

Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers

Cowles Foundation

5-1-1969

Increasing Risk: A Definition and Its Economic Consequences

Michael Rothschild

Joseph E. Stiglitz

Follow this and additional works at: <https://elischolar.library.yale.edu/cowles-discussion-paper-series>



Part of the [Economics Commons](#)

Recommended Citation

Rothschild, Michael and Stiglitz, Joseph E., "Increasing Risk: A Definition and Its Economic Consequences" (1969). *Cowles Foundation Discussion Papers*. 508.

<https://elischolar.library.yale.edu/cowles-discussion-paper-series/508>

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

**Box 2125, Yale Station
New Haven, Connecticut .**

COWLES FOUNDATION DISCUSSION PAPER NO. 275

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

INCREASING RISK: A DEFINITION AND ITS ECONOMIC CONSEQUENCES

M. Rothschild and J.E. Stiglitz

May 27, 1969

ERRATA FOR CFDP NO. 275

INCREASING RISK: A DEFINITION AND ITS ECONOMIC CONSEQUENCES

p. 4. 2nd line from bottom. " $H(X, Z)$ " should be " $H(x, z)$ "

p. 6. line 7, "differs" not "differe"

p. 7. Section 2. line 6, "conditions" not "condition"

p. 10. Proof: line 4, "By (7)" not "By (5)"

p. 11. Should read:

"then¹

$$(11) \quad ||F_n - F|| + ||G_n - G|| \leq \frac{4}{n}$$

and if"

p. 15. last line "partial ordering \leq_u "

p. 18. first line should read: $E \underset{Z}{U}(X + Z) \leq \underset{Z}{U}(E(X + Z)) = U(X)$ "

p. 22. Section 6, 5th line should read: "for all $y \in [0, 1]$ "

Section (a), line 10 should read: " $(X \underset{V}{\leq} Y \text{ if } EX = EY \text{ and } EX^2 \leq EY^2)$ "

p. 23. Section IV., line 5 and line 7 should read: " $F \underset{u}{\leq} G$ "

p. 25. line 3, "two parameter"

p. 26. fn. 2, "The generalization to the case..."

p. 27. Equation (25) should read:

$$(25) \quad \int U(X, \alpha) dF(X)$$

ERRATA FOR CDDP NO. 275 (continued)

p. 27 Equation (25) should read

$$(26) \quad \int \frac{\partial V(X, \alpha)}{\partial \alpha} dF(X) = 0$$

p. 30 line 10 should read: $P_1 = P_2 M'$

INCREASING RISK: A DEFINITION AND ITS ECONOMIC CONSEQUENCES*

by

M. Rothschild and J.E. Stiglitz

I. Introduction

This paper attempts to answer two questions:

1. When is a random variable Y "more variable" than another random variable X ?
2. What effect does an increase in variability or uncertainty¹ have on economic behavior, e.g., does an increase in the riskiness of investment opportunities lead to more or less saving?

There seem to be at least five possible approaches to answering the first question.

* 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 D. Ragozin, P. Diamond, D. Wallace, and D. 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 [3]. R.M. Solow used a device similar to our Mean Preserving Spread (Section II, below) to compare lag structures in [9]. The problem of "stochastic dominance" is a standard one in the (statistics) operations research literature. For other approaches to the problem, see, for instance, [2]. [6, 8] have recently provided an alternative proof to our theorem 2.b and its converse (p.18 below).

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 synonymously.

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" than the original. More formally, suppose Y and X are related as follows:¹

$$(1.i) \quad Y = X + Z$$

where Z is a r.v. with the property that

$$(1.ii) \quad E(Z|X) = 0 \text{ for all } X.$$

That is, Y is equal to X plus an uncorrelated disturbance term ("noise.")

If X and Y are discrete r.v.'s (1) has another natural interpretation.

Suppose X is a lottery ticket which pays off a_i with probability p_i ;

$\sum p_i = 1$. 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 (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 , that is, if

$$(2) \quad E U(X) \geq E U(Y) \text{ for all concave } U$$

then surely it is reasonable to say that X is less risky than Y .²

1 This equation means that the r.v. Y has the same distribution as the r.v. $Y' = X + Z$ where X and Z are jointly distributed r.v.'s with distribution function $H(X, Z)$. David Wallace suggested that we investigate this concept of greater riskiness.

2 Strictly speaking, we should perhaps limit ourselves to considering concave functions with non-negative first derivatives. Since little is gained by this extra restriction, we shall not impose it. Note that since $U(X) = X$ and $U(X) = -X$ are both concave functions, (2) implies that X and Y have the same mean.

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. X is a mixture of Y and a sure thing.

Let Y be r.v. with the mean μ and let Y' be the r.v. which is equal to μ with probability 1. Then if $X = \lambda Y + (1-\lambda)Y'$, $0 \leq \lambda < 1$, X is less risky than Y .

5. 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 four approaches lead to a single definition of greater riskiness, different from that of the fifth approach. We shall demonstrate the equivalence as follows. In section II, it is shown that the third and fourth approaches lead 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 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." [17] In section V 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 on¹ $[0, 1] \times [-1, 1]$

¹ \times denotes cartesian product.

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:

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

where

$$(3.ii) \quad 0 \leq a \leq a+t \leq a+d \leq a+d+t \leq b \leq b+t \leq b+e \leq b+e+t \leq 1$$

and

$$(3.iii) \quad \beta e = \alpha d$$

Such a function is pictured in figure 2. It is easy to verify that

$$\int_0^1 s(x)dx = \int_0^1 xs(x) dx = 0. \quad \text{Thus if } f \text{ is a density function and if}$$

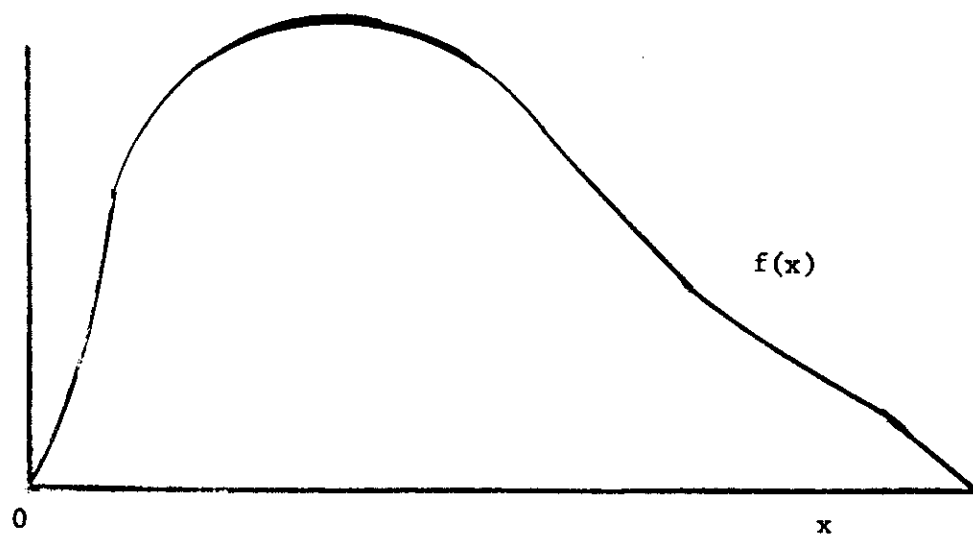


FIGURE 1

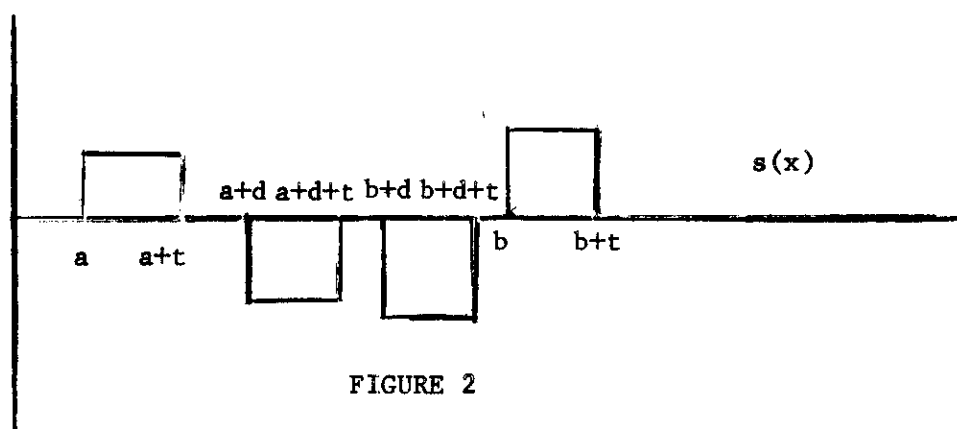


FIGURE 2

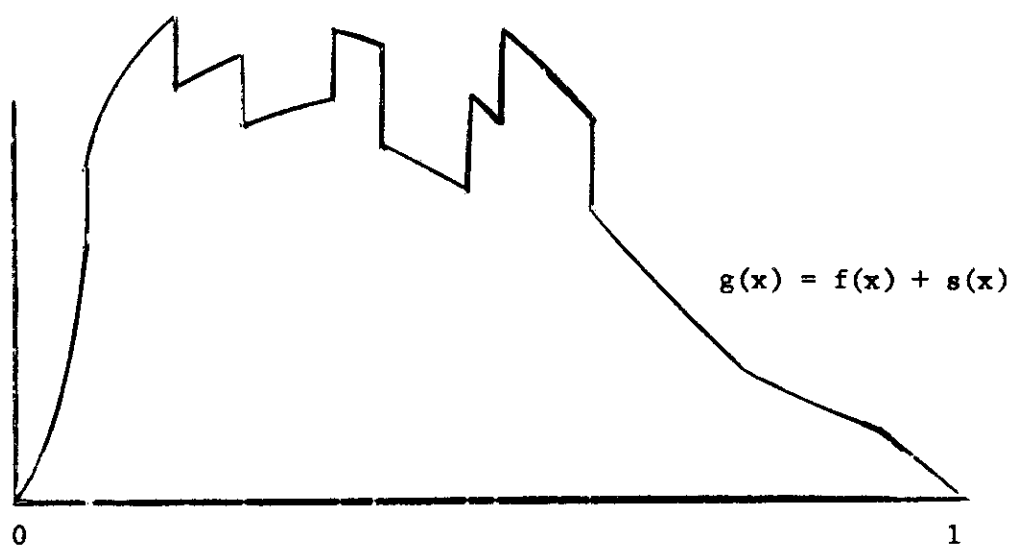


FIGURE 3

$$g = f + s, \text{ then } \int_0^1 g(x)dx = \int_0^1 f(x)dx + \int_0^1 s(x)dx = 1 \text{ and } \int_0^1 xg(x)dx$$

$$= \int_0^1 x(f(x) + s(x))dx = \int_0^1 xf(x)dx. \text{ It follows then that if } g(x) \geq 0$$

for all x^1 , g is a density function with the same mean as f . Adding a function like s to f shifts probability weight from the center to the tails. We shall call a function which satisfies (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 differences 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:

$$\Pr(X = \hat{a}_i) = \hat{f}_i; \quad \Pr(Y = \hat{a}_i) = \hat{g}_i$$

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

$$(4.1) \quad \gamma_1 = -\gamma_2 \geq 0 \text{ and } \gamma_4 = -\gamma_3 \geq 0$$

1 That is, if $f(x) \geq \alpha$ for $a + d < x < a + d + t$ and $f(x) \geq \beta$ for $b < x < b + t$.

Y has more weight in the tails than X and if

$$(4.ii) \quad \sum_{k=1}^4 a_k \gamma_k = 0$$

the means of X and Y will be the same. See Figure 4. If two discrete r.v's X and Y attribute the same weight to all but four points and if their differences satisfy (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, $s = g - f$ implies $S = G - F$ where

$$S(x) = \int_0^x s(u)du. \quad S, \text{ which is drawn in figure 5, has several interesting}$$

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 condition. First $S(0) = S(1) = 0$. Second, there is a z such that

$$(5) \quad S(x) \leq 0 \text{ if } x \leq z \text{ and } S(x) \geq 0 \text{ if } x > z$$

$$\text{Finally, if } T(y) = \int_0^y S(x)dx$$

then

$$(6) \quad T(1) = 0$$

$$\text{since } T(1) = \int_0^1 S(x)dx = \left[xS(x) \right]_0^1 - \int_0^1 x s(x)dx = 0.$$

Finally (5) and (6) together imply that

$$(7) \quad T(y) \geq 0; \quad 0 \leq y < 1$$

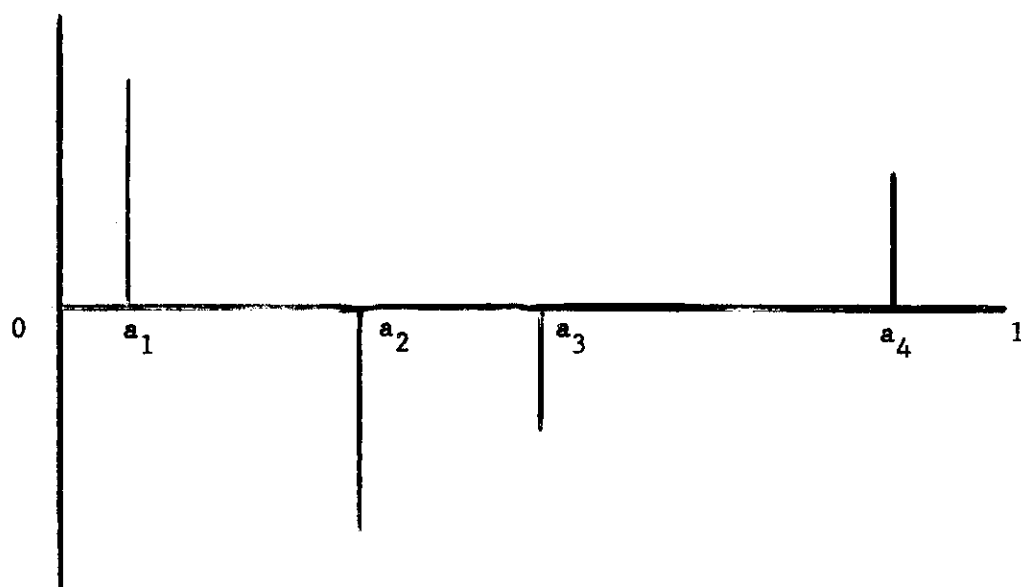


FIGURE 4

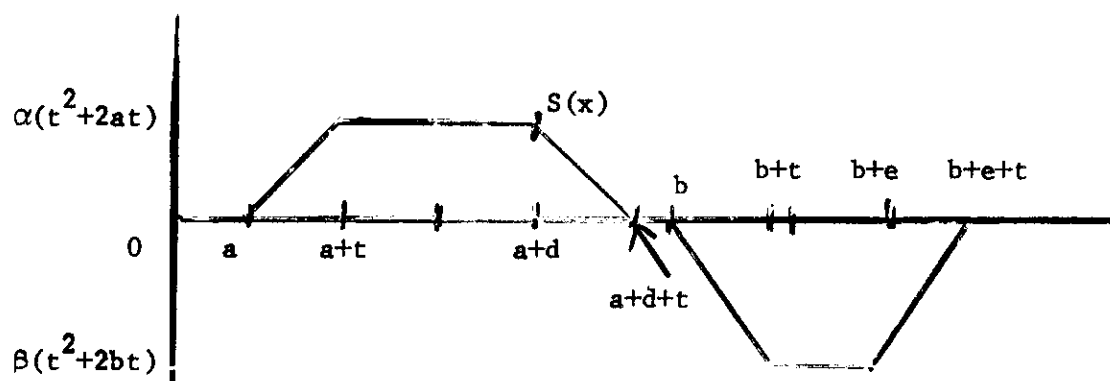


FIGURE 5

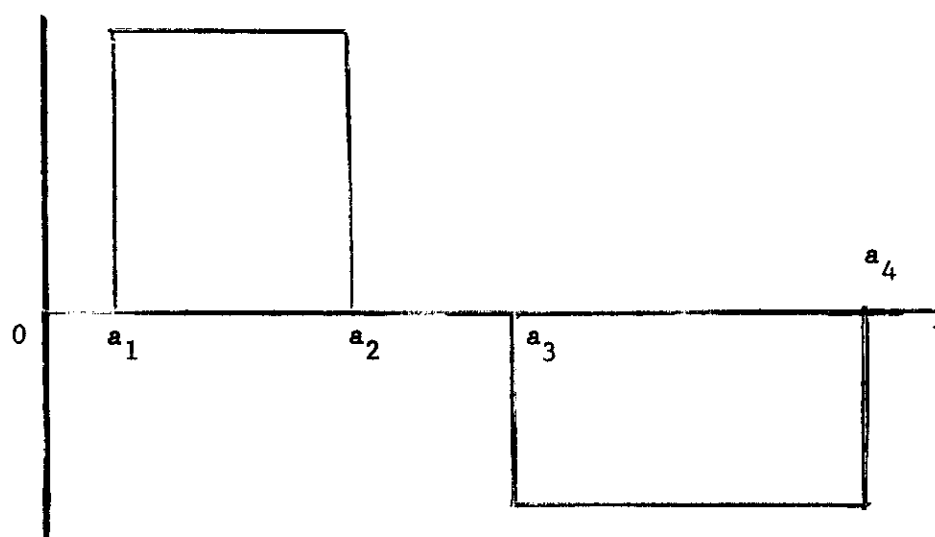


FIGURE 6

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 figure 6.

It should be noted that if X and Y are r.v's such that X is a mixture of Y and a sure thing (see I.4) above, then the difference of the c.d.f's corresponding to Y and X also satisfies (5), (6) and (7). See figures 7 and 8.

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.¹

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 $G - F$

¹ 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).

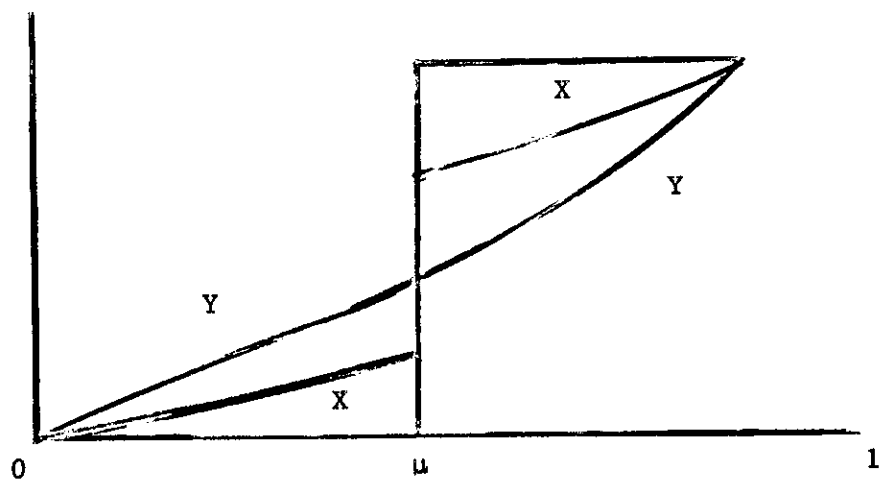


FIGURE 7

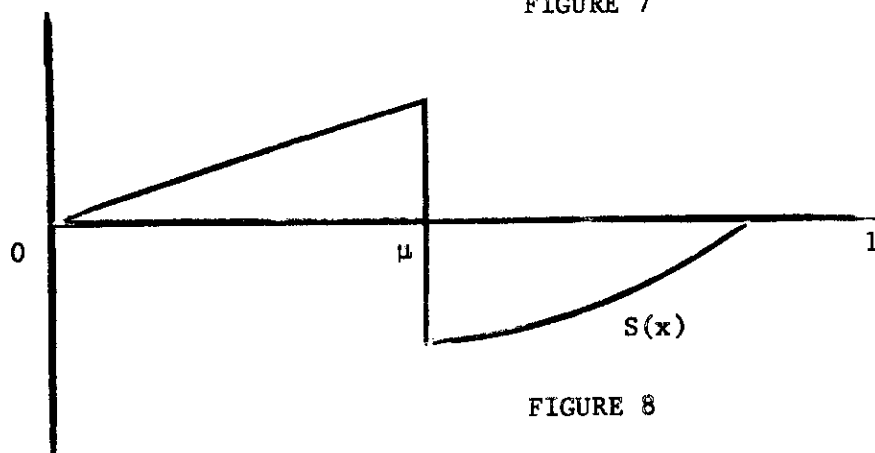


FIGURE 8

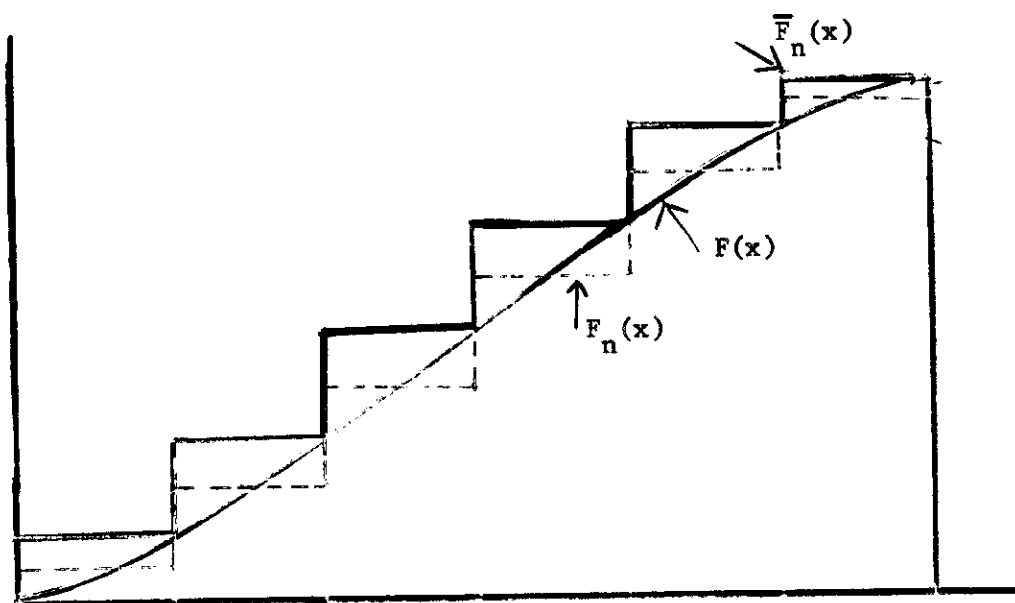


FIGURE 9

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 $F_n \rightarrow G$)¹ 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_n 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.

Theorem 1(b). If $F - G$ satisfies the integral conditions (6) and (7)), then there exist sequences F_n and G_n , $F_n \rightarrow F$, $G_n \rightarrow G$, 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),

1 Let $E(u) = \int_0^1 u(x) dG(x)$ and $E_n(u) = \int_0^1 u(x) dG_n(x)$. Then $F_n \rightarrow G$ if and only if $E_n(u) \rightarrow E(u)$ for all continuous functions u on $[0, 1]$. See Feller II, p. 243.

then there exists c.d.f's, F_0, \dots, 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 (6), $a_2 < a_3$. 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

$$(8) \quad \gamma_1(a_2 - a_1) \geq \gamma_2(a_4 - a_3)$$

or

$$(9) \quad \gamma_1(a_2 - a_1) < \gamma_2(a_4 - a_3)$$

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

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

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 $T(y) =$

$$\int_0^y (G(x) - F(x))dx. \text{ If}$$

$$(6) \quad T(y) \geq 0, \quad 0 \leq y \leq 1 .$$

and

$$(7) \quad T(1) = 0$$

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

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

$$(11) \quad ||F_n - F|| + ||G_n - G|| \leq \frac{4}{n}$$

then¹

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

$$(12) \quad T_n(y) \geq 0$$

$$(13) \quad T_n(1) = 0$$

Proof: We prove this by constructing F_n and G_n for fixed n .

¹ Condition (11) implies weak convergence. See Feller II, p. 243.

For $i = 1, \dots, n$ let $I_i = (\frac{i-1}{n}, \frac{i}{n})$. Let

$\bar{f}_i = F(\frac{i}{n})$ and define \bar{F}_n by $\bar{F}_n(x) = \bar{f}_i$ for $x \in I_i$. (See figure 9)

Since F is monotonic $\bar{F}_n(x) \geq F(x)$. It follows also from monotonicity

that $||\bar{F}_n - F|| \leq \frac{1}{n}$. 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 $||\hat{F}_n - \bar{F}_n|| \leq \frac{1}{n}$ and

$$||\hat{F}_n - F|| \leq ||\hat{F}_n - \bar{F}_n|| + ||\bar{F}_n - F|| \leq \frac{2}{n}$$

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 $||\hat{G}_n - G|| \leq \frac{2}{n}$.

For every i there exists $f_i \in F(I_i)$ and $g_i \in G(I_i)$ such

that $\frac{g_i - f_i}{n} = \int_{I_i} (G(x) - F(x))dx$. 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 $\hat{T}_n(y) \geq 0$. If

$y = \frac{j}{n}$ for $j = 0, 1, \dots, n$, then $\hat{T}_n(y) = T(\frac{j}{n}) \geq 0$ so we need only

examine the case where $y = \frac{j}{n} + \alpha$, $0 < \alpha < \frac{1}{n}$. Then, $\hat{T}_n(x) =$

$T_n(\frac{j}{n}) + \alpha (g_j - f_j)$. If $g_j > f_j$ both terms of the sum are positive.

If $g_j < f_j$ then $T(\frac{j}{n}) + \alpha(g_j - f_j) > T(\frac{j}{n}) + \frac{1}{n}(g_j - f_j) = T(\frac{j+1}{n}) \geq 0$.

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 .

A similar argument, which we omit, can be used to show that the integral conditions are also implied by defining Y to be more variable than X if X is a mixture of Y and a sure thing with the same mean as Y and exploring the implication of transitivity. This establishes the essential equivalence of the third and fourth approaches to the definition of increasing risk.

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

four concepts of increasing risk set out in Section I and prove their equivalence.

1. Partial orderings.

A partial ordering \leq_p on a set is a binary, transitive, reflexive and antisymmetric relation.¹ The set over which our partial orderings are defined is the set of distribution functions on $[0, 1]$. We shall use $F \leq_p G$ interchangeably with $X \leq_p Y$ where F and G are the c.d.f's of the r.v's X and Y .

2. Definition of \leq_I .

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

Lemma 3. \leq_I is a partial ordering.

Proof. It is immediate that \leq_I is transitive and reflexive. We need only demonstrate antisymmetry. Define S_1 and S_2 as follows:

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

Thus $S_1 + S_2 = 0$. Furthermore, if $T_1(x) = \int_0^x S_1(x)dx$, then

¹ A relation \leq_p is antisymmetric if $A \leq_p B$ and $B \leq_p A$ implies $A = B$.

$T_1(y) \geq 0$, since $F \leq_I G$ and $G \leq_I F$. Since $0 = \int_0^y (S_1(x) + S_2(x))dx =$

$T_1(y) + T_2(y) = 0$ and $T_1(y) \geq 0$, $T_1(y) = 0$. 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.¹

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 \leq_u .

We define the partial \leq_u corresponding to the idea that X

¹ We shall follow the convention of considering two distribution functions to be equal if they differ only on a set of measure zero.

is less risky than Y if every risk averter prefers X to Y as follows.

$F \stackrel{\leq}{u} G$ if and only if for every bounded concave function U , $\int_0^1 U(x) dF(x) \geq$

$\int_0^1 U(x) dG(x)$. It is immediate that $\stackrel{\leq}{u}$ is transitive and reflexive. That

$\stackrel{\leq}{u}$ is antisymmetric is an immediate consequence of Theorem 2 below.

4. Definition of $\stackrel{\leq}{a}$.

Corresponding to the notion that X is less risky than Y if X equals Y plus some uncorrelated noise is ^{the} partial ordering $\stackrel{\leq}{a}$ which we now define. $F \stackrel{\leq}{a} G$ 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

$$J(y) = \Pr(X + Z \leq y)$$

then

$$(14) \quad F(x) = H(x, 1); \quad 0 \leq x \leq 1$$

$$(15) \quad G(y) = J(y) \quad 0 \leq y \leq 1$$

and

$$(16) \quad E(Z|X = x) = 0 \quad \text{for all } x.$$

That $\stackrel{\leq}{a}$ is a partial ordering follows from Theorem 2. However, to prove Theorem 2 we shall first need to prove $\stackrel{\leq}{a}$ is transitive.

Lemma 4. If $X^1 \stackrel{\leq}{a} X^2 \stackrel{\leq}{a} X^3$, then $X^1 \stackrel{\leq}{a} X^3$.

Proof. By hypothesis $X^2 = X^1 + Z^1$ and $X^3 = X^2 + Z^2$. Thus, $X^3 = X^1 + Z^1 + Z^2$. It remains to show that $E(Z^1 + Z^2 | X^1) = 0$. But,

$$E(Z^1 + Z^2 | X^1) = E E_{Z^1 Z^2} (Z^1 + Z^2 | X^1, Z^1) = E E_{Z^1 Z^2} (Z^2 | X^1, Z^1) =$$

$$E E_{Z^1 Z^2} (Z^2 | X^1 + Z^1) = 0.$$

5. Equivalence of $\frac{\leq}{I}$, $\frac{\leq}{a}$, $\frac{\leq}{u}$.

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

Theorem 2. The following statements are equivalent:

$$(A) \quad F \frac{\leq}{u} G$$

$$(B) \quad F \frac{\leq}{I} G$$

$$(C) \quad F \frac{\leq}{a} G$$

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.

$$a. \quad X \frac{\leq}{a} Y \Rightarrow X \frac{\leq}{u} Y.$$

By hypothesis there is an r.v. S such that $Y = X + Z$ and $E(Z|X) = 0$. For every fixed X and concave U we have by Jensen's inequality

$$\overline{E_{Z^1} (Z^2 | X^1, Z^1)} \text{ is the expectation with respect to } Z^1 \text{ of the r.v. } (Z^2 | X^1, Z^1).$$

$$E_Z U(X + Z) \leq U(E_Z (X + Z)) = U(X)$$

Taking expectations,

$$E_X E_Z U(X + Z) \leq E_X U(X)$$

or

$$E U(Y) \leq E U(X) .$$

$$b. \quad F \stackrel{\leq}{u} G \implies F \stackrel{\leq}{I} G^1 .$$

If $S = G - F$ then $F \stackrel{\leq}{u} G$ implies $\int_0^1 U(x) dS(x) \leq 0$ for

all concave U . Since the identity function and its negative are both

concave we have that $\int_0^1 x dS(x) + \int_0^1 (-x) dS(x) \leq 0$ or that

$\int_0^1 x dS(x) = 0$. Integrating by parts we find that $T(1) = 0$. It remains

to show that $T(y) \geq 0$ 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 $0 \leq$

$$\int_0^1 b_y(x) dS(x) = \int_0^y (y - x) dS(x) =$$

$y S(y) - \int_0^y x dS(x)$. Integrating the last term by parts we find that

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

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

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

$$c. \quad F \stackrel{\leq}{I} G \implies F \stackrel{\leq}{a} G.$$

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

$$(4.i) \quad \gamma_1 = -\gamma_2 \geq 0, \quad \gamma_4 = -\gamma_3 \geq 0$$

and

$$(4.ii) \quad \sum_{k=1}^4 \gamma_k a_k = 0$$

are the conditions that G differs from F by a single MPS. Define a r.v. Z conditional on X as follows:

if $X \neq a_2$ or $X \neq a_3$, $Z = 0$

if $X = a_2$, $\Pr(Z = 0) = \frac{g_2}{f_2} = \frac{f_2 + \gamma_2}{f_2}$, $\Pr(Z = a_1 - a_2) = \frac{C_{12}}{f_2}$, and

$\Pr(Z = a_4 - a_2) = \frac{C_{42}}{f_2}$ where C_{12} and C_{42} are the solutions to

$$(17) \quad C_{12} + C_{42} = -\gamma_2$$

$$(18) \quad (a_1 - a_2) C_{12} + (a_4 - a_2) C_{42} = 0$$

$$\text{If } X = a_3, \Pr(Z = 0) = \frac{g_3}{f_3} = \frac{f_3 + \gamma_3}{f_3}, \quad \Pr(Z = a_1 - a_3) = \frac{c_{13}}{f_3}$$

$$\Pr(Z = a_4 - a_3) = \frac{c_{43}}{f_3} \quad \text{where } c_{13} \text{ and } c_{43} \text{ are the solutions to}$$

$$(19) \quad c_{13} + c_{43} = -\gamma_3$$

$$(20) \quad (a_1 - a_3) c_{13} + (a_4 - a_3) c_{43} = 0.$$

Straightforward algebra will establish that c_{12} , c_{42} , c_{13} , c_{43} are non-negative and these conditions do define a r.v. Z . Conditions (18) and (20) imply $E(Z|X) = 0$. It remains to show that $Y = X + Z$. If $Y^1 = X + Z$, 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 $\Pr(Y^1 = a_2) = \Pr(X = a_2) \cdot \Pr(Z = 0 | X = a_2) = f_2 \cdot \frac{g_2}{f_2} = g_2 = \Pr(Y = a_2)$. 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, $Y = Y^1 = X + Z$.

Lemma 1 and Lemma 4 allow us to extend this result to all discrete distributions with a finite number of points of increase. We use Theorem 1, to extend it to all c.d.f.'s. If $F \leq_I G$, there exists sequences $\{F_n\}$

and $\{G_n\}$ of discrete distributions with a finite number of points of increase such that $F_n \rightarrow F$ and $G_n \rightarrow G$ and $F_n \leq G_n$. We have just shown $F_n \leq G_n$. 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 $J_n(y) = \Pr(X_n + Z_n \leq y)$, then

$$(21) \quad J_n(y) = G_n(y)$$

$$(22) \quad F_n(x) = H_n(x, 1)$$

and

$$(23) \quad E(X_n \mid Z_n) = 0.$$

Since H_n is a discrete distribution function (23) can be phrased as

$$(24) \quad \int_0^1 \int_{-1}^1 u(x)z \, dH_n(x, z) = 0$$

for all continuous functions u defined on $[0, 1]$. Since H_n is stochastically bounded, the sequence $\{H_n\}$ has a subsequence $\{H_{n_i}\}$ which converges to a distribution function¹ $H(x, z)$ of the r.v.'s X and Z . Since $H_{n_i}(x, 1) = F_{n_i}(x) \rightarrow F$, $H_{n_i}(x, 1) \rightarrow F$. Similarly, $J_{n_i} \rightarrow G$.

Let $M_{n_i} = \int_0^1 \int_{-1}^1 u(x)z \, dH_{n_i}(x, z)$. By the definitions of weak convergence $M_{n_i} \rightarrow \int_0^1 \int_{-1}^1 u(x)z \, dH(x, z)$. But $\{M_{n_i}\}$ is a sequence all

1 Feller II, pp. 247, 261

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 $E(Z | X) = 0$. 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 \geq_a , \geq_I , and \geq_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 exists two concave functions, U_1 and U_2 , such

that $\int_0^1 U_1 dF(x) > \int_0^1 U_1 dG(x)$ while $\int_0^1 U_2 dF(x) < \int_0^1 U_2 dG(x)$; i.e.

there is some risk averse individual who prefers F to G and another who

prefers G to F . On the other hand, the ordering \geq_V associated with

mean-variance analysis ($X \leq_V Y$ if $E_n X = EY$ and $EX^2 \geq EY^2$) is a complete ordering, i.e.

if X and Y have the same mean, either $X \leq_V Y$ or $X \geq_V Y$.

(b) Concavity. We have already noted that if U is concave, $X \leq_I Y$

implies $EU(X) \leq EU(Y)$. Similarly, given any differentiable function U

which over the interval $[0, 1]$ is neither concave nor convex, then there

exists distribution functions F , G , and H , $F \stackrel{\geq}{I} G \stackrel{\geq}{I} H$, such that

$$\int_0^1 U(x) dF \leq \int_0^1 U(x) dG, \text{ but } \int_0^1 U(x) dG \geq \int_0^1 U(x) dH.$$

In short, $\stackrel{\geq}{I}$ defines the set of all concave functions; a function

U is concave if and only if $X \stackrel{\leq}{I} Y$ implies $EU(X) \leq EU(Y)$.

IV. Mean-Variance Analysis.

The method most frequently used for comparing uncertain prospects for 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 $F \stackrel{\leq}{U} G \Rightarrow F \stackrel{\geq}{V} G$ (since variance is a convex function), $F \stackrel{\geq}{V} G$ does not imply $F \stackrel{\geq}{U} G$. Indeed by arguments closely analogous to those used earlier, it can be shown that a function U is quadratic if and only if $X \stackrel{\geq}{V} Y$ implies $EU(X) \geq EU(Y)$. An immediate consequence of this is that if $U(x)$ is any non-quadratic 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 [16], p. 20-21) -- valid only for distributions which differ only by "location parameters." (See Feller [4], 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 by 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.¹ It is clear

¹ Or some subset of E^2 ; we might restrict one or both of our parameters to be non-negative.

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 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, $\frac{\partial T(y)}{\partial \sigma^2}$ changes sign, where $T(y, \sigma^2, \mu) = \int_0^y F(x; \sigma^2, \mu)$; i.e. there exist individuals with concave utility functions who are better off with an increase in variance.¹

V. Applications

1. Choosing probability distributions. There are a number of situations in which an individual must, in effect, choose, for one purpose or another, a probability distribution from among a set of possible probability distri-

1 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 $[0, 1]$)

$$F(x; a, c) = \begin{cases} 0 & x \leq 1 - .25/a \\ ax + .25 - a & 1 - .25/a \leq x \leq 1 + (2c - .5)/(c - a) \\ cx + .75 - 3c & 1 + (2c - .5)/(c - a) \leq x \leq 3 + .25/c \\ 1 & x > 3 + .25/c \end{cases}$$

Two members of the family with the same mean but different variances are depicted in Figure 10(a). They clearly do not satisfy (6). The density functions are illustrated in Figure 10(b).

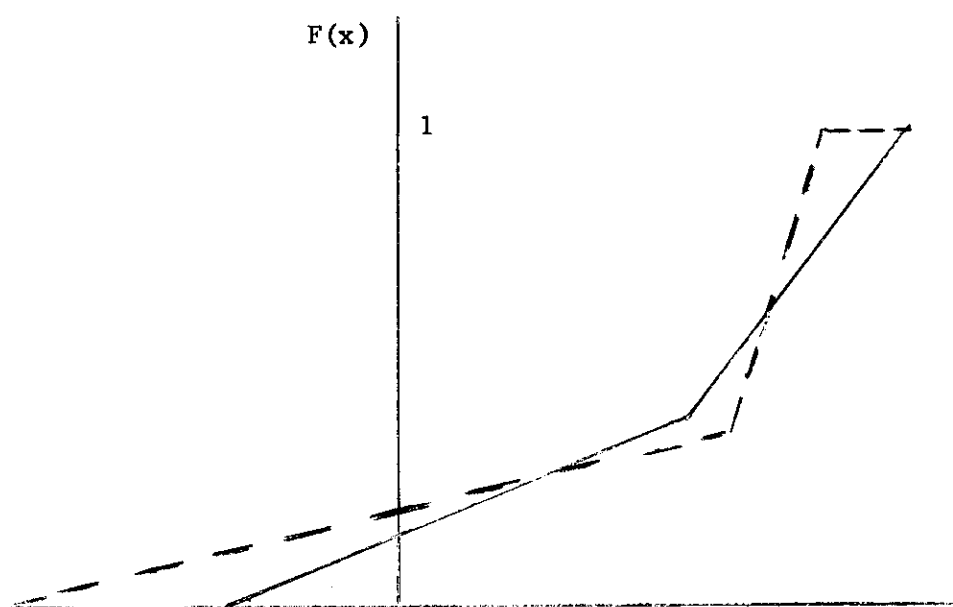


FIGURE 10(a)

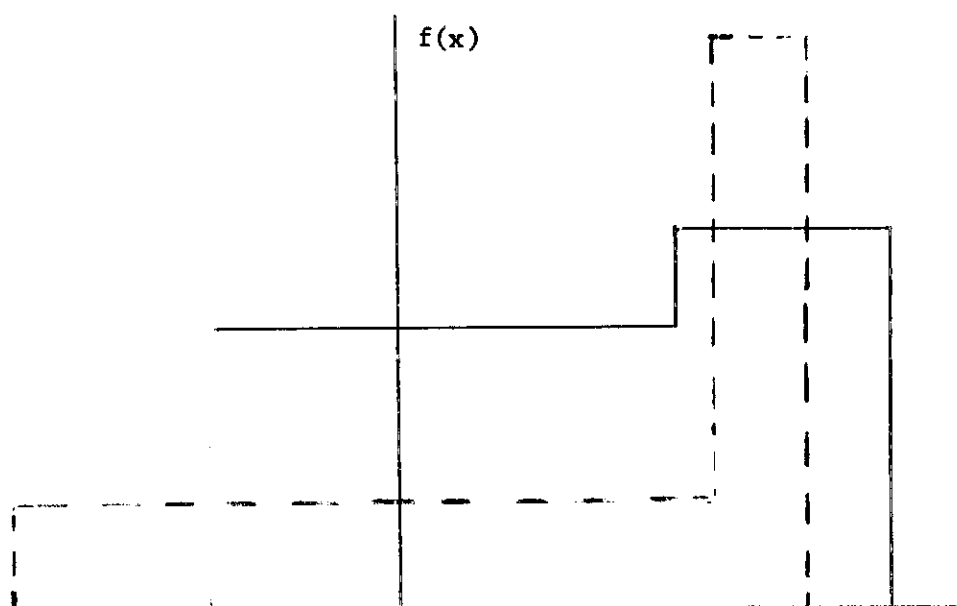


FIGURE 10(b)

butions. The following examples show how the preceding theorem may be applied to prove some general theorems about such situations.

(a) Diversification Theorem.¹ Assume an individual can purchase shares of two² 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

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

where U is a concave function.³ We prove that independent of the utility function b should be set at $1/2$. We can write

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

Since

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

$$y_b \leq y_{1/2} \quad \text{for all } b, \text{ by Theorem 2.}$$

(b). The Rao-Blackwell Theorem.⁴ This well-known theorem follows as an immediate corollary of our theorem 2. Let the distribution of X depend on some unknown parameter θ , and let S be sufficient statistic for θ . Let $d(X)$ be any nonrandomized decision rule, and let \hat{d} be the decision rule defined by

$$\hat{d}(t) = E(d(X) \mid S = t)$$

1 See [12] for an alternative proof and general discussion of this theorem.

2 The generalization to take case of n securities is straight forward.

3 We assume that $E e_1$ exists and is finite so EU exists and is finite.

4 Concepts and notation are borrowed from Ferguson [5].

(Because S is a sufficient statistic, \hat{d} is a function of t only.) Assume the loss function $L(\theta, y)$ (where y is the value of the decision rule) is convex. Then we wish to show that

$$E L(\theta, \hat{d}) \leq E L(\theta, d)$$

Observe that we can, without loss of generality, write

$$d(X) \equiv \hat{d}(t) + Z$$

with

$$E(d(X) \mid t) = \hat{d}(t) + E(Z \mid t)$$

By the definition of $\hat{d}(t)$,

$$E(Z \mid t) = 0.$$

The result follows immediately from Theorem 2.

2. Effects of Increasing Variability on Economic Decisions.

One of the original motivations of this study was to enable us to examine the effects of an increase in risk on some control parameter α , where α was chosen to maximize

$$(25) \quad \int U(X, \alpha) dF(x)$$

The first order condition for utility maximization is

$$(26) \quad \int \frac{\partial U(X, \alpha)}{\partial \alpha} dF(x) = 0$$

Assume there is a unique solution to (26), α^* and, that, in the neighborhood of α^* , U_{α} is monotone (decreasing) in α .¹ Then if $U_{\alpha}(x, \alpha)$ is

1 We assume $U_{\alpha\alpha} < 0$ for all (α, x) in the relevant region. In the examples below we shall have occasion to demonstrate the uniqueness of α^* .

a convex (concave) function of X , an increase in riskiness will lead to an increase (decrease) in $EU_{\alpha}(X, \alpha)$. But since $EU_{\alpha}(X, \alpha^*) = 0$, before the change, now $EU_{\alpha}(X, \alpha^*) > 0$ (< 0). Hence α^* is increased (decreased). In any particular problem, the question becomes one of ascertaining the conditions under which $U_{\alpha}(X, \alpha)$ is concave (convex) in X .

Earlier studies (see, e.g. [9,15]) 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 $\hat{\alpha} \geq \alpha^*$, where α^* is the solution to (25), depends simply on whether

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

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.

We examine below five problems of economic interest. Two are concerned with the consumer's allocation of his portfolio between safe and risky assets and of his income between savings and consumption. The remaining three are production problems involving the choice of technique and level of output under uncertainty.

(a). Savings and uncertainty.¹ There are at least two stories of how uncertainty about the rate of return on savings affects the savings rate. (i) A risk averse individual, in order to ensure his "minimum standard of living" saves more in the face of uncertainty. (ii) A risk averse individual is discouraged from saving by the uncertainty of the return -- "a bird in the hand is worth two in the bush." We shall show that whether the savings rate increases or decreases (in our simplified model) depends on whether relative risk aversion $[1,10]$ is less than or greater than unity.

We consider a risk averse 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 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,

$$E\{U(C_1) + (1-\delta)U(C_2)\} = E\{U((1-s)W_0) + (1-\delta)U(sW_0e)\}$$

where s is the savings rate and δ the pure rate of time preference. The necessary and sufficient condition for utility maximization is that

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

Whether s decreases or increases as risk increases depends on whether $U'(C_2)e$ is concave or convex in e ; i.e. whether

1 For a fuller discussion of this and related problems, see [7, 11].

$$U''(C_2)(1-R(C_2)) - U'(C_2)(R'(C_2)) \lesseqgtr 0$$

where $R(C) = -U''(C)C/U'(C)$ is the Arrow-Pratt measure of relative risk aversion and $C_2 = sW_0e$ is consumption in the second period. If relative risk aversion is constant, savings is unaffected if relative risk aversion is unity (the Bernoulli utility function), decreased if it is less than unity, increased if it is greater. If relative risk aversion is increasing, but less than or equal to one throughout the relevant range, then savings are increased.

Note that if we had assumed that individuals had a quadratic utility function,

$$U(C_1, C_2) = (C_1 - \bar{C})^2 + (1-\delta)(C_2 - \bar{C})^2$$

so that expected utility maximization yields

$$s = \frac{(1-\delta)\bar{C}Ee + W_0 - \bar{C}}{W_0(Ee^2(1-\delta) + 1)}$$

then we would have concluded that an increase in variance, with fixed mean, always lowers savings.

(b). A firm's production problem.¹ 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 (for simplicity) 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

¹ See [11] for a more complete analysis of this problem.

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

$$(27) \quad E[rK + wL(K, Q)] = rK + wE[L(K, Q)]$$

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 (27), the first order conditions may be written

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

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

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

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

1 If $L > 0$; $\frac{\partial L}{\partial K} = 0$ otherwise.

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

A sufficient condition for convexity is that $\rho \leq 0$, i.e. 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 function 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 figure 11) or, equivalently, whether the probability that Q will be greater than bK^* (the "capacity" of the original capital stock) increases or decreases.

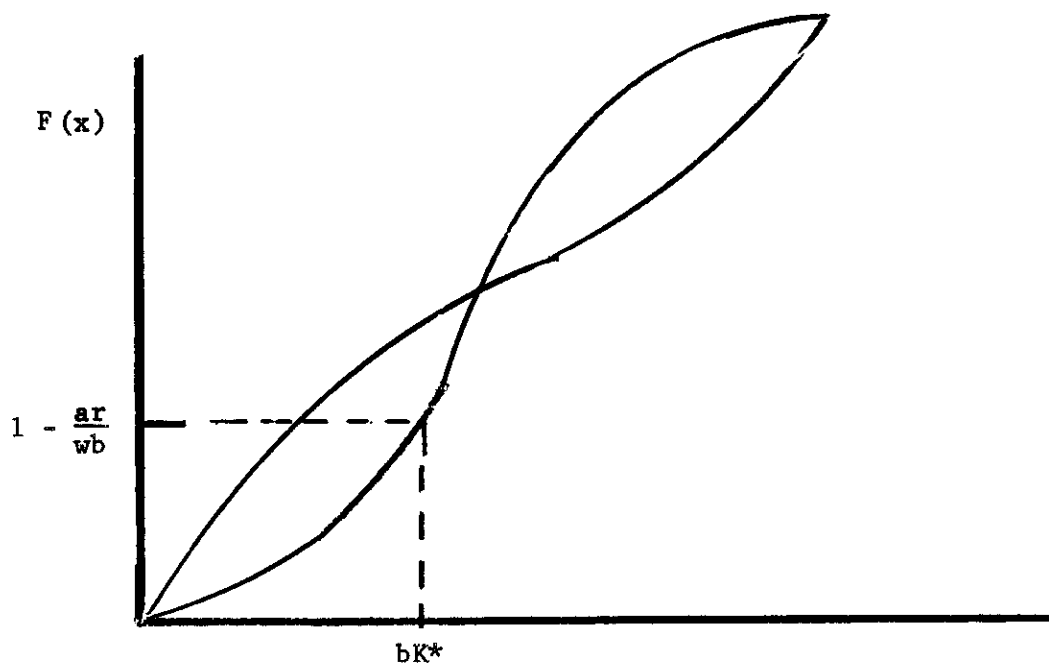


FIGURE 11

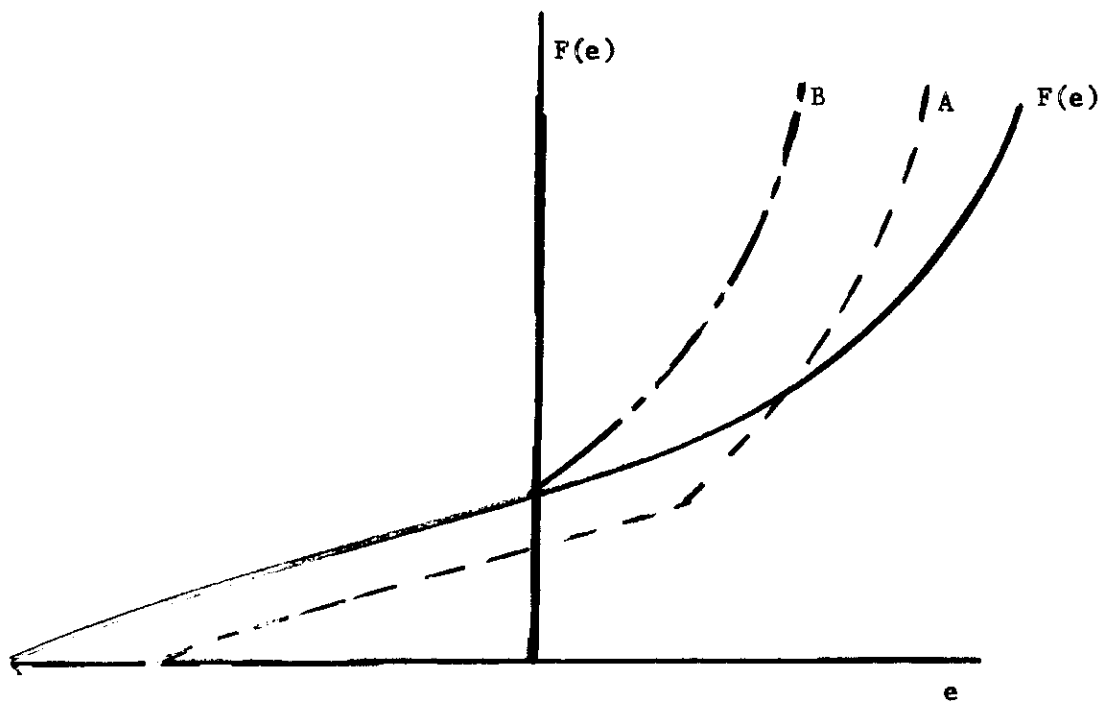


FIGURE 12

(c). A multi-stage planning problem.¹ 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

$$P = P_2 M'.$$

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

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

If e becomes more variable, what happens to L_1 (and L_2)? This depends on the sign of

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

¹ This problem was posed to us by M. Weitzman.

Assume that P is a constant elasticity of substitution production function:

$$P = (\delta L_2^\rho + (1-\delta)y^\rho)^{1/\rho} . \text{ Then}$$

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

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

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 $1 \geq \rho \geq 0$, i.e. the elasticity of substitution is greater than or equal to unity, $P_{122} \leq 0$ and $P_{222} \geq 0$, so L_2 decreases, L_1 increases, i.e. more labor is allocated to the earlier stage of production.

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

$$\begin{aligned} E(Q) &= \int_{-\infty}^{L_2 - M(L_1)} [M(L_1) + e] dF(e) + L_2(1 - F(L_2 - M(L_1))) \\ &= \int_{-\infty}^{\bar{L} - L_1 - M(L_1)} [M(L_1) + e] dF(e) + (\bar{L} - L_1)(1 - F(\bar{L} - L_1 - M(L_1))) \end{aligned}$$

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 $M''P - f(M' - 1)^2 < 0$, 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

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

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

(d). A portfolio problem. An individual with initial wealth of W_0 , wishes to allocate a fixed amount of wealth, W_0 , between money, which yields a zero rate of return, and a risky asset which yields a random rate of return e , so as to maximize the expected utility of his terminal wealth:

$$EU(W) = E(U(W_0(1 + ae)))$$

where a is the fraction of his wealth invested in the risky asset. U is assumed to be concave. A necessary and sufficient condition for utility maximization is

$$EU'e = 0$$

What happens to a if e becomes riskier depends on whether $U'e$ is concave or convex, i.e. whether

$$U''(1 - R + W_0 A) + U'(W_0 A' - R') \geq 0,$$

where, as before, $R = -U''W/U'$, the Arrow-Pratt measure of relative risk aversion and $A = -U''/U'$, the measure of absolute risk aversion. A sufficient condition for an increase in uncertainty leading to an increase allocation to the safe asset is that relative risk aversion be less than or equal to unity, and that absolute risk aversion be non-increasing and relative risk aversion be non-decreasing. The Bernoulli utility function clearly satisfies these conditions.

Note that once again if we had begun the analysis by assuming a quadratic utility function, we would have obtained misleadingly unambiguous results: If $U' = \alpha - \beta W$,

$$a = \frac{(\alpha - \beta W_0)Ee}{\beta W_0 Ee^2}$$

so

$$\frac{da}{dEe^2} < 0$$

An increase in variance, with mean fixed, lowers the allocation to the risky asset.

Taxation of earnings from investments amounts to a particular kind of change in the distribution of the payoff from an investment. The results we obtain here are much weaker than the corresponding results for the effect of an income tax with full loss offset, but they are identical to those obtained in [14] for an income tax with no loss offset. Such a tax can be viewed as a mean preserving reduction in risk (F to A in figure 12) plus a reduction in mean (A to B in figure 12), by shifting

the distribution to the left. The latter will lead to an increase in the demand for the safe asset if there is decreasing absolute risk aversion, a condition already included in the condition for a mean preserving reduction in risk leading to an increase in the demand for the safe asset.

(e). Choice of output level for a competitive firm. In the examples considered so far, the conditions we have obtained under which 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 weaker conditions 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 concave¹ and where

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

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

$$\frac{EU'P}{EU'} = C'(Q^*)$$

If the producer is risk neutral or if there is no variability in p , profit

1 For a discussion of the case of constant absolute risk aversion, see [9].

maximization requires that price equals marginal cost,

$$Ep = C'(\hat{Q}) .$$

$Q \geq \hat{Q}$ as $\frac{EU'p}{EU'} \geq Ep$, i.e., as $E[(U' - EU')(p - E(p))] \geq 0$. But since

$U'' < 0$, $U'(p) \geq U'(E(p))$ as $p \leq E(p)$, so $E[(U' - EU')(p - E(p))]$

$= E[(U' - U'(Ep))(p - E(p))] < 0$. Hence, there is always less output under uncertainty than under certainty.

REFERENCES

1. Arrow, K.J. "Aspects of the Theory of Risk Bearing", Yrjö Jahnsson Lectures, Helsinki, 1965.
2. Bessler, S.A. and A.F. Veinott, Jr. "Optimal Policy for a Dynamic Multi-Echelon Inventory Model", Naval Research Logistics Quarterly, (Dec. 1966), Vol. 13, No. 4, pp. 355-387.
3. Diamond, P.A., "Savings Decisions Under Uncertainty." Working Paper No. 71, Institute of Business Economic Research, University of California, Berkeley, June 1965.
4. Feller, William, An Introduction to Probability Theory and Its Application, Vol. II, N.Y., 1966.
5. Ferguson, T.S., Mathematic Statistics, New York: Academic Press, 1967.
6. Hader, J. and Russell, W.R., "Rules for Ordering Uncertain Prospects." American Economic Review, 59, March 1965, pp. 25-34.
7. Hahn, F.H., "Savings Under Uncertainty." LSE mimeo, May, 1968.
8. Hanoch, G. and C. Levy, "Efficiency Analysis of Choices Inventory Risk," mimeo. 1968.
9. McCall, J.J., "Competitive Production for Constant Risks Utility Function," Review of Economic Studies 34, (October 1964), pp. 417-419.
10. Pratt, J.W. "Risk Aversion in the Small and in the Large," Econometrica, Vol. 32, (Jan. - April 64), pp. 122-136.
11. Rothschild, M., "Changing Demand: Its Costs and Consequences," Paper presented to the Econometric Society, Evanston, Dec. 1968.
12. Samuelson, P.A., "General Proof that Diversification Pays," Journal of Financial and Quantitative Analysis, 2, March 1967, pp. 1 - 13.
13. Solow, R.M., "A Note on Dynamic Multiplier's" Econometrica, Vol. 19, (July 1951), pp. 306-16.
14. Stiglitz, J. "The Effects of Wealth, Income and Capital Gains Taxation on Risk Taking," Quarterly Journal of Economics, May 1969.
15. _____, "Liquidity Preference and the Term Structure of Interest Rates," Cowles Foundation Discussion Paper 252, August 1968.

16. Tobin, James, "The Theory of Portfolio Selection", in Hahn, F, and F. Brechling, The Theory of Interest Rates, London (MacMillan, 1965.)
17. _____, "Comment on Borch and Feldstein," Review of Economic Studies, 36, January 1965, pp. 13-14.