GENERAL PROOF THAT DIVERSIFICATION PAYS**

Paul A. Samuelson*

"Don't put all your eggs in one basket," is a familiar adage. Economists, such as Marschak, Markowitz, and Tobin,[1] who work only with mean income and its variance, can give specific content to this rule--namely, putting a fixed total of wealth equally into independently, identically distributed investments will leave the mean gain unchanged and will minimize the variance.

However, there are many grounds for being dissatisfied with an analysis dependent upon but two moments, the mean and variance, of a statistical distribution. I have long used the following, almost obvious, theorem in lectures. When challenged to find it in the literature, I was unable to produce a reference--even though I should think it must have been stated more than once.

<u>Theorem I:</u> If $U(X)$ is a strictly concave and smooth function that is monotonic for non-negative X, and $(X_1,\ldots,X_n)$ are independently, identically distributed variates with joint frequency distribution

$$\text{Prob}\{X_1 \le x_1,\ldots,X_n \le x_n\} = F(x_1)F(x_2)\ldots F(x_n)$$

$$\text{with } E[x_i] = \int_{-\infty}^{\infty} X_i dF(X_i) = \mu_1$$

$$E[x_i - \mu_i]^2 = \int_{-\infty}^{\infty} (X_i - \mu_1)^2 dF(X_i) = \mu_2$$

$$\text{with } 0 < \mu_2 < \infty ,$$

* Massachusetts Institute of Technology.

** My thanks go to the Carnegie Corporation for providing me with a reflective year and to Mrs. F. Skidmore for research assistance.

then

$$E\left[U\left(\sum_{1}^{n}\lambda_j X_j\right)\right] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} U\left(\sum_{1}^{n}\lambda_j X_j\right) dF(X_1)\ \ldots\ dF(X_n)$$

$$= \Psi(\lambda_1,\ \ldots,\ \lambda_n)$$

is a strictly concave symmetric function that attains its

unique maximum, subject to

$\lambda_1 + \lambda_2 + \ldots + \lambda_n = 1\ ,\quad \lambda_i \geq 0$

at $\Psi\left(\frac{1}{n}\ ,\ \ldots,\ \frac{1}{n}\right)$.

The proof is along the lines of a proof[2] used to show that equal distribution of income among identical Benthamites will maximize the sum of social utility.

$$\frac{\partial\Psi(\lambda_1,\ldots,\lambda_n)}{\partial\lambda_i} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} X_i\ U'\left(\sum_{1}^{n}\lambda_j X_j\right) dF(X_1)\ldots dF(X_n)$$

is independent of i at $(\lambda_1,\ldots\lambda_n) = (1/n,\ldots 1/n)$ by symmetry.

The Hessian matrix with elements

$$\frac{\partial^2\Psi}{\partial\lambda_i\partial\lambda_j} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} X_i X_j U''\left(\sum_{1}^{n}\lambda_k X_k\right) dF(X_1)\ldots dF(X_n)$$

is a Grammian negative definite matrix if $-U'' > 0$, $0 < \mu_2 < \infty$.

Hence, sufficient maximum conditions for a unique maximum are satisfied, namely

$$\frac{\partial\Psi\left(\frac{1}{n},\ \ldots,\ \frac{1}{n}\right)}{\partial\lambda_1} = \cdots = \frac{\partial\Psi\left(\frac{1}{n},\ \ldots,\frac{1}{n}\right)}{\partial\lambda_n}$$

$\sum_{1}^{n}\sum_{1}^{n}\frac{\partial^2\Psi}{\partial\lambda_i\partial\lambda_j}\ y_i y_j < 0$ for all non-negative λ's and not all y's vanishing.

Remarks: Differentiability assumptions could be lightened. It is not true, by the way, that $\frac{1}{2}X_1 + \frac{1}{2}X_2$ has a "uniformly more-bunched distribution" than X_1 or X_2 separately, as simple examples (even with finite μ_2) can show: still the risk averter will always benefit from diversification. The finiteness of μ_2 is important. Thus, for a Cauchy distribution $\frac{1}{2}X_1 + \frac{1}{2}X_2$ has the same distribution as either X_1 or X_2 separately; for the arc-sine Pareto-Lévy case, it has a worse distribution. The proof fails because the postulated E[U] cannot exist (be finite) for any concave U.

The General Case of Symmetric Interdependence

We can now drop the assumption of independence of distribution, replacing it by the less restrictive postulate of a symmetric joint distribution. I.e., we replace $F(x_1) \ldots F(x_n)$ by

$$\text{Prob}\{X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n\} = P(x_1, x_2, \ldots, x_n)$$

where P is a symmetric function in its arguments. We can rule out, as trivial, the case where the x's are connected by an exact functional relation, which in view of symmetry would have to take the form

$$x_1 = x_2 = \ldots = x_n = x$$

$$\text{Prob}\{X \leq x\} = P(x) = P(x, \ldots, x) \quad .$$

We do stipulate finite means, variances, and covariances

$$E[x_i] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} X_i \, dP(X_1, \ldots, X_n) = \mu$$

$$E[(x_i - \mu)(x_j - \mu)] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} (X_i - \mu)(X_j - \mu) \, dP(X_1, \ldots, X_n)$$

$$= \sigma_{ij}$$

the elements of a positive definite Grammian matrix. A generalization of our earlier theory on diversification can now be stated.

Theorem II: For U(x) a smooth, strictly concave function, the maximum of the symmetric, concave function

$$E\left[U\left(\sum_1^n \lambda_j x_j\right)\right] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} U\left(\sum_1^n \lambda_j X_j\right) dP(X_1, \ldots, X_n)$$

$$= \phi(\lambda_1, \ldots, \lambda_n) \quad ,$$

subject to

$$\lambda_1 + \lambda_2 + \cdots \lambda_n = 1, \quad \lambda_i \geq 0 \ ,$$

is given by $\phi\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$. Thus, diversification always pays.

The proof is exactly as before. By symmetry

$$\frac{\partial\phi\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)}{\partial\lambda_1} = \cdots = \frac{\partial\phi\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)}{\partial\lambda_n} ,$$

the necessary first-order conditions for the constrained maximum.

The Hessian matrix has elements

$$\frac{\partial^2 \phi}{\partial\lambda_i \, \partial\lambda_j} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} X_i X_j U''\left(\sum_1^n \lambda_k X_k\right) dP(X_1, \ldots, X_n)$$

which, being the coefficients of a negative-definite Grammian matrix, do confirm the concavity of ϕ and therefore the maximum value at $\phi(1/n, \ldots, 1/n)$.

If equal diversification is to be mandatory, symmetry or some assumption like it is of course needed. To verify this obvious fact, suppose (x_1, x_2, x_3) to be independently distributed, with x_2 and x_3 having the same distribution $P(x_i)$, but with x_1 having a distribution that is identical with that of $\frac{1}{2} x_2 + \frac{1}{2} x_3$, namely

$$\text{Prob}\{X_1 \leq x_1\} = Q(x_1) = \int_{-\infty}^{\infty} P(2x_1 - 2s)dP(2s) \quad .$$

In this case, symmetry tells us that wealth should be divided equally between the investments x_1 and $\frac{1}{2} x_2 + \frac{1}{2} x_3$, which is equivalent to investing in the (x_1, x_2, x_3) in the fractions $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$. Those who work with two moments, mean and covariance matrix, will find that minimum variance does not come at $(\lambda_1, \ldots, \lambda_n) = (1/n, \ldots, 1/n)$ when $\sigma_{ii} \neq \sigma_{jj}$, $\sigma_{ij} \neq \sigma_{rs}$.

It is possible, though, to prove that <u>some</u> positive diversification is mandatory under fairly general circumstances. Thus, in $(x_1, \ldots, x_n)$ let each have a common mean and each have finite but nonzero variance. Finally, suppose that one of the variables, say x_1, is independently distributed from the rest. Then an optimal portfolio must involve $\lambda_1^* > 0$, with some positive investment in x_1, as shown in the following.

Theorem III. Let $(x_1, x_2, \ldots, x_n)$ be jointly distributed as $P(x_1)Q(x_2, \ldots, x_n)$ with common mean and finite positive variances

$$E[x_i] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} X_i dP(X_1) dQ(X_2, \ldots, X_n) = \mu$$

$$0 < E[(x_i - \mu)^2] < \infty$$

and

$$E[U] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} U\left(\sum_1^n \lambda_j X_j\right) dP(X_1) dQ(X_2, \ldots, X_n)$$

$$= \theta(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

where, for $U'' < 0$, θ is a strictly concave function.

Then if

$$\theta(\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*) = \underset{\{\lambda_i\}}{\text{Max}}\ \theta(\lambda_1, \ldots, \lambda_n) \text{ s.t. } \sum_1^n \lambda_j = 1,\ \lambda_i \geq 0,$$

necessarily $\lambda_1^* > 0$ and $\lambda_1^* < 1$.

This will first be proved for $n = 2$, since the general case can be reduced down to that case. Denoting $\partial\theta(\lambda_1, \lambda_2)/\partial\lambda_i$ by $\theta_i(\lambda_1, \lambda_2)$, we need only show the following to be positive

$$\theta_1(0,1) - \theta_2(0,1) = \int_{-\infty}^{\infty} X_1 dP(X_1) \int_{-\infty}^{\infty} U'(X_2) dP(X_2) - \int_{-\infty}^{\infty} X_2 U'(X_2) dP(X_2)$$

$$= E[x_2]E[U'(x_2)] - E[x_2 U'(x_2)]$$

$$= -E\left[\{x_2 - \mu\}\{U'(x_2) - E[U'(x_2)]\}\right] > 0$$

if $U''(x_2) < 0$, since the Pearsonian correlation coefficient between any monotone-decreasing function and its argument is negative.

We reduce $n > 2$ to the $n = 2$ case by defining

$$\lambda_1 x_1 + \sum_2^n \lambda_j x_j = \lambda_1 x_1 + \lambda_{II} x_{II}$$

where

$$x_{II} = \sum_2^n \frac{\lambda_j}{\lambda_{II}} x_j, \quad \sum_2^n \frac{\lambda_j}{\lambda_{II}} = 1, \text{ as definition of } \lambda_{II}.$$

To show that the optimal portfolio has the property

$$\theta(\lambda_1^*,\lambda_2^*,\ldots,\lambda_n^*) > \theta(0,\lambda_2,\ldots,\lambda_n) \text{ for } \sum_2^n \lambda_j = 1,$$

it suffices to show that

$$\theta(\lambda_1^*,\lambda_2^*,\ldots,\lambda_n^*) > \theta(0,\lambda_2^{**} \cdots \lambda_n^{**})$$

$$= \underset{\{\lambda_2,\ldots,\lambda_n\}}{\text{Max}} \theta(0,\lambda_2,\ldots,\lambda_n) \text{ for } \sum_2^n \lambda_j = 1$$

But now if we define

$$\theta(\lambda_1,\lambda_{II}) = \theta(\lambda_1,\lambda_{II}\lambda_2^{**},\ldots,\lambda_{II}\lambda_n^{**}),$$

we have an ordinary n=2 case, for which we have shown that $\lambda_1^* > 0$ and $\lambda_{II}^* > 0$.

Having completed the proof of Theorem III, we can enunciate two easy corollaries that apply to risky investments.

Corollary I. If any investment has a mean at least as good as any other investment, and is independently distributed from all other investments, it must enter positively in the optimal portfolio.

Corollary II. If all investments have a common mean and are independently distributed, all must enter positively in the optimum portfolio.

Can one drop the strict independence assumption and still show that every investment, in a group with identical mean, must enter positively in the optimum portfolio? The answer is, in general, no. Only if, so to speak, the component of an investment that is orthogonal to the rest has an attractive mean can we be sure of wanting it. Since a single counterexample suffices, consider joint normal-distributions (x_1,x_2), with common mean and where optimality requires merely the minimization of the variance of $\lambda_1 x_1 + \lambda_2 x_2, \Sigma\Sigma\sigma_{ij}, \lambda_i\lambda_j$, subject to $\Sigma\lambda_j = 1, \lambda_i \geq 0$. If one neglects the non-negativity constraints and minimizes the quadratic expression, one finds for the optimum

$$\lambda_1^* = \frac{\sigma_{22} - \sigma_{12}}{(\sigma_{22} - \sigma_{12}) + (\sigma_{11} - \sigma_{21})} \quad .$$

If $\sigma_{12} = \sigma_{21} < 0$, λ_1^* is definitely a positive fraction. But, if σ_{12} is sufficiently positive, as in the admissible case $(\sigma_{11}, \sigma_{12}, \sigma_{22}) = (2, 1.1, 1)$, λ_1^* would want to take on an absurd negative value and would, of course, in the <u>feasible</u> optimum be zero, even though x_1's mean is equal to x_2's. Naturally this is a case of positive intercorrelation.

If the assumption of independence is abandoned in favor of positive correlation, we have seen that positive diversification need not be mandatory. However, as Professor Solow pointed out to me, abandoning independence in favor of negative correlation ought to improve the case for diversification. We saw that this was true in the case of negative linear correlation between two investments. It is easy to prove for any number of investments with common mean, among which all the intercorrelations are negative, that total variance is at a minimum when each investment appears with positive weight in the portfolio. (Although there is no limit on the degree to which all investments can be positively intercorrelated, it is impossible for <u>all</u> to be strongly negatively correlated. If A and B are both strongly negatively correlated with C, how can A and B fail to be positively intercorrelated with each other? For 3 variables, the maximum common negative correlation coefficient is $-1/2$; for 4 variables, $-1/3$;....for n variables $-1/(n-1)$.)

The whole point of this paper is to free the analysis from dependence on means, variances, and covariances. What is now needed is the generalization of the concept of negative linear correlation of the Pearsonian type. The natural tool is found in the concept of conditional probability of each variable, say x_i, and the requirement that increasing all or any other variables x_j be postulated to reduce this conditional probability. Thus, define

$$\text{Prob}\{X_i \leq x_i \mid \text{each other } X_j = x_j\} = P(x_i \mid \underline{x}_i),$$

$$\text{where } \underline{x}_i \text{ is the vector } (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

As always with conditional probabilities

$$P(x_i \mid \underline{x}_i) = P(x_1, x_2, \ldots, x_n) \div \int_{-\infty}^{\infty} dP(x_1, x_2, \ldots, X_i, \ldots, x_n)$$

where the last divisor $Q(\underline{x}_i) = Q(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$ is assumed not to vanish.

The appropriate generalization of pair-wise negative correlation or negative interdependence is the requirement

$$\frac{\partial P(x_i \underline{x})}{\partial x_j} < 0 \qquad j \neq i$$

Theorem IV, which I shall not prove, states that where the joint probability distribution has the property of negative interdependence as thus defined, and has a common mean expectation for every investment, $E[x_i] = \mu$, every investment must enter with positive weight in the optimal portfolio of a risk-averter with strictly concave $U(x)$. Buying shares in a coal and in an ice company is a familiar example of such diversification strategy.

Having now shown that quite general conclusions can be rigorously proved for models that are free of the restrictive assumption that only two moments count, I ought to say a few words about how objectionably special the 2-moment theories are (except for textbook illustrations and simple proofs). To do this, I must review critically the conditions under which it is believed the mean-variance theories are valid.

1. If the utility to be maximized is a quadratic function of x, $U(x) = a_0 + a_1x - a_2x^2$,

$$\begin{aligned} E[U(x)] &= a_0 + a_1E[x] + a_2E[x]^2 - a_2V(x) \\ &= a_0 + a_1\mu + a_2\mu^2 - a_2\sigma_X^2, \\ &= f(\mu,\sigma) \end{aligned}$$

where μ = mean of x and σ^2 = variance of x

However, as Raiffa, Richter, Hicks, and other writers[3] have noted, the behavior resulting from quadratic utility contradicts familiar empirical patterns. [E.g., the more wealth I begin with the less will I pay for the chance of winning ($0,$K) with probabilities (1/2,1/2) if I have to maximize quadratic utility. Moreover, for large enough x, U begins to decline -- as if having more money available begins to hurt a person.]

Anyone who uses quadratic utility should take care to ascertain which of his results depend critically upon its special (and empirically objectionable) features.

2. A quite different defense of 2-moment models can be given. Suppose we consider investment with less and less dispersion -- e.g., let

$$P(y_1,\ldots,y_n) \text{ have property } E[y_i] = 0 = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} Y_i\,dP(Y_1,\ldots,Y_n)$$

and

$$\text{Prob}\{X_1 \leq x_1,\ldots,X_n \leq x_n\} = P\left(\frac{x_1-\mu_1}{\alpha},\ldots,\frac{x_n-\mu_n}{\alpha}\right)_n$$

Then in the limit as $\alpha \to 0$, only the first 2 moments of $\Sigma_1 \lambda_j x_j$ will turn out to count in $E[U(x)] = f(\mu,\sigma,\ldots)$. In the extreme limit, even the second moment will count for less and less: for α small enough the mean <u>money</u> outcome will dominate in decision making. Similarly, when α is small, but not limitingly small, the third moment of skewness will still count along with the mean and variance; then the third-degree polynomial form of U(x) (its Taylor's expansion up to that point) will count. As Dr. M. Richter has shown in the cited paper, an n^{th} degree polynomial for U(x) implies, and is implied, by the condition that only the first n statistical moments count.

3. If each of the constituent elements of $(x_1,\ldots,x_n)$ is normally distributed, then so will be $z=\Sigma\lambda_j x_j$ and then it will be the case that only the mean and variance of z matter for E[U(z)]. However, with limited liability, no x_j can become negative as is required by the normal distribution. So some element of approximation would seem to be involved. Is the element of approximation, or rather of lack of approximation, ignorable? No, would seem to be the answer if there is some minimum of subsistence of z or x at which marginal utility becomes infinite. Thus, consider U = logx, the Bernoulli form of logarithmic utility

$$E[U(x)] = \int_{-\infty}^{\infty}\log X\,dN\left(\frac{X-\mu}{\sigma}\right) = \lim_{a\to b}\int_{a}^{\infty}\log X\,dN\left(\frac{X-\mu}{\sigma}\right)$$

$$= -\infty \text{ for } b \leq 0,$$

where N(t) stands for the normal distribution with zero mean and unit variance.

Suppose that each constituent x_i takes on only non-negative values with variances all bounded by $M < \infty$. The central-limit theorem will still apply, so that $\Sigma_1^n x_j$ or $z_n = \Sigma\lambda_j x_j$, with certain weak restrictions on the spread of the λ's around $1/n$, will approach a Gaussian distribution. Thus, let z_n have the distribution $P_n(z_n)$, with

$$\lim_{n\to\infty} P_n(z_n) = N\left(\frac{z_n - a_n}{b_n}\right) \quad .$$

Knowing how treacherous are double limits, we dare not infer

$$\lim_{n\to\infty} E[\log z_n] = \int_{-\infty}^{\infty} \log Z \lim_{n\to\infty} P_n(Z_n) \quad .$$

Actually, as $n\to\infty$ and each investment has its $\lambda_j > a/n$, each $\lambda_j x_j$ does have a smaller and smaller dispersion so that we might switch from reliance on the central limit theorem of normality to the 2-moment Taylor-expansion justification given in paragraph 2 above. The law of large numbers, which is even more basic than the central limit theorem involving normality, assures us that z_n becomes more and more tightly bunched around some positive value and this fact will make the quadratic approximation applicable in the limit.

4. A final defense of the mean-variance formulation, in which $E[U(x)]$ is replaced by $f(\mu,\sigma)$, comes when x belongs to a 2-parameter probability distribution $P(x;\theta_1,\theta_2)$.

Then

$$E[U(x)] = \int_{-\infty}^{\infty} U(X)dP(X;\theta_1,\theta_2) = g(\theta_1,\theta_2) \ ,$$

$$\mu = \int_{-\infty}^{\infty} XdP(X;\theta_1,\theta_2) = h_1(\theta_1,\theta_2) \ ,$$

$$\sigma = \left\{\int_{-\infty}^{\infty} (X-\mu)^2 dP(X;\theta_1,\theta_2)\right\}^{1/2} = h_2(\theta_1,\theta_2) \ .$$

Then, provided the Jacobian $\partial(\mu,\sigma)/\partial(\theta_1,\theta_2) \neq 0$, each θ_i can be solved for as a function $\theta_i(\mu,\sigma)$, with

$$f(\mu,\sigma) = g[\theta_1(\mu,\sigma),\ \theta_2(\mu,\sigma)] \quad .$$

So far so good, although even here one has to take care to verify that $P(x;\theta_1,\theta_2)$ has the properties needed to give $f(\mu,\sigma)$ the quasi-concavity properties used by the practitioners of the mean-variance techniques. And furthermore one cannot draw up the $f(\mu,\sigma)$ indifference contours once and for all from knowledge of the decision-makers risk preferences, but instead must redraw them for each new probability distribution $P(x;\theta_1,\theta_2)$ upon which the f functional depends.

But waiving these last matters, we must point out that the Markowitz efficient-portfolio frontier need not work to screen out (or rather in!) optimal portfolios. For even when each constituent x_i belongs to a common 2-parameter family, the resulting $z = \Sigma\lambda_j x_j$ will not belong to that family or to any 2-parameter family that is independent of the λ weightings. It suffices to show this in the case of statistical independencies. Thus, define the rectangular distribution

$$\begin{aligned} R(x;a,b) &= \frac{x-a}{b-a} \quad , \quad && a \le x \le b \\ &= 0 \quad , && x \le a \\ &= 1 \quad , && x \ge b \end{aligned}$$

and

$$P(x_1,\ldots,x_n) = \prod_{i=1}^{n} R(x_i;a_i,b_i) \quad .$$

Then, of course, $\Sigma\lambda_j x_j = z$ does not satisfy an R distribution but rather a 3n parameter distribution $P(z;\lambda_1,\ldots,\lambda_n;a_1,\ldots,a_n,b_1,\ldots,b_n)$. Let $(\lambda_1^{\dagger},\ldots,\lambda_n^{\dagger})$ be _a_ point on the Markowitz efficiency frontier, with minimum variance $\Sigma\lambda_j^2\sigma_j^2$ subject to $\Sigma\lambda_j\mu_j = \mu$, $\Sigma\lambda_j = 1$. Then it need not be the case that the optimum $(\lambda_1^*,\ldots,\lambda_n^*)$ that maximizes $E[U(z)]$ will belong to the efficiency set $(\lambda_1^{\dagger},\ldots,\lambda_n^{\dagger})$! I do not recall this fact's being mentioned by those who speak of 2-parameter-family justifications of mean-variance analysis. Some quite different argument, such as that $n \to \infty$ and quadratic approximation to U then becomes increasingly good, will be needed to bring back the Markowitz frontier into more general applicability.

I do not wish to end on a nihilistic note. My objections are those of a purist, and my demonstrations in this paper have shown that even a purist can develop diversification theorems of great generality. But in practice, where crude approximations may be better than none, the 2-moment models may be found to have pragmatic usefulness.

FOOTNOTES

1. H. Makower and J. Marschak, "Assets, Prices, and Monetary Theory," Economica N.S. (Vol. V, 1938), pp. 261-88. Harry M. Markowitz, "Portfolio Selection," The Journal of Finance (Vol. VII, 1952), pp. 77-91 and Portfolio Selection: Efficient Diversification of Investments (New York: John Wiley & Sons, 1959); James Tobin, "Liquidity Preference as Behavior Towards Risk," Review of Economic Studies (Vol. XXV, 1958), pp. 65-86. The path-breaking article, E. D. Domar and R. A. Musgrave, "Proportional Income Tax and Risk-taking," Quarterly Journal of Economics (Vol. LVII, 1944), pp. 389-422 replaces variance by risk of loss (mean absolute loss) as dispersion parametersto be avoided. (This paper is reprinted in A.E.A. Selected Readings in Fiscal Policy and Taxation (Homewood, Ill.: Irwin, 1959).

2. P. A. Samuelson, "A Fallacy in the Interpretation of Pareto's Law of Alleged Constancy of Income Distribution," Essays in Honor of Marco Fanno, ed., Tullio Bagiotti, (Padua, Cedam-Casa Editrice Dott. Antonio Milani, 1966), pp. 580-584.

3. Howard Raiffa, unpublished Harvard Business School memos; Marcel K. Richter, "Cardinal Utility, Portfolio Selection and Taxation," Review of Economic Studies (Vol. XXVII, 1959), pp. 152-66; E. C. Brown, "Mr. Kaldor on Taxation and Risk Bearing," Review of Economic Studies (Vol. XXV, 1957), pp. 49-52; J. R. Hicks, "Liquidity," Economic Journal (Vol. LXXII, 1962), pp. 787-802, depicts a rediscovery of some of the Markowitz theory. Also see John Lintner, "Valuation of Risk Assets," Review of Economics and Statistics (Vol. XLVII, 1965), pp. 13-37, and "Optimum Dividends and Uncertainty," Quarterly Journal of Economics (Vol. LXXVIII, 1964), pp. 49-95 and unpublished appendix.