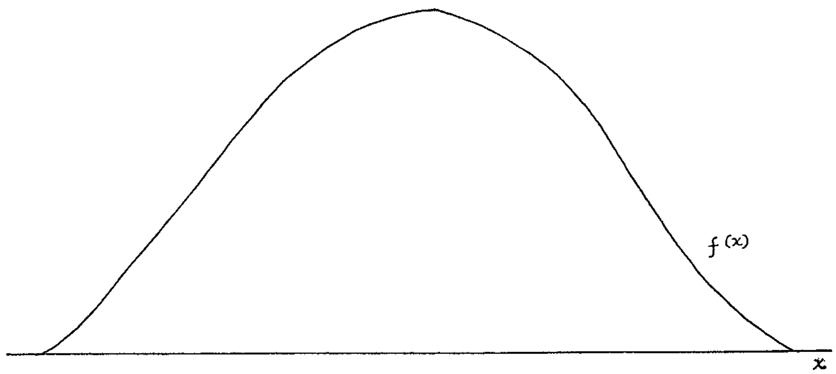

FIGURE 1

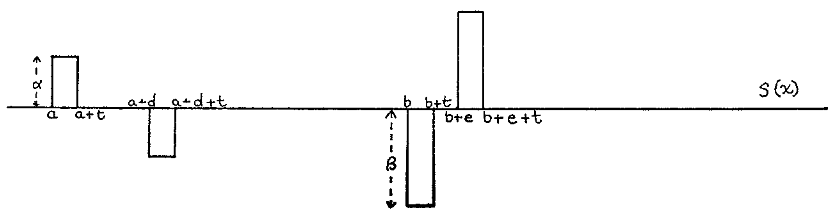

FIGURE 2

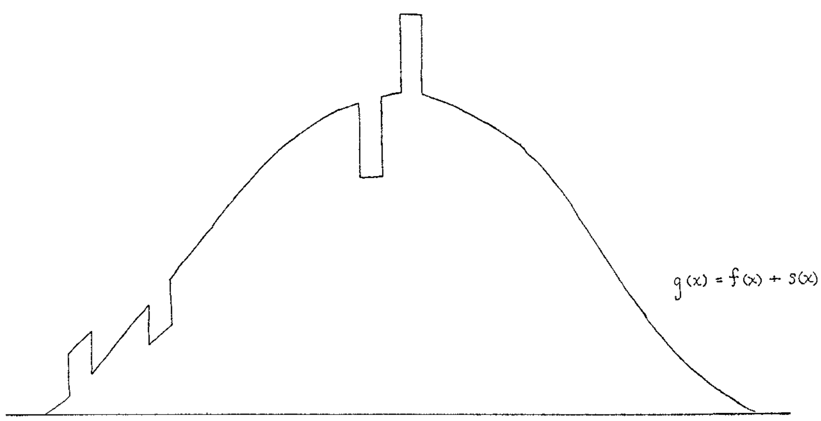

FIGURE 3

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

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

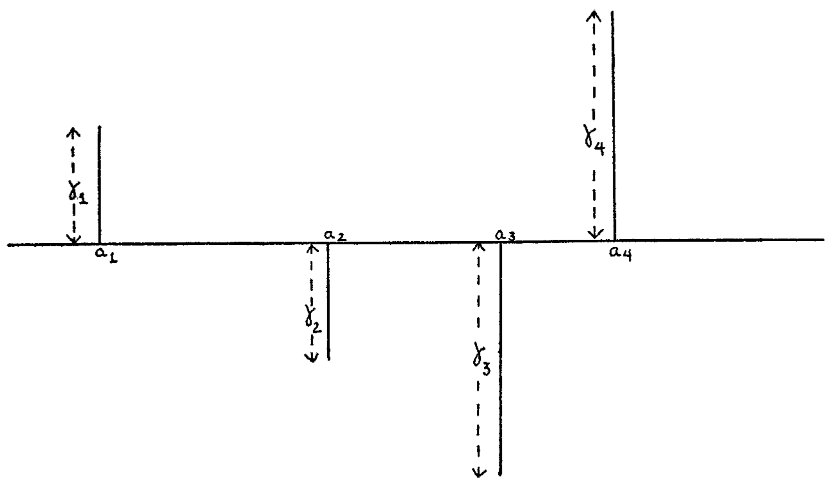

FIGURE 4

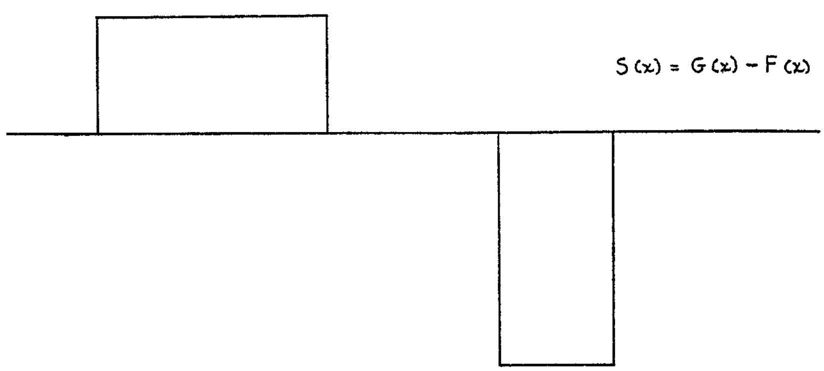

FIGURE 5

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.

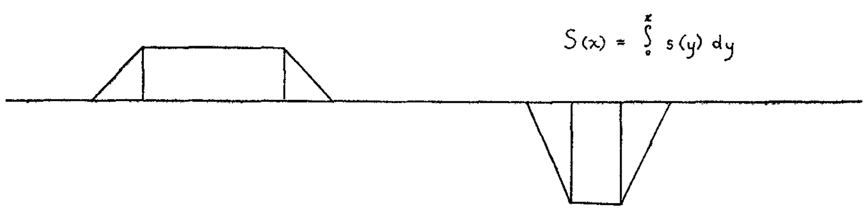

FIGURE 6

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* $F_0 \equiv F$, *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 trivially 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

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

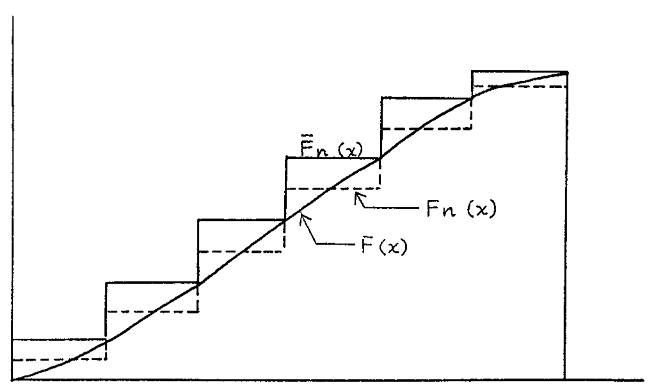

FIGURE 7

[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 $j = 0, 1,..., n$, 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

$$\begin{aligned} F(x) &= H(x, 1), \qquad 0 \leqslant x \leqslant 1, \\ G(y) &= J(y), \qquad 0 \leqslant y \leqslant 1, \end{aligned}$$

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 $b_y(x) = \text{Max}(y - x, 0)$. 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.

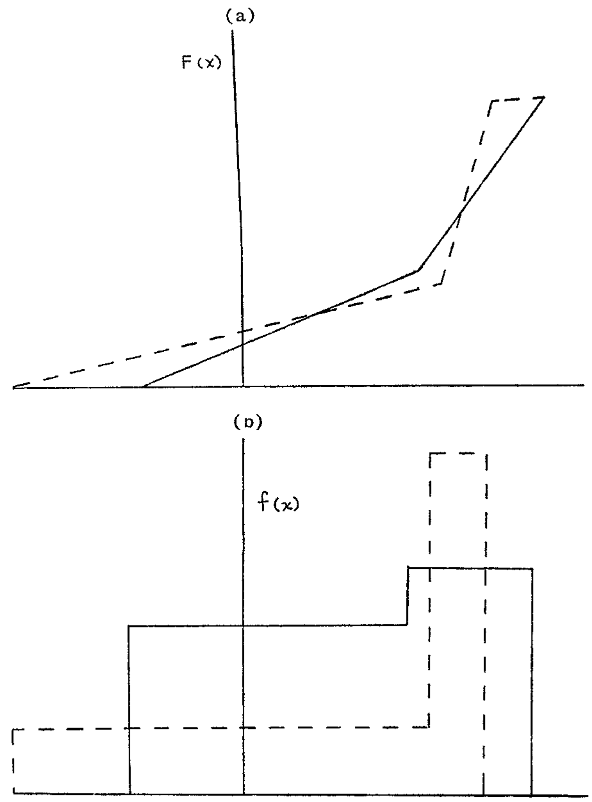

FIGURE 8

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.

Cowles Foundation Paper 341b

Reprinted from JOURNAL OF ECONOMIC THEORY
Vol. 3, No 1, March 1971

Increasing Risk II: Its Economic Consequences*

MICHAEL ROTHSCHILD

Harvard University, Cambridge, Massachusetts 02138

AND

JOSEPH E. STIGLITZ

Cowles Foundation, Yale University, New Haven, Connecticut 06520

Received April 27, 1970

1. INTRODUCTION

The first part of this paper [11] established the equivalence of the following three statements comparing random variables X and Y with the same expected value:

(a) "All risk averters—those with concave utility functions—prefer X to Y."

(b) "Y is equal to X plus some noise."

(c) "Y has more weight in its tails than X."[1]

The equivalence of these three apparently different concepts seemed to us a good argument that our definition was the natural definition of increasing risk. It also suggested a useful multiplicity of approaches to investigations of the effect of risk on economic decisions. This paper investigates two such approaches.

* Support from the National Science Foundation and the Ford Foundation is gratefully acknowledged.

[1] Formal versions of (a), (b), and (c) are:

(a) $EU(X) \geqslant EU(Y)$ for all concave Y.

(b) There exists a random variable Z such that Y has the same distribution as $X + Z$, where $E[Z \mid X] = 0$ for all X.

(c) If the points of increase of F and G, the distribution functions of X and Y, are confined to the closed interval $[a, b]$, then if

$$T(y) = \int_a^Y (F(x) - G(x))\,dx,\ T(y) \geqslant 0 \text{ and } T(b) = 0.$$

The first concerns the economic effects of increasing risk. Our paper contained some sharp criticism of the conventional mean-variance treatment of this problem.[2] Mindful that counsels of perfection are best accompanied by demonstrations of the possibility of attaining virtue, we committed ourselves to demonstrating that our approach could yield "the kind of comparative static results that economists are interested in [16]." The next section of our paper is our attempt to deliver on that promise.

If an individual's utility is a function of some control parameter α, and a random variable θ, then the individual will choose α to maximize his expected utility

$$\max_{\alpha} \int U(\theta, \alpha)\, dF\theta. \tag{1}$$

The optimal α must satisfy

$$\int \frac{\partial U(\theta, \alpha)\, dF(\theta)}{\partial \alpha} = EU_{\alpha}(\theta, \alpha) = 0. \tag{2}$$

Let α^* be the unique solution to (2), and assume that in the neighborhood of α^*, U is monotone decreasing in α. If $U_{\alpha}(\theta, \alpha)$ is a concave function of θ, an increase in riskiness will decrease α^*. This is so because (*a*) states exactly that increasing risk decreases $EU_{\alpha}(\theta, \alpha)$; lowering α restores the first order conditions of (2). Similarly, if $U_{\alpha}(\theta, \alpha)$ is convex in θ, increases in the variability of θ raise α^*. If $U_{\alpha}(\theta, \alpha)$ is neither convex nor concave, then the effect of an increase in risk is ambiguous.[3]

[2] As we emphasized in Part I [11], our approach yields only a partial ordering of distributions. The mean-variance approach gives rise to a complete ordering of distributions (with the same expected value) and, thus, often gives answers as to the effects of changes in uncertainty when our approach cannot. But this is an argument for rather than against our approach. The answers of mean-variance analysis are spurious; they hold only if the utility function or the class of distributions is arbitrarily restricted. Furthermore, mean-variance analysis does not seem to provide clues as to what restrictions must be imposed if its results are to hold. As we shall demonstrate repeatedly in what follows, the necessary restrictions are a byproduct of our method of analysis.

[3] Earlier studies (see, e.g., [4,14]) have made comparisons between perfectly certain and risky situations. In a "certain" situation, we choose α so that

$$U_{\alpha}(X, \hat{\alpha}) = 0.$$

Whether $\hat{\alpha} \gtrless \alpha^*$, where α^* is the solution to (1), depends simply on whether

$$EU_{\alpha}(X, \hat{\alpha}) \gtrless U(EX, \hat{\alpha}).$$

Jensen's inequality allows us to make unambiguous statements whenever U_{α} is concave or convex in X; but this is the same condition under which we are able to make unambiguous statements for a wider class of problems.

The next section applies this idea. We try to determine under what conditions functions which arise in some simple economic models are concave or convex. The first three models concern savings and investment behavior while the next three treat the choice of technique and the level of output when production or demand conditions are uncertain.[4] Our general conclusions are that the mean-variance approach gives misleadingly general results, and that the conditions for the relevant functions to be concave or convex can be usefully phrased in terms of the Arrow–Pratt concepts of relative and absolute risk aversion [1, 9]. When this approach fails to give unambiguous results of the effects of increases in riskiness, it is natural to ask what sorts of increases in riskiness do have determinate effects. We have found that the emphasis on the interplay between utility functions and changes in distribution functions which underlies the integral conditions of [11] often suggests a useful approach to this problem. An example is our discussion of the portfolio problem below.

The paper concludes with a different demonstration of the usefulness of the results of [11]. Choice under uncertainty can be conceived of as the choice of a random variable. The selection of a portfolio, e.g., is the selection of a random stream of returns. We contrast this view with that discussed above by formulating the individual's problem as one of choosing β to maximize

$$\int U(x)\, dF(x, \beta). \tag{3}$$

That is, he chooses the most favorable random variable. Condition (b) gives us a criterion for determining which random variables risk averters prefer. The final section of the paper gives two examples of how this condition may be used.

2. The Effects of Increasing Risk

A. *Savings and Uncertainty*

We begin with an analysis of the effect of uncertainty in the rate of return on savings.[5] Consider an individual who has a given wealth W_0, which he wishes to allocate between consumption today and consumption tomorrow. What he does not consume today he invests; at the end of the

[4] A further important application of the concepts developed in [11] is to the problem of income inequality which is discussed at length in the article by A. B. Atkinson [2].

[5] For a fuller discussion of this and related problems, see [4, 5, 7, 14].

period his investment yields the random return e per dollar invested. He wishes to allocate W_0 between the two periods to maximize two period expected utility:

$$E[U(C_1) + (1 - \delta)\, U(C_2)] = U((1 - s)\, W_0) + (1 - \delta)\, EU(sW_0e), \tag{4}$$

where s is the savings rate, and δ the pure rate of time discount. We assume the man is risk averse so that his utility function satisfies

$$U' > 0; \quad U'' < 0. \tag{5}$$

The necessary (and, because of (5), sufficient) condition for utility maximization is found by setting the derivative of (4) with respect to s equal to zero. The optimal savings rate must satisfy

$$U'((1 - s)\, W_0) = E[U'(sW_0e)](1 - \delta)\, e. \tag{6}$$

With its customary ambiguity, intuition suggests that increased uncertainty in the return on savings will either lower savings because "a bird in the hand is worth two in the bush" or raise it because a risk-averse individual, in order to insure his minimum standard of living, saves more in the face of increased uncertainty. Mean-variance analysis—equivalent to the assumption that U is quadratic—suggests that only the first argument has validity. If $U(C) = aC - \frac{1}{2}bC^2$, then the *RHS* of (6) is equal to

$$(1 - \delta)(aE(e) - bsW_0E(e^2)),$$

which decreases as e becomes riskier. A decrease in s will compensate for this disturbance of the first-order conditions as the *LHS* of (6) falls and the *RHS* of (6) rises as s decreases.

But this result is not general. Whether increasing variability increases or decreases s^* depends on whether $eU'(sW_0e)$ is convex or concave in e. Thus, increasing risk increases savings if

$$2U''(C) + U'''(C)\, C > 0, \tag{7}$$

and decreases it if

$$2U''(C) + U'''(C)\, C < 0 \tag{8}$$

for all positive C. Otherwise the effect of an increase in risk is ambiguous. Clearly, a nonpositive third derivative is sufficient for increasing risk to decrease savings. The Arrow–Pratt concept of relative risk aversion ($R = -\,U''C/U'$) can be used to put these results in a somewhat different form. Since R' has the same sign as $-(U'''C + U''(1 + R))$, if R is nonincreasing and greater than unity, (7) will hold, while, if R is nondecreasing and less than unity, (8) will hold. Thus for the class of constant relative risk aversion utility functions, $U(W) = (1 - a)\, W^{1-a}$, $(a > 0, a \neq 1)$, whether or not increased risk increases or decreases savings depends on whether or not $a > 1$ or $a < 1$.[6]

B. *A Portfolio Problem*

The conventional portfolio problem admits a similar, albeit somewhat more complicated analysis. Consider an individual who must allocate his portfolio between two assets: money, which yields a zero rate of return, and a risky asset which yields a random rate of return e. If he puts α of his initial wealth W_0 in the risky asset, his terminal wealth is $W(\alpha) = W_0(\alpha e + 1)$. The optimal portfolio mix is chosen to maximize the expected utility of terminal wealth, $EU(W(\alpha))$,[7] where U, an increasing concave function, satisfies (5). The chosen α is, therefore, the solution to

$$H(\alpha) = W_0 EU'e = W_0 \int U'(W(\alpha))\, e\, dF(e) = 0, \tag{9}$$

where F is the distribution function of e. Note that (9) is a necessary and sufficient condition, as (5) implies that

$$H'(\alpha) < 0. \tag{10}$$

What happens to the demand for risky assets—which we identify with α—as e becomes riskier? Once again, mean-variance analysis gives misleadingly general results which appear to confirm the common sense proposition that increasing the variability of the risky asset makes it less attractive to risk averse investors and reduces demand. If U is of the form $U(W) = aW - \frac{1}{2}bW^2$ then (9) becomes

$$0 = W_0[(a - bW_0)\, E(e) - \alpha b W_0 E(e^2)].$$

Thus, $\alpha = (a - bW_0)\, E(e)/E(e^2)\, bW_0$. As e becomes riskier, $E(e^2)$ increases and $E(e)$ remains constant; α must increase.

This is not true in general. If the distribution of e is changed from F to

[6] Arrow [1] has argued that normally relative risk aversion is increasing and—for low incomes—it is less than unity. Stiglitz [13] has questioned this assertion.

[7] Since in this subsection, we shall always assume that W_0 is constant, it makes no difference whether we formulate the problem as Tobin does in terms of the rate of return ($r = (W - W_0)/W_0$) or in terms of the value of terminal wealth. The interpretation of certain characteristics of the utility function (e.g., its elasticity) differs if utility is viewed as a function of r rather than of W. (In the former case, since r may be both positive and negative, U' cannot have constant elasticity.)

G, where F is less variable than G, then the new allocation parameter $\tilde{\alpha}$ satisfies $\int U'(W(\tilde{\alpha}))\, edG(e) = 0$. Thus, $\tilde{\alpha} \gtreqless \alpha$, as $\int U'(W(\alpha))\, edSe \gtreqless 0$, where $S = G - F$. If increasing risk is to decrease *all* risk-averse investors' demand for risky assets, then, if $V(e) = U'(W(\alpha))$ and if the points of increase of F and G are confined to (a, b), we must have

$$\int_a^b V(e)\, edS(e) \leqslant 0, \tag{11}$$

for all positive and decreasing V and all S satisfying the integral conditions of [11]; i.e., all S satisfying

$$\int_a^b S(e)\, de = \int_a^b edS(e) = 0 \tag{12a}$$

and

$$\int_a^t S(e)\, de \geqslant 0, \qquad \text{for all} \quad t \in (a, b). \tag{12b}$$

Using (12b) and the second mean value theorem of the integral calculus, we find the integral in (11) is equal to $[V(a) - V(b)]\, h(c)$ for some $c \in (a, b)$ where $h(c) = \int_a^c edS(e)$. A sufficient condition for (11) therefore is that

$$h(c) \leqslant 0 \tag{13}$$

for all $c \in (a, b)$. This is also a necessary condition. Suppose there is a $\bar{c}$ such that $h(c) > 0$. For (11) to hold we must have $\int_a^b V(e)\, dS(e) < 0$ for all positive and decreasing V. One such V is given by

$$\bar{V}(e) = \begin{cases} V_1 & \text{for} \quad a \leqslant e < \bar{c} \\ V_2 & \text{for} \quad \bar{c} \leqslant e \leqslant b, \end{cases}$$

where $V_1 > V_2 > 0$. Then $\int_a^b \bar{V}(e)\, dS(e) = (V_1 - V_2)\, h(\bar{c}) > 0$.

Still the presumption is that increasing risk decreases the demand for risky assets. It is possible both to exhibit increasing concave utility functions for which this is always true, and to show that *no* such utility function has the property that increasing risk always increases α. An increase in risk will always decrease demand for risky assets if $Z(e) = eU'(W(\alpha))$ is concave in e. Since

$$Z''(e) = [(1 - R + W_0 A)\, U'' + (W_0 A' - R')\, U']\, W_0 a,$$

where $R = -U''W/U'$ and $A = -U''/U'$, the measure of absolute risk aversion, a sufficient condition for an increase in uncertainty to increase the allocation to the safe asset is that relative risk aversion be less than or equal to unity and nondecreasing, and that absolute risk aversion be nonincreasing. The Bernoulli or logarithmic utility function as well as all constant relative risk aversion functions where the degree of risk aversion is less than one satisfies these conditions; other members of the constant relative risk aversion class do not.

No risk averse investor will always increase his holdings of risky assets when their variability increases. To prove this, it suffices to exhibit an increase in risk which satisfies (13). One such is

$$\hat{S}(e) = \begin{cases} S_1 > 0 & \text{for } \; e_1 \leqslant e < 0 \\ S_2 > 0 & \text{for } \; 0 \leqslant e \leqslant e_2 \\ 0 & \text{otherwise} \end{cases},$$

where the e_i and S_i are chosen to satisfy (12b). This kind of reasoning can be used to demonstrate that increases in the riskiness of any member of a rather important class of assets, bets, always lowers the demand for it.

A bet is an investment which can have only two outcomes. If all goes well, it pays off $1 + E(e) + \gamma$ per dollar invested. Otherwise, only $1 + E(e) - \lambda$ is returned. We call the gain from the bet γ, and the loss, λ, while $E(e)$ is the expected rate of return. A "fair" bet is one for which $E(e) = 0$. As risk averters will not take fair bets and as a bet for which $E(e) - \lambda \geqslant 0$ represents no risk at all, the portfolio problem is only of interest if

$$E(e) - \lambda < 0 < E(e). \tag{14}$$

If P is the probability of winning the bet then $P\gamma - (1 - P)\lambda = 0$ or $\lambda = [P/(1 - P)]\gamma$. All bets with the same rate of return may be represented by points in (P, γ) space (Fig. 1).

Figure 1 permits a graphic depiction of the difference between the use of our criterion and the variance criterion to rank risky prospects. The point B represents a typical bet with a possible gain of γ_B and a probability of attaining it of P_B. A straightforward application of the definition of a mean preserving spread, or of the integral conditions of [11], shows that a bet C with gain γ_C and loss λ_C is riskier than B if and only if

$$\gamma_C > \gamma_B \tag{15a}$$

and

$$\lambda_C > \lambda_B\,. \tag{15b}$$

Clearly, all bets to the right of B satisfy (15a). The solid line in Fig. 1

represents the locus of all bets with loss λ_B. The slope of this line is $d\gamma/dP|_{\lambda=\lambda_B} = -\gamma/P(1-P)$. Only bets above it satisfy (15b). Thus points in regions III and IV are riskier than B. Similarly, those in I and VI are safer. Those in regions II and V cannot be compared with B by our criterion.

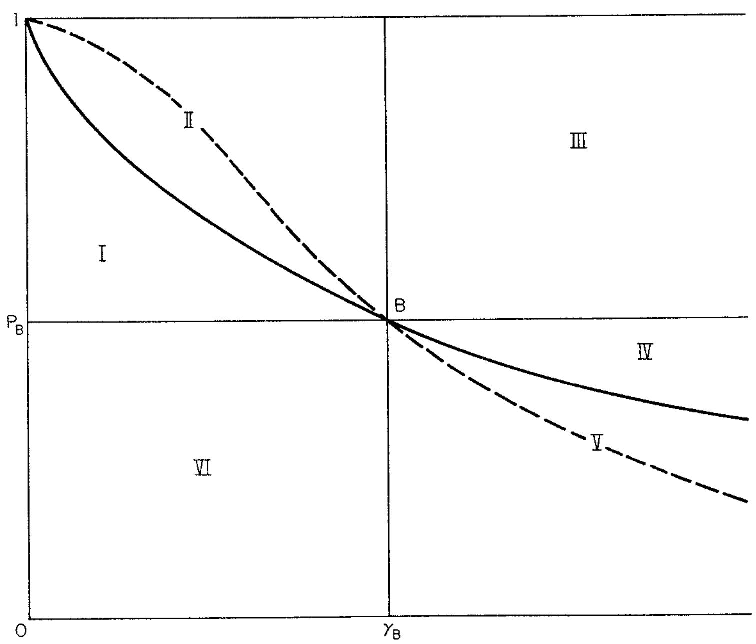

FIG. 1

The variance criterion ranks all the bets in Fig. 1. The variance of a bet is $\sigma^2 = P\gamma^2 + (1-P)\lambda^2 = \gamma^2(P/(1-P))$. The dotted line in the figure is the locus of all bets with the same variance as B. All bets above it have greater variance than B; all below it less. Its slope is

$$\frac{d\gamma}{dP}\bigg|_{\sigma^2=\sigma_B^2} = -\frac{1}{2}\frac{\gamma}{P(1-P)} = \frac{1}{2}\frac{d\gamma}{dP}\bigg|_{\lambda=\lambda_B}.$$

The dotted line cuts the broken line from below at B and lies entirely in regions II and V. There are bets in each region which have greater variance than B but which are not riskier than B.

It is now quite straightforward to use condition (13) to demonstrate that if C is riskier than B a risk averter will always hold more of B than C in his portfolio.[8] Define e_i, $i = 1, 2, 3, 4$, as follows:

$$e_1 = E(e) - \lambda_C \qquad e_2 = E(e) - \lambda_B$$
$$e_3 = E(e) + \gamma_B \qquad e_4 = E(e) + \gamma_C .$$

Then, combining (14) and (15) we see that

$$e_1 < e_2 < 0 < e_3 < e_4 . \tag{16}$$

Since B and C have the same expected return

$$(1 - P_C)\, e_1 - (1 - P_B)\, e_2 - P_B e_3 + P_C e_4 = \sum_{i=1}^{4} S_i e_i = 0. \tag{17}$$

For this case, (13) becomes

$$h_k = \sum_{i=1}^{k} S_i e_i \leqslant 0; \qquad k = 1, 2, 3, 4. \tag{18}$$

Since $e_1 < 0$ and $S_1 = (1 - P_C) > 0, h_1 < 0$; (17) states that $h_4 = 0$; furthermore, $h_3 = h_4 - P_C e_4 = -P_C e_4 < 0$. It remains to show that $h_2 \leqslant 0$, or that

$$(1 - P_C)\, e_1 - (1 - P_B)\, e_2 = -(P_C e_4 - P_B e_3) \leqslant 0. \tag{19}$$

But, either $P_B \leqslant P_C$ or $P_B > P_C$. If $P_B \leqslant P_C$, then (16) implies the *RHS* of (19) is negative. If $P_B > P_C$, then $(1 - P_B) < (1 - P_C)$ which with (16) implies the *LHS* of (19) is negative.

C. *A Combined Portfolio-Savings Problem*

Levhari and Srinivasan[9] [5] have recently analyzed the optimal portfolio and savings decision of an individual who wishes to maximize the expected value of the discounted utility of consumption,

$$E \sum_{t=0}^{\infty} (1 - \delta)^t\, U(C_t), \tag{20}$$

[8] The comparison is, of course, only between a portfolio consisting of B and a safe asset and of C and a safe asset.

Another class of changes which satisfies (13) is given by the partial ordering $\leqslant_m$ mentioned in Part I [11]. If Y is a random variable with mean $\bar{Y}$ and X is a random variable with the same distribution as the mixture $\alpha Y + (1 - \alpha)\bar{Y}, 0 \leqslant \alpha \leqslant 1$, then we write $X \leqslant_m Y$. It is straightforward to show that if $X \leqslant_m Y$, then the difference between the c.d.f.'s of the two distribution functions satisfies (13). We have analyzed the case of bets at length because of its greater importance and because of the intrinsic interest of the diagrammatic representation of different concepts of increasing risk which is given in Fig. 1 below.

[9] Henceforth *LS*. For other discussions of dynamic portfolio problems, see [4].

where δ is the discount rate and C_t is consumption at time t, subject to the stochastic constraints

$$W_{t+1} = (W_t - C_t)\, r^t; \qquad W_t \geqslant C_t \geqslant 0. \tag{21}$$

W_t is the individual's wealth at t and $r^t - 1$ is the stochastic rate of return on the investment of $(W_t - C_t)$ in period t. At time t, the individual may invest either in asset 1 with a rate of return of $r_1{}^t - 1$, or in asset 2 with a rate of return of $r_2{}^t - 1$. If α is the fraction invested in the first asset and $(1 - \alpha)$ is the fraction invested in the second, then r^t is given by

$$r^t = \alpha r_1{}^t + (1 - \alpha)\, r_2{}^t. \tag{22}$$

It is readily apparent that in the special case where (a) the $r_j{}^t$ are independently and identically distributed in time[10] and (b) the utility function has constant relative risk aversion

$$U(C) = \frac{C^{1-a}}{1-a} \quad \begin{matrix} a > 0 \\ a \neq 1 \end{matrix} \qquad \text{or} \quad U(C) = \ln C, \tag{23}$$

the savings decision, the determination of C_t (given W_t) is independent of the portfolio decision (the choice of α). Moreover, the savings rate (C_t/W_t) is a constant, independent of W_t, i.e., optimal behavior entails following the rule of saving a constant proportion of wealth

$$C_t = (1 - s)\, W_t \tag{24}$$

for all $W_t > 0$. *LS* show that for any choice of α, the optimal savings ratio $s(\alpha)$[11] must satisfy

$$s(\alpha)^a = (1 - \delta)\, E(r^{1-a}), \tag{25}$$

where

$$r = \alpha r_1 + (1 - \alpha)\, r_2\,. \tag{26}$$

They also show that, for any initial wealth W_0, when s is chosen to satisfy Eq. (25), the expected discounted utility of an optimal savings program is given by

$$V(W_0, \alpha) = \frac{1}{1-a} E\Sigma(1 - \delta)^t\, C_t^{1-a} = \frac{(1 - s(\alpha))^{-a}\, W_0^{1-a}}{1-a}\,. \tag{27}$$

[10] This assumption allows us to omit the superscript on $r_j{}^t$ in most of the following discussion.

[11] We assume $a \neq 1$. The case of $a = 1$ needs special treatment detailed below. *LS* do not appear to recognize that their analysis does not apply when $a = 1$.

It follows from Eqs (25) and (26) that s is a function of α, and so is the *RHS* of Eq. (27). Hence, α is selected to maximize (27). For an interior solution, $0 < \alpha < 1$, we must have

$$0 = \frac{\partial V}{\partial \alpha} = \frac{\partial V}{\partial s} \frac{\partial s}{\partial \alpha},$$

where

$$\frac{\partial V}{\partial s} = \frac{a(1-s)^{-(1+a)} W_0^{1-a}}{1-a} \gtrless 0 \qquad \text{as} \quad a \lessgtr 1.$$

It can be shown that in the neighborhood of α^*, where $\partial s/\partial \alpha = 0$, $(1-a)\, \partial^2 s/\partial \alpha^2 < 0$ for $a \neq 1$, so that the only critical points of (27) are relative maxima, implying a single critical point.[12] Thus, the condition $\partial s/\partial \alpha = 0$ determines α^*.

It is natural to ask what effect an increase in variability of the return of one of the assets will have on portfolio choice and on savings. *LS* have shown that, under special circumstances, an increase in the variance of one asset, with the mean held constant, decreases the proportion invested in that asset. To see whether this is true in general, we note that for $a \neq 1$, $\partial s/\partial \alpha = 0$ if, and only if,

$$E(r_1 - r_2)(\alpha r_1 + (1-\alpha)\, r_2)^{-a} = 0. \tag{28}$$

The effect of an increase in variability will be unambiguous if, and only if, $H(r_1) = (r_1 - r_2)(\alpha r_1 + (1-\alpha)\, r_2)^{-a}$ is convex or concave in r_1. But

$$\begin{aligned} H''(r_1) = &- \{\alpha(1-a)\, r_1 + r_2(\alpha(1+a) + 2(1-\alpha))\} \\ &\times [a\alpha(\alpha r_1 + (1-\alpha)\, r_2)^{-(a+2)}]. \end{aligned}$$

This expression is negative for $a < 1$ (assuming $\alpha > 0$), so that an increase in the variability of r_1 reduces the demand for this risky asset. The same result holds if $a = 1$ although the argument is more complicated. For $a > 1$, the sign of $H''(r_1)$ is ambiguous and an increase in variability could have the opposite effect.[13]

The effects of a change in variability on the savings rate are much easier to analyze than the effects on portfolio allocation. By (25), what happens to s depends on what happens to

$$E\{\alpha r_1 + (1-\alpha)\, r_2\}^{1-a}, \tag{29}$$

[12] *LS*'s argument is incorrect here as their evaluation of $\partial^2 S/\partial \alpha^2$ is in error.

[13] This generalizes the result obtained for the lognormal distribution in [5].

where α is chosen optimally. But, since, using Eq. (28),

$$\frac{\partial E(\alpha r_1 + (1-\alpha) r_2)^{1-a}}{\partial \alpha} = (1-a) E(r_1 - r_2)(\alpha r_1 - (1-\alpha) r_2)^{-a} = 0,$$

we need consider only what happens to (29) at any fixed value of α. But,

$$\frac{\partial^2(\alpha r_1 + (1-\alpha) r_2)^{1-a}}{\partial (r_1)^2}$$
$$= \alpha^2(1-a) a[\alpha r_1 + (1-\alpha) r_2]^{-(1+a)} \gtrless 0 \qquad \text{as} \quad a \lessgtr 1,$$

from which it immediately follows that an increase in the variability of r increases the savings rate if $a < 1$ and decreases it if $a > 1$.

The case of the logarithmic utility function ($a = 1$) needs some special attention. We can write the functional equation for this case as

$$V(W_0) = \ln(1-s) W_0 + (1-\delta) EV(sW_0 r). \tag{30}$$

The solution of Eq. (30) is[14]

$$V(W_0) = \frac{1}{\delta} \ln(1-s) W_0 + \frac{1-\delta}{\delta^2} [\ln(s) + E \ln r] \tag{31}$$

and the optimal value of s is given by[15]

$$s = 1 - \delta \tag{32}$$

[14] $V(W_0)$ may be calculated directly as follows: If s is a constant, utility from a single realization of $r^i (r^0 \equiv 1$, by convention) is

$$V(W_0) = \ln(1-s)W_0 + (1-\delta)\ln(1-s)W_0 s r^1 + (1-\delta)^2 \ln(1-s) W_0 s^2 r^1 r^2 + \cdots$$
$$= \textstyle\sum_{i=0}^{\infty} (1-\delta)^i \ln(1-s) W_0 s^i \prod_{j=0}^{i} r^i$$
$$= \textstyle\sum_{i=0}^{\infty} (1-\delta)^i \ln(1-s) W_0 + \sum_{i=0}^{\infty} (1-\delta)^i \, i \ln s + \sum_{i=0}^{\infty} (1-\delta)^i \ln \prod_{j=0}^{i} r^j .$$

Thus,

$$EV(W_0) = \frac{1}{\delta} \ln((1-s)W_0) + \frac{(1-\delta)}{\delta^2} \ln s + \frac{(1-\delta)}{\delta^2} E \ln r.$$

The second term is evaluated by observing that if

$$f(x) = \frac{1}{1-x} = \sum_{i=0}^{\infty} x^i, f'(x) = \sum_{i=0}^{\infty} i x^i = \frac{1}{(1-x)^2}.$$

Since the r^j are independently and identically distributed, $E \ln \prod_{j=0}^{i}(r^j) = iE \ln r$. Hence,

$$E \sum_{i=0}^{\infty} (1-\delta) \ln \prod_{j=0}^{i} r^j = E \sum_{i=0}^{\infty} (1-\delta)^i \, i \ln r = \frac{1-\delta}{\delta^2} E \ln r.$$

[15] If s is chosen to maximize (31), then $1/(1-\delta) = (1-\delta)/\delta^2 s$ from which (32) follows.

independent of r. To find the optimal portfolio allocation (the optimal value of α), we maximize $V(W_0)$ with respect to α; i.e., from Eq. (31), we must maximize $E \ln r$. Thus, α satisfies

$$E \left\{ \frac{r_1 - r_2}{\alpha r_1 + (1 - \alpha) r_2} \right\} = 0. \tag{33}$$

Since

$$\frac{d^2 E \ln r}{d\alpha^2} = -E \left\{ \frac{(r_1 - r_2)}{\alpha r_1 + (1 - \alpha) r_2} \right\}^2 < 0,$$

(33) is both necessary and sufficient for an interior solution. Since

$$\frac{r_1 - r_2}{\alpha r_1 + (1 - \alpha) r_2}$$

is a concave function of r_1, an increase in the variability of r_1 always increases the proportion of the portfolio allocated to r_2 but leaves the savings rate unaffected.

D. *A Firm's Production Problem*[16]

Consider a firm whose output Q next period is uncertain (e.g., a public utility which must meet all demands at a fixed price). It wishes to minimize the expected cost of producing Q. Q is produced by a two-factor concave production function $Q = P(K, L)$, where K is, say, capital, a factor which cannot be varied in the short run, and L is, say, labor, the variable factor. What happens to expected costs as Q becomes more variable? If r is the cost of capital and w that of labor, expected costs are given by

$$E[rK + wL(K, Q)] = rK + wE[L(K, Q)], \tag{34}$$

where $L(K, Q)$ is the labor required to produce the given output Q with capital K. Since F is concave, it is easy to show that $L(K, Q)$ is convex in Q, for any given K. Hence, an increase in variability of Q always leads to an increase in expected cost.

A somewhat more difficult problem is: What happens to the optimum level of K? Not surprisingly, the answer depends on the elasticity of substitution between K and L. We choose K to minimize expected costs. From (34), the first order conditions may be written

$$\frac{r}{w} = E \frac{\partial L(Q, K)}{\partial K},$$

[16] See [10] for a more complete analysis of this problem.

i.e., the factor-price ratio must be equal to the average marginal rate of substitution. Let us assume that the production function has constant elasticity of substitution. Then[17]

$$Q = (\delta K^{\rho} + (1 - \delta) L^{\rho})^{1/\rho}$$

$$\frac{\partial L}{\partial K} = \frac{\delta}{1-\delta}\left(\frac{K}{L}\right)^{\rho-1} = \left(\frac{\delta}{1-\delta}\right)\left(\frac{Q^{\rho} - \delta K^{\rho}}{1-\delta}\right)^{(1-\rho)/\rho} K^{\rho-1}$$

$$\frac{\partial^2(\partial L/\partial K)}{\partial Q^2} = \left[\frac{\delta K^{\rho-1}}{(1-\delta)^{1/\rho}}(1-\rho)\, Q^{\rho-2}(Q^{\rho} - \delta K^{\rho})^{(1-3\rho)/\rho}\right] \times (-\rho Q^{\rho} + (1-\rho)\,\delta K^{\rho}).$$

A sufficient condition for convexity is that $\rho \leqslant 0$, i.e., that the elasticity of substitution be less than or equal to unity. Thus, if the elasticity of substitution is less than or equal to unity, the optimal level of K increases with an increase of variability in Q.

To show that for other production functions K may decrease with an increase in variability of output, consider the extreme case of a constant elasticity of substitution production function with infinite elasticity:

$$Q = bK + aL.$$

If the capital stock is given by K, expected costs are given by

$$rK + \frac{w}{a}\int_{bK}^{\infty} (Q - bK)\, dG(Q),$$

where $G(Q)$ is the distribution function for Q. Expected cost minimization requires (for an interior solution)

$$r - \frac{wb}{a}(1 - G(bK)) = 0,$$

so that

$$K = \frac{G^{-1}(1 - (ar/wb))}{b}.$$

Whether K increases or decreases depends solely on whether $G^{-1}(1 - (ar/wb))$ increases or decreases (see Fig. 2) or, equivalently, whether the probability that Q will be greater than bK (the "capacity" of the original capital stock) increases or decreases.

[17] If $L > 0$; $\partial L/\partial K = 0$ otherwise.

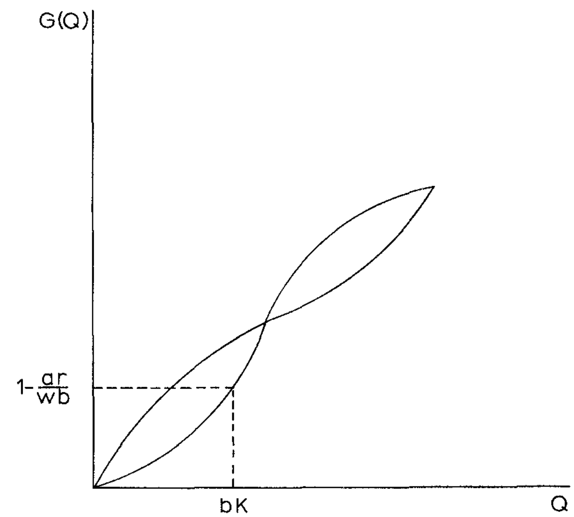

FIG. 2

E. *A Multi-stage Planning Problem*[18]

Consider a simple economy in which the final consumption good is produced by labor and an intermediate commodity y:

$$Q = P(L_2, y),$$

while y is produced by labor alone:

$$y = M(L_1).$$

The economy faces an overall labor constraint L, so

$$L_1 + L_2 = L.$$

In the absence of uncertainty, maximization of Q simply requires

$$P_1 = P_2 M'.$$

Assume that there is uncertainty associated with the production of y:

$$y = M(L_1) + e,$$

where e has mean zero and distribution function F. We wish to maximize EQ; we require

$$E[P_1 - P_2 M'] = 0.$$

[18] This problem was posed to us by M. Weitzman.

If e becomes more variable, what happens to the allocation of labor between the two sectors depends on the sign of

$$P_{122} - M'P_{222} .$$

Assume that P is a constant elasticity of substitution production function: $P = (\delta L_2{}^\rho + (1 - \delta)\, y^\rho)^{1/\rho}$. Then

$$P_{122} = \frac{A}{L_2}((1 - \rho)\, \delta L_2{}^\rho + \rho(1 - \delta)\, y^\rho)$$

$$P_{222} = \frac{A((\rho - 2)\, \delta L_2{}^\rho - (1 + \rho)(1 - \delta)\, y^\rho)}{y},$$

where

$$A = \delta(1 - \delta)(\rho - 1)\, L_2{}^\rho y^{\rho-2}(\delta L_2{}^\rho + (1 - \delta)\, y^\rho)^{(1-3\rho)/\rho} < 0.$$

If $1 \geqslant \rho \geqslant 0$, i.e., the elasticity of substitution is greater than or equal to unity, $P_{122} \leqslant 0$ and $P_{222} \geqslant 0$, so L_2 decreases and L_1 increases; more labor is allocated to the "earlier" stage of production (to producing y).

Consider the other extreme case, where Q is produced by a fixed-coefficients production function, $Q = \min(L_2 , y)$. Then

$$E(Q) = \int_{-\infty}^{L_2-M(L_1)} [M(L_1) + e]\, dF(e) + L_2(1 - F(L_2 - M(L_1)))$$

$$= \int_{\infty}^{\bar{L}-L_1-M(L_1)} [M(L_1) + e]\, dF(e) + (\bar{L} - L_1)(1 - F(\bar{L} - L_1 - M(L_1)))$$

so that maximization of EQ requires

$$[M'(L_1) + 1]\, F(\bar{L} - L_1 - M(L_1)) = 1.$$

The second-order conditions are satisfied, since $M''F - f(M' + 1)^2 < 0$, where f is the density function corresponding to F; hence, there is a unique maximum. Whether L_1 increases or decreases depends solely on whether $F(L - L_1 - M(L_1))$ increases or decreases, i.e., whether the probability that (at the old allocation) the y constraint will be binding is increased or decreased; either is clearly possible. Note that if y is also produced by a constant returns-to-scale production function

$$y = L_1 ,$$

then the optimal value of L_1 is simply given by

$$F(L - 2L_1) = \tfrac{1}{2}.$$

So what happens to L_1 depends completely on whether the *median* of e increases or decreases.

F. *Choice of Output Level for a Competitive Firm*

In the examples considered so far, the conditions we have obtained under which it is possible to make unambiguous statements about the effects of increases in variability have been essentially identical to those obtained earlier in comparisons between safe and risky situations. There are, however, problems in which the latter comparisons can be made under conditions weaker than the former. In the following example, we can, for instance, make unambiguous statements even when the first-order condition is neither concave nor convex.

Consider a competitive firm which must decide today on the level of output tomorrow, although the price p of output Q is uncertain. It wishes to maximize expected utility of profits $U(\pi)$, where U is concave[19] and

$$\pi = pQ - C(Q),$$

where $C(Q)$, the cost function, is convex. A necessary and sufficient condition for an optimum is that

$$\frac{EU'p}{EU'} = C'(Q^*). \tag{35}$$

If the producer is risk neutral, or if there is no variablility in p, profit maximization requires that price equals marginal cost, $Ep = C'(\hat{Q})$. $Q^* \gtrless \hat{Q}$ as $EU'p/EU' \gtrless Ep$, i.e., as $[E(U' - EU')(p - E(p))] \gtrless 0$. But since $U'' < 0$, $U'(p) \gtrless U'(E(p))$, as $p \lessgtr E(p)$; so

$$E[(U' - EU')(p - E(p))] = E[U' - U'(E(p))(p - E(p))] < 0.$$

Hence, there is always less output under uncertainty than under certainty.

Not surprisingly, the comparative statistics of the behavior of this firm differ significantly from that of the competitive firm with no uncertainty:

(a) In the absence of uncertainty, a proportional profits tax at rate t leaves output unchanged. Here it will increase or decrease output as relative risk aversion is increasing or decreasing. It is easy to see that dQ/dt has the same sign as $ERU'(p - C'(Q^*))$; from (35) this last quantity has the same sign as R' (once again, $R = -U''\pi/U'$ is the measure of relative risk aversion).

(b) In the absence of uncertainty, a (uniform) upward shift in the total cost curve leaves output unchanged. Here, if

$$C(Q, \tau) = C(Q) + \tau,$$

[19] For a discussion of the case of constant absolute risk aversion, see [6]. Except under very stringent conditions, how one ought to describe the behavior of a competitive firm in the presence of uncertainty remains an open question.

then

$$\frac{dQ}{d\tau} \gtreqless 0, \qquad \text{as} \quad -EU''(p - C') = EA(p - C')\, U' \gtreqless 0,$$

where $A = -U''/U'$ is the measure of absolute risk aversion. But $EA(p - C')\, U'$ has the same sign as A'. Thus output increases or decreases as absolute risk aversion is increasing or decreasing.

3. Choosing Probability Distributions

The following examples show how our definition of variability and our basic theorem on the equivalence of the three alternative approaches discussed in [11] may be applied to prove some general theorems about situations where one must choose a probability distribution from among a set of possible probability distributions:

A. *Diversification theorem.*[20]

Assume an individual can purchase shares of two[21] securities whose value next period (per dollar invested) is described by identical but independent distributions. How should he allocate his given initial wealth, i.e., how should he choose b to maximize[22]

$$EU(W) = EU((be_1 + (1 - b)\, e_2)\, W_0),$$

where U is a concave function? We prove that independent of the utility function, b should be set at $\frac{1}{2}$. We can write

$$y_b = (be_1 + (1 - b)\, e_2)\, W_0 = y_{1/2} - (b - \tfrac{1}{2})(e_1 - e_2)\, W_0 .$$

Note that

$$E(e_1 - e_2 \mid y_{1/2}) = 0,$$

i.e., the expected value of the difference between two identically distributed independent random variables, given only that their sum be a particular number, is zero. Since y_b has the same distribution as $y_{1/2}$ plus a random variable whose expectation conditional on $y_{1/2}$ is zero, by Theorem 2 of [11], y_b is more variable than $y_{1/2}$ for all $b \neq \frac{1}{2}$, i.e., $y_{1/2}$ is preferred to y_b by all individuals with concave utility functions.

B. *The Rao–Blackwell Theorem.*[23]

Suppose a sample of random variables $x = (x_1, ..., x_n)$ is generated by

[20] See [12] for an alternative proof and general discussion of this theorem.
[21] The generalization to the case of n securities is straightforward.
[22] We assume that $E(e_i)$ exists and is finite so EU exists and is finite.
[23] Concepts and notation are borrowed from Ferguson [3].

distribution depending on an unknown parameter θ. An estimating procedure for θ is a mapping $d(x)$ from the sample x to the real line. Often a statistician tries to find an estimating procedure which minimizes the expected value of a convex loss function $L(d(x))$. The Rao–Blackwell theorem states that for any estimator $d(x)$, and any convex L, if there is a sufficient statistic T for θ, there is an estimator $d^*(x)$ at least as good as $d(x)$ in the sense that $EL(d^*(x)) \leqslant EL(d(x))$.

To prove this, it follows from Theorem 2 of Part I that it is clearly necessary and sufficient to find a $d^*(x)$ such that $d^*(x) \leqslant_a d(x)$.

For every T let $d^*(x) = E(d(x)| \; T)$. Then consider the *r.v.* z defined by $d(x) = d^*(x) + z$. Clearly, $E(z \mid T, x) = E(z \mid d^*(x)) = 0$ and $d^*(x) \leqslant_a d(x)$ as was to be shown.

References

1. K. J. Arrow, "Aspects of a Theory of Risk Bearing," Yrjö Jahnsson Lectures, Helsinki, 1965.
2. A. B. Atkinson, On the measurement of inequality, *J. Econ. Theor.* **2** (1970), 244–263.
3. T. S. Ferguson, "Mathematical Statistics," Academic Press, New York/London, 1967.
4. F. H. Hahn, Savings under uncertainty, *Rev. Econ. Studies* **37** (1970), 21–24.
5. D. Levhari and T. N. Srinivasan, Optimal savings under uncertainty, *Rev. Econ. Studies* **36** (1969), 153–164.
6. J. J. McCall, Competitive production for constant risk utility functions, *Rev. Econ. Studies* **31** (1964), 417–419.
7. J. A., Mirrlees, "Optimum Accumulation under Uncertainty," December 1965, Cambridge, unpublished.
8. E. S. Phelps, The accumulation of risky capital: A sequential utility analysis, *Econometrica* **30** (1962), 729–743.
9. J. W. Pratt, Risk aversion in the small and in the large, *Econometrica* **32** (1964), 122–136.
10. M. Rothschild, "Changing Demand: Its Costs and Consequences," paper presented to the Econometric Soc., Evanston, December, 1968.
11. M. Rothschild, and J. E. Stiglitz, Increasing risk I: A definition, *J. Econ. Theory* **2** (1970), 225–243.
12. P. A. Samuelson, General proof that diversification pays, *J. Finan. Quant. Anal.* **2** (1967), 1–13.
13. J. E. Stiglitz, The effects of wealth, income and capital gains taxation on risk taking, *Quart. J. Economics* **83** (1969), 263–283.
14. J. E. Stiglitz, A consumption oriented theory of the demand for financial assets and the term structure of interest rates, *Rev. Econ. Studies*, **37**, 321–351.
15. James Tobin, The theory of portfolio selection, *in* "The Theory of Interest Rates" (E. Hahn and F. Brechling, Eds.), Macmillan, London, 1965.
16. James Tobin, Comment on Borch and Feldstein, *Rev. Econ. Studies* **36** (1969), 13–14.