Approximating Expected Utility by a Function of Mean and Variance

Article *in* American Economic Review · January 1979

Source: RePEc

Haim Levy
Hebrew University of Jerusalem

Approximating Expected Utility by a Function of Mean and Variance

By H. LEVY AND H. M. MARKOWITZ*

Suppose that an investor seeks to maximize the expected value of some utility function $U(R)$, where R is the rate of return this period on his portfolio. Frequently it is more convenient or economical for such an investor to determine the set of mean-variance efficient portfolios than it is to find the portfolio which maximizes $EU(R)$. The central problem considered here is this: would an astute selection from the E,V efficient set yield a portfolio with almost as great an expected utility as the maximum obtainable EU?

A number of authors have asserted that the right choice of E,V efficient portfolio will give precisely optimum EU if and only if all distributions are normal or U is quadratic.[1] A frequently implied but unstated corollary is that a well-selected point from the E,V efficient set can be trusted to yield almost maximum expected utility if and only if the invector's utility function is approximately quadratic, or if his a priori beliefs are approximately normal. Since statisticians frequently reject the hypothesis that return distributions are normal, and John Pratt and Kenneth Arrow have each shown us absurd implications of a quadratic utility function, some writers have concluded that mean-variance analysis should be rejected as the criterion for portfolio selection, no matter how economical it is as compared to alternate formal methods of analysis.

Consider, on the other hand, the following evidence to the contrary. Suppose that two investors, let us call them Mr. Bernoulli and Mr. Cramer, have the same probability beliefs about portfolio returns in the forthcoming period; while their utility functions are, respectively,

$$U(R) = log(1 + R) \tag{1}$$

$$U(R) = (1 + R)^{1/2} \tag{2}$$

Suppose that Mr. Cramer and Mr. Bernoulli share beliefs about exactly 149 portfolios. In particular suppose that the 149 values of E, V, $Elog(1 + R)$ and $E(1 + R)^{1/2}$ they share happen to be the same as that of the annual returns observed for 149 mutual funds during the period 1958 through 1967, as reported below. (We are not necessarily recommending unadjusted past data as predictors of the future; rather we are using these observations as one example of "real world" moments.)

Now let us suppose that Mr. Bernoulli, having read William Young and Robert Trent, decides that when he knows the E and the V (or the standard deviation σ) of a distribution he may guess its expected utility to him by the formula:

$$Elog(1 + R) \cong (log(1 + E + \sigma) + log(1 + E - \sigma))/2 \tag{3}$$

He would find that there is a .995 correlation between the pairs (actual mean log$(1 + R)$, estimated mean log$(1 + R)$) for the 149 such pairs provided by the 149 historical distributions. Furthermore, the regression relation (over the sample of 149) between the actual mean $log(1 + R)$ and the estimate provided by (3) is

$$\text{actual} = 0.002 + 0.996 \cdot \text{estimated} \tag{4}$$

As it happens, the portfolio which maximized the approximation (3) also maximized the expected value of the true utility (1). If Mr.

*Jerusalem School of Business Administration, Hebrew University, and Thomas J. Watson Research Center, IBM Corporation, respectively. We have benefited greatly from technical assistance and the helpful comments of Y. Kroll, and from discussions with Marshall Sarnat.

[1] Analyses of relationships between E, V efficiency, on the one hand, and quadratic utility and/or normal distributions, on the other hand, may be found for example in James Tobin (1958, 1963), Markowitz, Martin Feldstein, Giora Hanoch and Levy (1969, 1970), and John Chipman.

Bernoulli selected among the 149 portfolios on the basis of (3) he would, in this instance, do precisely as well as if he had used the true criteria (1). Finally, as will be shown later, (3) always increases with E and decreases with σ, thus is always maximized by an E,V efficient portfolio.

Mr. Cramer, seeing the good fortune of Mr. Bernoulli in finding an approximation to his expected utility based on E and V alone, might try the corresponding approximation to his own utility function, namely:

$$(5)\quad EU \cong (U(1 + E + \sigma) + U(1 + E - \sigma))/2$$

where U is now given by equation (2). Mr. Cramer would be delighted to find that the correlation between predicted and actual for his utility function is .999; the regression relationship is

$$(6)\quad \text{actual} = -.013 + 1.006 \cdot \text{estimated}$$

The portfolio, among the 149, which maximized the approximation (5) also maximized the true expected utility (2); and (2) is always maximized by a portfolio in the E,V efficient set.

Suppose that a third investor, a Mr. X, does not know his current utility function—has just not taken the time recently to analyze it as prescribed by John von Neumann and Oskar Morgenstern—but does know that equation (5) provides about as good an approximation to his utility function as it does to those of Mr. Bernoulli and Mr. Cramer. He also knows, from certain properties which he is willing to assume concerning his utility function, that equation (5) is maximized by an E,V efficient portfolio. If Mr. X can carefully pick the E,V efficient portfolio which is best for him, then Mr. X, who still does not know his current utility function, has nevertheless selected a portfolio with maximum or almost maximum expected utility.

In this paper we present a class of approximations $f_k(E, V, U(\cdot))$ where $k \geq 0$ is a continuous parameter distinguishing one method of approximation from another. For $k = 1$ we get equation (5); for $k = 0$ we have a method proposed by Markowitz. We shall examine some empirical relationships between EU and $f_k(E,V,U(\cdot))$ for various utility functions, empirical distributions, and values of k. We shall explain these empirical results in terms of a simple analysis of the expected difference between a utility function and an approximating function. We shall also consider certain objections to E, V analysis, due to Karl Borch, Pratt, and Arrow in light of our empirical results, our analysis of expected difference, and a reconsideration of Pratt's analysis of risk aversion for the kinds of quadratic approximations we use.

I. A Class of Approximations

Markowitz used two methods to approximate EU by a function $f(E, V)$ depending on E and V only. The first is based on Taylor-series around zero:

$$(7)\quad U = U(0) + U'(0)R + .5U''(0)R^2 \dots$$

hence

$$(8)\quad EU \cong U(0) + U'(0)E + .5U''(0)(E^2 + V)$$

The second approximation is based on a Taylor-series around E:[2]

$$(9)\quad U = U(E) + U'(E)(R - E) + .5U''(E)(R - E)^2 \dots$$

hence

$$(10)\quad EU \cong U(E) + .5U''(E)V$$

In tests with empirical distributions and the logarithmic utility function (by Markowitz, and by Young and Trent) the approximation in (10) performed markedly better than the approximation in (8).

Both approximations involve fitting a quadratic to $U(R)$ based on properties of U (i.e., U, U', and U'') at only one value of R ($=0$ or E, respectively). The present authors conjectured that a better approximation perhaps could be found by fitting the quadratic to three judiciously chosen points

[2]A somewhat different use of this Taylor-series to justify mean-variance analysis is presented by S. C. Tsiang. See also Levy regarding the Tsiang analysis.

on $U(R)$. To produce a mean-variance approximation the three points must themselves be functions of at most E, V and the function $U(\ )$. A class of such functions was selected in which the quadratic was fit to the three points

$$(E - k\sigma, U(E - k\sigma)), (E, U(E)), (E + k\sigma, U(E + k\sigma)) \tag{11}$$

The quadratic passing through these three points can be written as

$$Q_k(R) = a_k + b_k(R - E) + c_k(R - E)^2 \tag{12}$$

To simplify notation we will often write Q, a, and b for Q_k, a_k, and b_k, the subscript k being understood. Equation (12) implies

$$EQ = a + cV \tag{13}$$

Solving

$$\begin{aligned} U(E - k\sigma) &= a + b((E - k\sigma) - E) + c((E - k\sigma) - E)^2 \\ &= a - bk\sigma + ck^2\sigma^2 \\ U(E) &= a + b0 + c0^2 \\ U(E + k\sigma) &= a + bk\sigma + ck^2\sigma^2 \end{aligned} \tag{14}$$

we find that

$$\begin{aligned} a &= U(E) \\ b &= \frac{U(E + k\sigma) - U(E - k\sigma)}{2k\sigma} \\ c &= \frac{U(E + k\sigma) + U(E - k\sigma) - 2U(E)}{2k^2\sigma^2} \end{aligned} \tag{15}$$

hence

$$\begin{aligned} f_k(E, V, U(\cdot)) &= EQ \\ &= U(E) + \frac{U(E + k\sigma) + U(E - k\sigma) - 2U(E)}{2k^2} \end{aligned} \tag{16}$$

If we substitute $k = 1$ and simplify we obtain equation (5). If we define the approximation in (10) as f_0, and let $k \to 0$ in (16) we find that[3]

$$f_k \to f_0 \quad \text{as} \quad k \to 0 \tag{17}$$

For given k and U, (16) clearly depends on only E and V. It is not immediately clear that f is always maximized by an E,V efficient portfolio. Each of the utility functions used in our experiments has $U' > 0$, $U'' < 0$, and $U''' \geq 0$ for all rates of return $R > -1.0$. These three properties are sufficient to assure us that f is maximized by an E,V efficient portfolio, provided that $E - k\sigma > -1$ for all portfolios considered.[4]

II. Analysis of Error Functions

For a given k, and for any probability distribution for which the specified moments exist, the difference between EU and $f(E, V, U(\cdot))$ may be written as

$$\begin{aligned} D_k &= EU - EQ_k \\ &= Ed_k(R; E, V, U(\cdot)) \end{aligned} \tag{18}$$

where Q_k is given in (12) and

$$d_k(R) = U(R) - Q_k(R) \tag{19}$$

[3]That $f_k \to f_0$ as $k \to 0$ follows readily if we show that

$$\frac{U(E + k\sigma) + U(E - k\sigma) - 2U(E)}{k^2} \to U''(E)V$$

This may be shown by computing the second derivative of the numerator with respect to k, the second derivative of the denominator, and applying L'Hospital's rule.

[4]Differentiating (16) with respect to σ we find, since $U'' < 0$, that

$$\frac{\partial f_k}{\partial \sigma} = \frac{U'(E + k\sigma) - U'(E - k\sigma)}{2k} < 0$$

Differentiating (16) with respect to E, and substituting $\xi = k\sigma$, we find that

$$\text{(a)} \quad \partial f_k/\partial E = U'(E) + [U'(E + \xi) + U'(E - \xi) - 2U'(E)]V/2\xi^2$$

For positive or negative η Taylor's theorem implies

$$\text{(b)} \quad U'(E + \eta) = U'(E) + U''(E)\eta + .5U'''(E + \theta)\eta^2$$

for some θ between 0 and η. Substituting (b) for $U'(E + \xi)$ and $U'(E - \xi)$ in (a), and using $U' > 0$ and $U''' \geq 0$, we find (for some θ_1 and θ_2 between 0 and ξ) that $\partial f_k/\partial E = U'(E) + [U'''(E + \theta_1) + U'''(E - \theta_2)]V/4 > 0$.

TABLE 1—$d_k(R) = U(R) - Q_k(R)$ FOR $U = log_e(1 + R)$; $E = .1$; $k = 0$ or 1; AND $\sigma = .15$

R	U(R)	$Q_0(R)$	$U - Q_0$	$Q_1(R)$	$U - Q_1$
-.70	-1.20397	-.89643	-.30755	-.90348	-.30049
-.50	-.69315	-.59891	-.09424	-.60373	-.08942
. . .					
-.30	-.35667	-.33444	-.02223	-.33735	-.01933
-.25	-.28768	-.27349	-.01419	-.27596	-.01172
-.20	-.22314	-.21461	-.00854	-.21667	-.00648
-.15	-.16252	-.15779	-.00473	-.15946	-.00306
-.10	-.10536	-.10304	-.00232	-.10433	-.00103
-.05	-.05129	-.05035	-.00094	-.05129	.00000
.00	.00000	.00027	-.00027	-.00034	.00034
.05	.04879	.04882	-.00003	.04853	.00026
.10	.09531	.09531	-.00000	.09531	-.00000
.15	.13976	.13973	.00003	.14001	-.00024
.20	.18232	.18209	.00023	.18262	-.00030
.25	.22314	.22238	.00077	.22314	.00000
.30	.26236	.26060	.00176	.26158	.00078
.35	.30010	.29676	.00335	.29794	.00217
.40	.33647	.33085	.00563	.33221	.00427
.45	.37156	.36287	.00869	.36439	.00718
.50	.40546	.39283	.01263	.39449	.01098
.55	.43825	.42072	.01753	.42250	.01576
.60	.47000	.44655	.02345	.44842	.02158
. . .					
1.00	.69315	.57878	.11437	.58075	.11240
1.50	.91629	.55812	.35817	.55846	.35783
2.00	1.09861	.33086	.76775	.32761	.77100
2.50	1.25276	-.10301	1.35577	-.11179	1.36455
3.00	1.38629	-.74349	2.12978	-.75975	2.14604

For example, Table 1 presents $d_k(R)$ for $U = log(1 + R)$, for $k = 0$ and $k = 1$, for $E = .1$, for $\sigma = .15$, and for various values of R. Much about the joint distribution of EU and $f_k(E, V, U(\cdot))$ is explained by such tables plus general properties of the distributions involved. Consider the fourth column of Table 1 showing $d_k(R)$ for $k = 0$. Among the 149 mutual fund distributions mentioned earlier, those with E in the neighborhood of .10 all have every year's return between a 30 percent loss and a 60 percent gain for the year. (For example, 64 distributions had $.08 \leq E \leq .12$; all were within the range indicated.) $d_0(R)$ goes from $-.022233$ at $R = -.3$, to $+.023454$ at $R = .6$, with substantially smaller values of $|d_0|$ in between. If we imagine spreading a probability distribution throughout the interval $-.3$ to $+.6$, keeping $E = .1$, there is a limit to how large we can make $|Ed_0|$. In fact, if we assume that the distribution can take on only the values listed in the table ($-.3$ to $+.6$ by steps of .05), then a little linear programming will show us that the worst distribution, the one with the greatest $|E(d)|$, is one with a probability of 3/8 of $R = -.3$, a probability of 5/8 of $R = +.35$, and with $E(d) = -.00649$. None of the 149 historical distributions had this worst possible distribution. Insofar as they were less skewed, positive errors above E tended to cancel negative errors below E. Insofar as they clustered closer to the mean than in our worst case, the absolute value of the deviations were smaller.

If we recomputed Table 1 for some other E, say $E = .15$, the new table would appear very much like the old. The principal difference in column 4 would be that the smallest values of $|d_0(R)|$ would be centered around the new mean, $R = .15$. An analysis similar to that above would again explain why $|Ed|$ was small for those historical distributions which had E in this new neighborhood.

We should examine the concept of $|Ed|$ being small. A statement to the effect that the difference between EU and EQ is less than some number (like .0065) is, in itself, of absolutely no value either in explaining the correlation between EU and EQ, or in judging whether Q is a good approximation to U in practice. For a utility function is only defined up to an arbitrary choice of unit and scale. In particular, if we multiply U by an arbitrary positive constant, obtaining a precisely equivalent utility function, we also multiply by the same constant the approximations (8), (10), and (16), and therefore multiply by the same constant the difference between EU and each of these. Thus we can make $|Ed|$ arbitrarily close to zero by using the utility function $V = bU(R)$ for sufficiently small b.

The arbitrary choice of unit and scale, however, does not change certain measurements: the correlation and regression coefficients are unaffected; and the appearance of a comparison between the plot of U against R vs. a plot of Q against R is also unaffected in the following sense. Suppose that we plot U in Table 1 against R for $R = -.3$ to $+.6$. Since U rises from about $-.36$ to about $+.47$ we might allow .1 utility units per inch of vertical scale. If we also plot on the same graph $Q_0(R)$ from the third column of the table, Q would be two-tenths of an inch above U at

$R = -.3$, about two-tenths of an inch below U at $R = .6$; but for much of their lengths the two curves would be virtually indistinguishable as they rose from the lower left to the upper right-hand corner of the page. Suppose now that we change the origin and scale of the utility measure. If we still want the two curves to fill the page we simply relabel the vertical axis leaving the two curves unchanged.

If we define σ_{Ed} to be the standard deviation of Ed over a set of distributions, where Ed is the mean value of d for a given distribution, and similarly define $\sigma_f = \sigma_{EQ}$ as the standard deviation of f over the set of distributions, then it can be shown[5] that the correlation $\rho_{EU,f}$ between EU and f over the set of distributions is at least

$$\rho_{EU,f} \geq (1 - \gamma^2)^{1/2} \tag{20}$$

where

$$\gamma = \sigma_{Ed}/\sigma_f \tag{21}$$

Thus if f is ten times as variable as Ed then $\rho_{EU,f}$ is at least $(.99)^{1/2} = .995$.

Column 6 of Table 1 presents $d_1(R)$. This equals zero at $R = E - \sigma$, $R = E$, and $R = E + \sigma$ as planned. Just considering $-.3 \leq R \leq .6$, the f_1 approximation is definitely superior to f_0 in the range from $R = -.05$ through $-.3$, and from $+.25$ to $.6$. On the other hand, f_0 fits better near the mean, i.e., from .00 through .20 among the values listed. The empirical results presented in the following sections indicate, for the distributions and utility functions explored, whether it was more beneficial to approximate a bit better near the mean, as does f_0, or to hold up well over wider range as does f_1.

Table 1 also shows values of $d_0 = U - Q_0$ and $d_1 = U - Q_1$ for more extreme values of $|R - E|$ than discussed thus far. We see for example that, for an investor with a logarithmic utility function, f_0 is likely to be a poor approximation to EU for a distribution with $E = .1$ and with nontrivial probabilities of, say, $R = -.7$ and $R = 1.5$. More generally, the empirical results reported below would be less favorable to mean-variance approximations if we were dealing with much more speculative distributions. The mean-variance analysis is thus more suitable for investor's and opportunity sets in which such extremes have very low probabilities in the portfolios which maximize both EU and f_k.

III. Empirical Results

For annual returns of 149 investment companies, 1958–67,[6] Table 2 shows the correlation between $EU(R)$ and f_k $(E, V, U(\cdot))$ for $k = .01, .1, .6, 1.0$, and 2.0 and for various utility functions. (We also computed correlation coefficients for a few other values of the a and b coefficients in the utility functions, with results which one might expect by interpolating or extrapolating the results reported in Table 2. For example, the exponential utility with $b = 20$ was even more of an exception to the general rule than is the case with $b = 10$ reported here.) A row of Table 2 presents correlation ρ as a function of k, for some given utility function. In every case reported in Table 2, with the exception of the exponential utility function with $b = 10$, ρ is a nonincreasing function of k; hence $\rho_{.01} \geq \rho_k$ for all k. Note ρ_k, as a function of k, is frequently quite flat between $k = .01$ and 1.0, but drops faster from $k = 1$ to $k = 2$. We did not calculate the $\rho_{0.0}$ correlations but, from continuity considerations, we assume

[5]From (19) and $f = EQ$ it follows that

$$\rho_{EU,f} = cov\ (f, f + Ed)/(\sigma_f \sigma_{f+Ed}) \tag{a}$$

A short calculation shows that $cov\ (f, f + Ed) = var\ (f) + \rho_{f,Ed}\sigma_f\sigma_{Ed} = var\ (f)(1 + \gamma\rho_{f,Ed})$, and that $\sigma_{f+Ed} = \sigma_f(1 + \gamma^2 + 2\gamma\rho_{f,Ed})^{1/2}$. Substituting these two formulas into (a) we get

$$\rho_{EU,f} = (1 + \gamma\rho_{f,Ed})/(1 + \gamma^2 + 2\gamma\rho_{f,Ed})^{1/2} \tag{b}$$

For $\gamma < 1$ this is a continuous, positive function of $\rho_{f,Ed}$ in the range $-1 \leq \rho_{f,Ed} \leq +1$. If we differentiate (b) with respect to $\rho_{f,Ed}$, set the resulting expression equal to zero, assume $0 < \gamma < 1$, and solve for $\rho_{f,Ed}$, we find that $\partial\rho_{EU,f}/\partial\rho_{f,Ed} = 0$, only at $\rho_{f,Ed} = -\gamma$. Substituting this into (b) we find that at this stationary point $\rho_{EU,f} = (1 - \gamma^2)^{1/2}$. We may confirm that is a minimum rather than a maximum or inflection point by noting, from (b), that $\rho_{EU,f} = +1 > (1 - \gamma^2)^{1/2}$ for $\rho_{f,Ed} = \pm 1$.

[6]The annual rate of return of the 149 mutual funds are taken from the various annual issues of A. Wiesenberger and Company. All mutual funds whose rates of return are reported in Wiesenberger for the whole period 1958–67 are included in the analysis.

TABLE 2—CORRELATION BETWEEN $EU(R)$ AND $f_k(E, V, U(\cdot))$ FOR ANNUAL RETURNS OF 149 MUTUAL FUNDS, 1958–67

Utility Function		k = 0.01	0.1	0.6	1.0	2.0
Log(1 + R)		0.997	0.997	0.997	0.995	0.983
$(1+R)^a$	a=0.1	0.998	0.998	0.997	0.997	0.988
	a=0.3	0.999	0.999	0.999	0.998	0.995
	a=0.5	0.999	0.999	0.999	0.999	0.998
	a=0.7	0.999	0.999	0.999	0.999	0.999
	a=0.9	0.999	0.999	0.999	0.999	0.999
$-e^{-b(1+R)}$	b=0.1	0.999	0.999	0.999	0.999	0.999
	b=0.5	0.999	0.999	0.999	0.999	0.999
	b=1.0	0.997	0.997	0.997	0.996	0.995
	b=3.0	0.949	0.949	0.941	0.924	0.817
	b=5.0	0.855	0.855	0.852	0.837	0.738
	b=10.	0.447	0.449	0.503	0.522	0.458

that they are close to those found for $k = .01$. For most cases considered $\rho_{.01} > .99$.

The correlations for the exponential with $b = 10$ are much lower than those of the other utility functions reported in Table 2. In our 1977 paper we analyze this utility function at a greater length than space permits here, and arrive at two conclusions. The first conclusion is that an investor who had $-e^{-10(1+R)}$ as his utility function would have some very strange preferences among probabilities of return. Reasonably enough, he would not insist on certainty of return. For example, he would prefer (a) a 50–50 chance of a 5 percent gain vs. a 25 percent gain rather than have (b) a 10 percent gain with certainty. On the other hand there is no R which would induce the investor to take (a) a 50-50 chance of zero return (no gain, no loss) vs. a gain of R rather than have (b) a 10 percent return with certainty. Thus a 50-50 chance of breaking even vs. a 100 percent, or 300 percent, or even a 1000 percent return, would be considered less desirable than a 10 percent return with certainty. We believe that few if any investors have preferences anything like these. A second conclusion, more important than the first as far as the present discussion is concerned, is that even if some unusual investor did have the utility function in question, if he looked at his $d_k(R)$ in advance he would be warned of the probable inapplicability of mean-variance analysis. The corresponding version of Table 1 (scaled to have about the same σ_f for the two approximations, as (20) suggests) shows d_k to generally be more than an order of magnitude greater for the exponential with $b = 10$ than for the logarithmic utility function.

TABLE 3—CORRELATION BETWEEN $EU(R)$ AND $f_{.01}(E, V, U(\cdot))$ FOR 3 HISTORICAL DISTRIBUTIONS

Utility Function		Annual returns on 97 stocks[8]	Monthly returns on 97 stocks[8]	Random portfolios of 5 or 6 stocks[9]
Log(1 + R)		0.880	0.995	0.998
$(1+R)^a$	a=0.1	0.895	0.996	0.998
	a=0.3	0.932	0.998	0.999
	a=0.5	0.968	0.999	0.999
	a=0.7	0.991	0.999	0.999
	a=0.9	0.999	0.999	0.999
$-e^{-b(1+R)}$	b=0.1	0.999	0.999	0.999
	b=0.5	0.961	0.999	0.999
	b=1.0	0.850	0.997	0.998
	b=3.0	0.850	0.976	0.958
	b=5.0	0.863	0.961	0.919
	b=10.	0.659	0.899	0.768

Table 3 shows the correlation between EU and $f_{.01}$ for three more sets of historical distributions. While ρ_k was computed for the same values of k reported in Table 2, we confine our attention to $k = .01$ since, almost without exception, ρ_k was a nonincreasing function of k. The first column of data in Table 3 shows $\rho_{.01}$ for annual returns on 97 *U.S.* common stocks during the years 1948–68.[7] It is understood, of course, that mean-variance analysis is to be applied to the portfolio as a whole rather than individual investments taken one at a time. Annual returns on individual stocks were used in this example, nevertheless, as an example of historic distributions with greater variability than that found in the portfolios reported in Table 2. As expected, the correlations are clearly poorer for the individual stocks than they are for the mutual fund portfolios. For $U = log(1 + R)$, for example, the correlation is .880 for the annual returns on stocks as

[7]This data base of 97 *U.S.* stocks, available at Hebrew University, had previously been obtained as follows: a sample of 100 stocks was randomly drawn from the CRSP (Center for Research in Security Prices, University of Chicago) tape, subject to the constraint that all had reported rates of return for the whole period 1948–68. Some mechanical problems reduced the usable sample size from 100 to 97. The inclusion only of stocks which had reported rates of return during the whole period may have introduced selection bias into the sample. It might prove worthwhile to experiment with alternate methods of handling the appearance and disappearance of stocks.

compared to .997 for the annual returns on the mutual funds.

Since monthly returns tend to be less variable than annual returns we would expect the correlations to be higher for the former than the latter. The $\rho_{.01}$ for monthly returns on the same 97 stocks are shown in the second column of data in Table 3. For the logarithmic utility function, for example, the correlation is .995 for the monthly returns on individual stocks as compared to .880 for annual returns on the stocks and .997 for annual returns on the mutual funds. On the whole, the $\rho_{.01}$ for the monthly returns on individual stocks are comparable to the annual returns on the mutual funds.

Annual returns on individual stocks (i.e., on completely undiversified portfolios) have perceptibly smaller $\rho_{.01}$ than do the annual returns on the well diversified portfolios of mutual funds. The third column of data in Table 3 presents $\rho_{.01}$ for "slightly diversified" portfolios consisting of a few stocks. Specifically, it shows the correlations between EU and $f_{.01}$ on the annual returns for nineteen portfolios of 5 or 6 stocks each randomly drawn (without replacement) from the 97 *U.S.* stocks.[8] We see that for the logarithmic utility function $\rho_{.01} = .998$ for the random portfolios of 5 or 6, up from .880 for individual stocks. The $\rho_{.01}$ for the annual returns on the portfolios of 5 or 6 were generally comparable to those for the annual returns on the mutual funds. These results were perhaps the most surprising of the entire analysis. They indicate that, as far as the applicability of mean-variance analysis is concerned, at least for joint distributions like the historical returns on stocks for the period analyzed, a little diversification goes a long way.

In addition to the correlation coefficient, we examine in our 1977 paper other measures of the ability of f_k to serve as a surrogate for EU, and conclude that $f_{.01}$ does as well in these comparisons as it does in terms of correlations. For example, we computed the frequency with which any other available portfolio was better than the portfolio which maximized f_k. We found, in particular, that in every case with $\rho > .9$ in Tables 2 or 3 the portfolio with maximum $f_{.01}$ was also the portfolio (among the 149 or 97 or 19 considered) with the greatest EU. We cannot say with any precision how high a correlation between f_k and EU is high enough. Be that as it may, for many of the utility functions and distributions considered (chosen in advance as representative of utility functions frequently postulated, or distributions clearly "real world") $f_{.01}$ was an almost faultless surrogate for EU. Where $f_{.01}$ performed poorly, the user would have been warned in advance by an analysis of expected error.

[8]We randomly drew 5 stocks to constitute the first portfolio; 5 different stocks to constitute the second portfolio, etc. Since we have 97 stocks in our sample, the eighteenth and nineteenth portfolios include 6 stocks each. Repetition of this experiment with new random variables produced negligible variations in the numbers reported, except for the case of $U=e^{-10(1+R)}$. A median figure is reported in the table for this case.

IV. Some Objections Reconsidered

Since $\rho_{EU,f} < 1$ it can happen that portfolio A has a higher f_k, than portfolio B, while portfolio B has a higher EU. In fact, given any function f of E and V, Borch presents a method for finding distributions A and B such that $f(E_A, \sigma_A) = f(E_B, \sigma_B)$, yet clearly $EU_A > EU_B$ because distribution A stochastically dominates distribution B. (In terms of equation (20), this example has $\gamma = \infty$.)

Borch's argument shows that it is hopeless to seek an f that will be perfectly correlated with EU for all collections of probability distributions. The evidence on the preceding pages nevertheless supports the notion that the imperfect approximations $f_k(E, V)$ are frequently good enough in practice to choose almost optimally among realistic distributions of returns; and the d_k function can be used in advance to judge the suitability of f_k.

The approximation f_k was obtained by fitting the quadratic (12) to three points. We have seen that for certain utility functions and historical distributions of returns, f_k is highly correlated with EU. Pratt and Arrow, on the other hand, have shown that any quadratic utility function had highly undesirable theoretical characteristics. How do we reconcile these two apparently contradictory observations?

It is essential here to distinguish between three types of quadratic approximations:

1) Assuming that utility as a function of wealth $V(W)$ remains constant from period to period, a quadratic $q(W)$ is fit to $V(W)$ once and for all. As W changes from time to time, the same $q(W)$ function is used to select portfolios. (Note that V expresses utility as a function of wealth W, in contrast to the previously defined U function which expressed utility as a function of rate of return R. Note also that an unchanging $V(W)$ implies the existence of some $U(R)$ at each point in time, though not necessarily the same $U(R)$ each time. The converse is not true: the existence of a $U(R)$ at each point in time does not necessarily imply an unchanging $V(W)$.)

2) At the beginning of each time period a $q(W)$ function is fit to $V(W)$ at the point $W =$ current wealth; or equivalently $Q(R)$ is fit to $U(R) = V((1 + R)W_{t-1})$ at $R = 0$. Even if $V(W)$ remains constant through time, the quadratic fit $Q(R)$ is changed as wealth changes.

3) The quadratic fit (of $Q(R)$ to $U(R)$) depends on at least E and perhaps σ as well. In this case the fit varies between one distribution and another being evaluated at the same time.

The Pratt and Arrow objections apply to quadratic approximations of type 1.[9] The approximation in (8) is of type 2. Approximations (10) and (16) are of type 3. Each of these three classes of approximations have different risk-aversion properties. We shall confine our attention here to a comparison between the first and third of these. To illustrate the difference between the first and third, we shall compare their "risk premiums" for small risks as defined by Pratt.

Pratt's results may be expressed as follows. Let $Pr_i(R)$ $i = 1, 2, 3, \ldots$ be a sequence of probability distributions such that each has the same mean:

$$(22a)\quad \int R\,dPr_i(R) = E_0 \qquad i = 1, 2, 3, \ldots$$

and such that their standard deviations approach zero:

$$(22b)\qquad \lim \sigma_i = 0$$

where the limit, here and elsewhere in this section unless otherwise specified, is taken as $i \rightarrow \infty$. The risk premium for the ith distribution is defined implicitly by

$$(22c)\quad U(E_0 - \pi_i) = \int U(R)\, dPr_i(R) \qquad i = 1, 2, 3, \ldots$$

where the right-hand side equals the expected utility of the ith distribution. Pratt shows that, under certain general conditions,

$$(23)\quad \begin{aligned} \lim \pi_i/\sigma_i^2 &= -1/2V''((1 + E_0)\, W_{t-1}) \\ &\qquad /V'((1 + E_0)W_{t-1}) \\ &= -1/2U''(E_0)/U'(E_0) \\ &= 1/2r((1 + E_0)W_{t-1}) \end{aligned}$$

where r is defined to be the investor's risk aversion at the wealth associated with return $R = E_0$. Pratt reasonably asserts that π/σ^2, and hence $r((1 + E_0)W_{t-1})$ should be a decreasing function of its one argument, $W = (1 + E_0)W_{t-1}$. He notes that for a quadratic (of type 1) $r(W)$ is an increasing function of W and concludes:

> Therefore a quadratic utility cannot be decreasingly risk-averse on any interval whatever. This severely limits the usefulness of quadratic utility, however nice it would be to have expected utility depend only on the mean and variance of the probability distribution. Arguing "in the small" is no help: decreasing risk aversion is a local property as well as a global one. [p. 132]

That Pratt's conclusion is not correct for an approximation of type 3 will be shown in the following way. If an investor maximized some mean-variance approximation $f(E,\sigma)$, such as f_k for fixed nonnegative k, then his risk premium would be given implicitly by

$$(24)\qquad f(E - \pi, 0) = f(E, \sigma)$$

[9]Pratt correctly asserts that his analysis does not require constant $V(W)$ over time. We must distinguish, however, between a V function varying with time (but not depending on W itself) vs. approximations of type 2 or 3 in which the choice of the quadratic function depends on W or even E and σ. For simplicity, we describe approximations of type 1 in terms of a fixed $V(W)$, rather than try for greater generality here.

Below we define a class of type 3 approximations including, in particular, f_k for all $k \geq 0$. We consider a sequence of probability distributions $Pr_i(R)$ $i = 1, 2, 3, \ldots$ satisfying (22a) and (22b), and find that

$$(25) \qquad \lim \pi_i/\sigma_i^2 = -1/2U''(E_0)/U'(E_0)$$

for all approximations in the class. Thus the risk aversion "in the small" of f_k is precisely the same as that of U itself. We have seen that not all f_k are equally good "in the large." But in the small, they are asymptotically the same as the utility function which they approximate.

We assume that the investor maximizes an approximation $f(E,\sigma)$ which is the expected value of a quadratic of type 3 such that $f = EQ$ in (13) satisfies

$$(26a) \qquad a = U(E)$$

$$(26b) \qquad c \rightarrow .5U''(E) \text{ as } \sigma \rightarrow 0$$

It follows immediately from (10) that f_0 satisfies (26a) and (26b), and from (15) that f_k $k > 0$ satisfies (26a). That f_k also satisfies (26b) follows from L'Hopital's rule applied to the expression for c in (15).

From (13) and (26a), (24), (13) again, and (26b) we infer that

$$\begin{aligned} U(E_0 - \pi) &= f(E_0 - \pi, 0) \\ &= f(E_0, \sigma) \\ &= U(E_0) + c(E_0, \sigma)\sigma^2 \\ &\quad \rightarrow U(E_0) \text{ as } \sigma \rightarrow 0 \end{aligned}$$

But $U(E_0 - \pi) \rightarrow U(E_0)$, and $U' > 0$ throughout, implies

$$(27) \qquad \lim \pi_i = 0$$

Using Taylor's theorem we may write

$$\begin{aligned} (28) \quad f(E_0 - \pi, 0) &= U(E_0 - \pi) \\ &= U(E_0) - U'(E_0)\pi \\ &\quad + .5U''(\xi)\pi^2 \end{aligned}$$

where ξ is between $E_0 - \pi$ and E_0. Hence, from (27), $\xi \rightarrow E_0$ as $i \rightarrow \infty$. Using (28) on the left of (24), (13) on the right of (24), and rearranging terms we get

$$(29) \quad \pi/\sigma^2 = -c(E_0,\sigma)/(U'(E_0) - .5U''(\xi)\pi)$$

for each distribution in the sequence. This together with (26b) and (27) implies (25).

V. The *E,V* Investor

Let us return to Mr. X who has not analyzed his utility function recently. Suppose that, when presented with probability distributions of E,V efficient portfolios, he can pick that portfolio which has greater EU than any other E,V efficient portfolio. By definition, his choice will be at least as good as the portfolio which maximizes $f_{.01}$. In addition to the functions of E and V discussed above there may very well be others—not yet explored, or perhaps not yet conjectured—which perform better than does any f_k as a surrogate for EU. Mr. X's choice of portfolio will also be at least as good as the best of all of these in each particular situation.

REFERENCES

Kenneth Arrow, *Aspects of the Theory of Risk Bearing*, Helsinki 1965.

K. Borch, "A Note on Uncertainty and Indifference Curves," *Rev. Econ. Stud.*, Jan. 1969, *36*, 1–4.

J. S. Chipman, "The Ordering of Portfolios in Terms of Mean and Variance," *Rev. Econ. Stud.*, Apr. 1973, *40*, 167–90.

M. S. Feldstein, "Mean Variance Analysis in the Theory of Liquidity Perference and Portfolio Selection," *Rev Econ. Stud.*, Jan. 1969, *36*, 5–12.

G. Hanoch and H. Levy, "The Efficiency Analysis of Choices Involving Risk," *Rev. Econ. Stud.*, July 1969, *36*, 335–46.

______ and ______, "Efficient Portfolio Selection with Quadratic and Cubic Utility," *J. Bus., Univ. Chicago*, Apr. 1970, *43*, 181–89.

H. Levy, "The Rationale of the Mean-Standard Deviation Analysis: Comment," *Amer. Econ. Rev.*, June 1974, *64*, 434–41.

______ and H. Markowitz, "Mean-Variance Approximations to Expected Utility," T. J. Watson Res. Ctr., IBM, Aug. 1977.

Harry Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, New York 1959; New Haven 1970.

J. W. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica*, Jan. 1964, *32*, 122–36.

J. Tobin, "Liquidity Preference as Behavior Towards Risk," *Rev. Econ. Stud.*, Feb. 1958, *25*, 65–87.

———, "The Theory of Portfolio Selection" in Frank H. Hahn and Frank P. R. Brechling, eds., *Theory of Interest Rates*, New York 1963.

S. C. Tsiang, "The Rationale of the Mean Standard Deviation Analysis, Skewness Preference, and the Demand for Money," *Amer. Econ. Rev.*, June 1972, *62*, 354–71.

John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior*, 3d ed., Princeton 1953.

W. E. Young and R. H. Trent, "Geometric Mean Approximation of Individual Security and Portfolio Performance," *J. Finan. Quant. Anal.*, June 1969, *4*, 179–99.

A. Wiesenberger and Company, *Investment Companies*, New York, annual editions.